EDITED BY
EMILY BALCETIS AND
GORDON B. MOSKOWITZ

THE HANDBOOK OF IMPRESSION FORMATION

A Social Psychological Approach

ROUTLEDGE

The Handbook of Impression Formation

Presenting diverse perspectives from eminent scholars and contemporary researchers, *The Handbook of Impression Formation* contextualizes current and future areas of research in the social psychology of impression formation within a rich historic framework.

Affirming that impression formation is at the core of human experience, chapters explore how and why people form snap judgments about others and when those impressions update. They examine the processes through which people infer the reasons for the events they encounter, allowing people to plan for appropriate behavioral responses to social contexts. The research reviewed is informed by the foundational theory of unconscious automatic processes involved in making judgments of other people, pioneered by Professor Jim Uleman who contributes a chapter that suggests important new directions, and concludes the volume by reflecting on the state of the field more broadly. This book explores how certain attributes stimulate categorization, examining current issues around implicit bias, stereotypes, and social media. Chapters cover a range of approaches, featuring personal narratives, presentation of new data and discoveries, comprehensive literature reviews, and contemplations on where the field must go and what questions require focus for progress to be made, calling for even the most advanced scholars to contribute more to the collective investigation of impression formation.

This fascinating work provides a solid foundation from which all researchers can build a new and unique program of research, and arms the reader with the intellectual tools they need to chart new theoretical territory and discover aspects of the human experience we have yet to even wonder about. It is essential reading for students and academics in social psychology, and the social sciences more broadly.

Emily Balcetis, director of the New York University Social Perception Action and Motivation research lab, earned her PhD at Cornell University and leads an international team to uncover strategies that increase, sustain, and direct people's efforts to meet their goals.

Gordon B. Moskowitz conducts research on social cognition, with a focus on stereotyping, impression formation, minority influence, and the implicit influence of goals on judgment and behavior. His research program more recently has examined interventions to control/reduce implicit bias, with implications for group disparities in health care.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

The Handbook of Impression Formation

A Social Psychological Approach

Edited by Emily Balcetis and Gordon B. Moskowitz

Routledge
Taylor & Francis Group
NEW YORK AND LONDON

Cover image: Kate Uleman artwork

First published 2023
by Routledge
605 Third Avenue, New York, NY 10158

and by Routledge
4 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

Routledge is an imprint of the Taylor & Francis Group, an informa business

Library of Congress Cataloging-in-Publication Data
Names: Balcetis, Emily, editor. | Moskowitz, Gordon B., editor.
Title: The handbook of impression formation : a social psychological approach / edited by Emily Balcetis, Gordon B. Moskowitz.
Description: New York, NY : Routledge, 2022. | Includes bibliographical references and index. |
Identifiers: LCCN 2022016962 (print) | LCCN 2022016963 (ebook) | ISBN 9780367493141 (paperback) | ISBN 9780367493158 (hardback) | ISBN 9781003045687 (ebook)
Subjects: LCSH: Impression formation (Psychology)
Classification: LCC HM1081 .H36 2022 (print) | LCC HM1081 (ebook) | DDC 153.6--dc23/eng/20220425
LC record available at https://lccn.loc.gov/2022016962
LC ebook record available at https://lccn.loc.gov/2022016963

ISBN: 978-0-367-49315-8 (hbk)
ISBN: 978-0-367-49314-1 (pbk)
ISBN: 978-1-003-04568-7 (ebk)

DOI: 10.4324/9781003045687

Typeset in Goudy
by MPS Limited, Dehradun

Contents

Contributors

Angel D. Armenta, Psychology Department, The University of Texas at El Paso, El Paso, Texas, United States of America

Amy Arndt, Department of Psychology, Univesity of Texas at Austin, Austin, Texas, United States of America

Emily Balcetis, Department of Psychology, New York University, New York, New York, United States of America

Mahzarin R. Banaji, Department of Psychology, Harvard University, Cambridge, Massachusetts, United States of America

John A. Bargh, Department of Psychology, Yale University, New Haven, Connecticut, United States of America

Skylar M. Brannon, Department of Psychology, Univesity of Texas at Austin, Austin, Texas, United States of America

Jessica R. Bray, Psychology Department, The University of Texas at El Paso, El Paso, Texas, United States of America

Kimberly E. Chaney, Department of Psychological Sciences, University of Connecticut, Storrs, Connecticut, United States of America

Jacqueline M. Chen, Department of Psychological Sciences, University of Connecticut, Storrs, Connecticut, United States of America

Cydney H. Dupree, School of Management, University College London, London, England, United Kingdom

Melissa Ferguson, Department of Psychology, Yale University, New Haven, Connecticut, United States of America

Mário B. Ferreira, Center for Research in Psychological Science (CICPsi), Faculdade de Psicologia – Universidade de Lisboa, Lisbon, Portugal

Leonel Garcia-Marques, Center for Research in Psychological Science (CICPsi), Faculdade de Psicologia – Universidade de Lisboa, Lisbon, Portugal

Bertram Gawronski, Department of Psychology, Univesity of Texas at Austin, Austin, Texas, United States of America

Grace S. R. Gillespie, Department of Psychology, University of California, Los Angeles, Los Angeles, California, United States of America

Sara Hagá, Center for Research in Psychological Science (CICPsi), Faculdade de Psicologia – Universidade de Lisboa, Lisbon, Portugal

David L. Hamilton, The Department of Psychological & Brain Sciences, University of California, Santa Barbara, Santa Barbara, California, United States of America

J. Kiley Hamlin, Department of Psychology, University of British Columbia, Vancouver, British Columbia, Canada

Lasana T. Harris, Division of Psychology and Language Sciences, University College London, London, England, United Kingdom

Marlone Henderson, Department of Psychology, Univesity of Texas at Austin, Austin, Texas, United States of America

Kerri L. Johnson, Deptartments of Communication and Psychology, University of California, Los Angeles, Los Angeles, California, United States of America

Benedek Kurdi, Department of Psychology, Yale University, New Haven, Connecticut, United States of America

Ethan Ludwin-Peery, Department of Psychology, New York University, New York City, New York, United States of America

Dillon M. Luke, Department of Psychology, Univesity of Texas at Austin, Austin, Texas, United States of America

Keith B. Maddox, Department of Psychology, Tufts University, Medford, Massachusetts, United States of America

Daniel Marcelo, Oswald DigitalLisbon, Portugal

Arthur D. Marsden III, Department of Psychology, Syracuse University, Syracuse, New York, United States of America

Randy J. McCarthy, Department of Psychology, Northern Illinois University, DeKalb, Illinois, United States of America

David E. Melnikoff, The College of Science psychology program, Northeastern University, Boston, Massachusetts, United States of America

Gordon B. Moskowitz, Department of Psychology, Lehigh University, Bethlehem, Pennsylvania, United States of America

Leonard S. Newman, Department of Psychology, Syracuse University, Syracuse, New York, United States of America

Irmak Olcaysoy Okten, Department of Psychology, Florida State University, Tallahassee, Florida, United States of America

Diana Orghian, Feedzai, Lisbon, Portugal

Kimberly A. Quinn, Department of Psychology, DePaul University, Chicago, Illinois, United States of America

Tânia Ramos, Springer Nature Group – SN Digital, Lisbon, Portugal

Kate A. Ratliff, Department of Psychology, University of Florida, Gainesville, Florida, United States of America

Jessica D. Remedios, Department of Psychology, Tufts University, Medford, Massachusetts, United States of America

Diana T. Sanchez, Department of Psychology, Rutgers University, New Brunswick, New Jersey, United States of America

Emily Sands, Experimental Psychology, University College London (UCL), London, England, United Kingdom

Erica Schneid, TPX (Technology, Product, and Experience) Team, Comcast, Philadelphia, Pennsylvania, United States of America

Xi Shen, Department of Psychology, Cornell University, Ithaca, New York, United States of America

Jeffrey W. Sherman, Department of Psychology, University of California, Davis, Davis, California, United States of America

Jessica L. Shropshire, Cornell Engineering Leadership Program, Cornell University, Ithaca, New York, United States of America

John J. Skowronski, Department of Psychology, Northern Illinois University, Dekalb, Illinois, United States of America

Joel A. Thurston, Biocomplexity Institute and Initiative – Division of Social and Decision Analytics, University of Virginia, Charlottesville, Virginia, United States of America

Alexander Todorov, The University of Chicago Booth School of Business, The University of Chicago, Chicago, Illinois, United States of America

Yaacov Trope, Department of Psychology, New York University, New York, New York, United States of America

James S. Uleman, Department of Psychology, New York University, New York, New York, United States of America

Brandon M. Woo, Department of Psychology, Harvard University, Cambridge, Massachusetts, United States of America

Michael A. Zárate, Psychology Department, The University of Texas at El Paso, El Paso, Texas, United States of America

Preface: Impression Formation in Social Psychology

Gordon B. Moskowitz[1] *and Emily Balcetis*[2]

[1]*Department of Psychology, Lehigh University, Bethlehem, Pennsylvania, United States of America*

[2]*Department of Psychology, New York University, New York, New York, United States of America*

About two decades before the turn of the century, two rival universities in New York City held very different reputations. On the north end of town, Columbia University housed the legendary giants of social psychology, whose names filled textbooks, and whose work commanded large audiences. They were the famous ones whose names included Stanley Schachter, Walter Mischel, Bob Krauss, Mort Deutsch, and Richard Christie. On the south end of town, at New York University, sat a young set of professorial scholars generally in their 20s and 30s. Their work had yet to enter the scientific cannon, was not (yet) on required reading lists for post-graduate studies, and did not show up in introductory psychology lectures. Their names included John Bargh, Shelly Chaiken, Tory Higgins, Diane Ruble, Jeff Tanaka, and James Uleman, who were soon joined by Susan Andersen and Yaacov Trope.

Psychological scientists at the time categorized Columbia University's intellectual community as one respecting and venerating past contributions to the discipline. It followed the Ivy League model of collecting scholars of great acclaim whose revered contributions were rooted in methodologies of the past. To the extent that young scholars entered the ranks, they were to be recycled every six years. In contrast, New York University's community—or at least philosophy on how to develop a community—was a speculative one. NYU, as it is known, wagered on forthcoming productivity in a specific content area, and at the time this area was impression formation coupled with new and cutting-edge methodologies for exploring this topic. Its interests were vested in the potential of early career professors, some fresh from their doctoral hooding ceremonies. Interestingly, as history has shown, NYU's return on investment was a large one, as this fledging flock of intellectual entrepreneurs came to be an important intellectual force in the field of social psychology from the 1980s through today, leading and nurturing the development (and eventual dominance) of the discipline of social cognition.

Of course, impression formation, and cognition's role in it, were central aspects of social psychological inquiry prior to the 1980s. In the 1940s, Fritz Heider, Gustav Ichhesier, and Solomon Asch all began examining how perceivers form coherent impressions about others, and developing models

regarding how perceivers reason about the behaviors and inferred qualities they observe in others. In the 1950s, Gordon Allport extended this to the study of a particular type of impression—the stereotype. In the 1960s, scholars including Ned Jones and Hal Kelley developed formalized sets of rules that they believed perceivers follow when forming impressions. These rules included examining the effects of an action and of alternative possible actions to determine those effects that are not common across the alternatives, and the use of information concerning an action's consistency, consensus, and distinctiveness. In all these early approaches, however, the mechanisms that produce impressions and attributions were not the concern. A research participant would be asked to read about the qualities of another person, and then asked to explicitly report what they thought or felt about the person. Speculation about whether these qualities were averaged in some way, perhaps weighted by some "central" traits, or added together to produce the final impression was the extent to which there was a concern with the process of impression formation as opposed to the final product being self-reported. Hastorf et al. (1970) provided an excellent review of this early work. In the 1970s a separate approach revolted against this methodology of explicit measures, arguing that perceivers cannot accurately report what they think and feel or how they produce thoughts or feelings about the people they perceive. These scholars turned to using measures of implicit memory to explore processing strategies used by perceivers and more accurate measures of impressions that did not rely upon self-report. A subset called themselves the Person Memory Interest Group (Hastie et al., 1980), while others used the term implicit psychology (e.g., Wegner & Vallacher, 1977).

Our tale of two universities reflects this historical dynamic. To the north the revered scholars of Columbia with (the still valuable) tools of the past, and to the south, the NYU revolutionaries trying to change the discipline with implicit measures and a concern with processing mechanisms. Whereas the "Person Memory" scholars were a minority scattered around the United States, NYU took the novel approach of building an entire program of scholars with shared methodological approaches that were new and creative, who held complementary content interests relating to the topic of impression formation. A striking summary of their impact is seen in the edited volume *Unintended Thought* (Uleman & Bargh, 1989), with chapters from (among others) the entire NYU group. With cognitive processing mechanisms that drive impression formation as its main area of focus, NYUs advocacy and leadership for the emerging discipline of social cognition saw its influence spread far beyond the island of Manhattan to permeate psychological studies around the globe. It is hard to predict how the field would have evolved if left to isolated scholars scattered across the country. After all, the field had already grown enough to allow Fiske and Taylor (1984) and Wyer and Srull (1984) to produce important summary volumes in the early 1980s. Yet, it is our own subjective assessment that the creation of the NYU team helped to solidify, organize, and unify the discipline of social cognition.

This assessment is based on the fact that (no need to guess—it was Gordon) lived through this period at NYU in the 1980s, and saw every major figure pass through NYU for colloquium talks, with Manhattan itself being vital to drawing these scholars to visit NYU. He saw the enthusiasm that being in this special environment generated and the appeal of being in such a novel community of like-minded academics all of the same age cohort (so much so that he documented each visit with a "polaroid" photograph of each visitor that still hung on the walls at NYU for decades to follow). This like-minded study of impression formation drew together scholars from the traditionally isolated fields of attitudes and person perception. In the social cognition era, impression formation encompassed affective reactions or attitudes that social perceivers hold about others, as well as semantic inferences about people about such things as who they are and why they act the way they do. Lower-level, basic, and primary psychological processes such as attention, perception, and categorization contribute to the impressions perceivers form of others, while higher-level, emergent, and abstract processes such as attribution, intentionality, self-regulation, decision making, and stereotyping depend on having formed an impression of others. And that special environment, the team of scientists at New York University, was centered around the seminal and foundational contributions of James Uleman. This edited volume is a reflection on the 40-year history of research he inspired in social cognition on impression formation.

The Uleman-led group at NYU united disparate lines of research on memory, attention, attitudes, perception, categorization, and judgment. For the first time in empirical research, an entire scholarly unit was built to study how and why we form impressions of people. Initially, the field (including the NYU group) studied these issues in a vacuum, holding constant and ignoring motivational variables that might shape impression formation processes. Much like the sister discipline of cognitive psychology, social cognition saw goals as a variable to be controlled. This would allow processing mechanisms to be studied in their "pure" state. A quote from the chapter Uleman submitted for this volume summarizes that sentiment that pervaded this early history of social cognition: "spontaneous inferences seem not to be for doing anything; they simply occur unintentionally… They can be affected by the perceiver's goals, even unconsciously primed goals… however, they do not seem to be purposive, goal directed, or functional in any immediate sense… My interest is in *how* this occurs, mechanistically, not *why* it occurs, teleologically. Although top-down goals affect social inferences, theorists attributing motives or goals to cognition is both dangerous and slippery." Leadership from others in the NYU group helped to shift this sentiment in the field more broadly (if not in Uleman), and returned motivated reasoning processes to the center of social cognition (where factors such as interactions goals, hedonic relevance, and ego enhancement had pervaded the earlier work of Ned Jones; e.g., Jones & Davis, 1965; Jones & Thibaut, 1958). Prominent in guiding this shift was the *Handbook of Motivation and Cognition*

(Higgins & Sorrentino, 1986) and Chaiken's heuristic-systematic model of persuasion (e.g., Chaiken, 1987; Eagly & Chaiken, 1984).

If one desires to understand human social behavior, one must first be able to understand how each individual human construes the social world they inhabit at that moment. The snap judgments formed about interaction partners and the inferences perceivers make about the causes and reasons for the events people encounter are the basis for emotional reactions, predictions about what will happen next, and plans for appropriate behavioral responses to social contexts. This is to say, at the core of human experience is impression formation. It is this elemental nature of impression formation coupled with Uleman's retirement, that inspired the idea for collecting the chapters contained here. This volume does not aim to pay tribute to Uleman as the generous colleague, strong mentor, and prolific scholar. Instead, the volume looks back on the vast field of impression formation that his decades of scholarship and guidance helped to nurture. The volume is also forward looking, foreshadowing the next generation of methodological innovations and theoretical pivots that will offer answers to perplexing questions about social interactions. The goal is to review a broad discipline that emerged from a particular place and time, with the benefit of asking many of its founding figures and contemporary luminaries to explain how Uleman's classic work on spontaneous inference spawned their own empirical program of research, synthesize their current questions of interest and results, and presage the next investigations and innovations they expect to achieve. Finally, the volume uniquely allows one of those figures, James Uleman himself, to reflect on the state of the past and future state of the field.

As editors, our goal was to collect leading scholars who do research on a wide set of processes that relate to impression formation, and invite them to relate their work to an audience of people with an emerging interest in impression formation. Some of the chapters take the form of a personal narrative, allowing the reader to see the origins of inspiration, to see how a scholar thinks about questions over the arc of a career, and how specific events or findings lead to important turns and new discoveries. Other chapters were written as literature reviews that comprehensively detail the developments in an area over time, with relevant empirical examples to support the arguments. Finally, some chapters focus on new directions and contemplate where the field must go and what questions require focus for progress to be made.

As a guiding structure, we organized the volume around three themes. The first theme explores sources of input to impression formation processes, by probing the interaction of the person and the situation that is central to social psychology. These chapters explore how features of the surrounding world stimulate categorization. It probes processes like how perceivers' mental representations inform their perceptual experience. It explores important features of the environment that exert a strong influence on impression formation including nonverbal aspects of the people we perceive such as their facial expressions and facial features, as well as their

physiognomy, skin color, gait, posture, age, and gender. It also includes verbal features like language and prosody.

The second theme explores the process of forming impressions. It examines the mechanisms by which observations of human behavior produce inferences, categorizations, and eventually impressions. These chapters explore Bayesian reasoning, propositional and associative processing from which meaning is made, and the implicit nature of the inference process that allows for efficient and effective plans for action in response to the meaning that is implicitly made. Importantly, meaningful biases in impression formation, such as stereotyping, arise from such processing, and this section allows several authors to explore how social stereotyping can arise from basic processes of impression formation that were not explicitly intended to produce bias.

The third theme explores the flexibility of the impression formation process. Traditionally, scholars assumed that first impressions are difficult to change, but, simultaneously, that impressions are malleable. This appears at first blush to be a logical inconsistency. However, the chapters in this section offer theoretically derived, empirically supported resolutions to what only appears to be a paradox of ideas. Contributors to this section reflect on the ability for perceivers to update their first impressions, the ease with which impressions update, and the differences in updating affective compared to semantic impressions. As with the second section of the book, a focus is placed on stereotyping and prejudice as types of impressions that may change.

With this volume, we wish to inspire new generations of scholars to gain a deeper understanding of the impetus for the probative thoughts classic thinkers mused over, James Uleman among them, decades ago. We aim to offer traction for novel ideas to catch foot. And, after reading these chapters, we hope we have armed an ever-expanding next group of researchers with the intellectual tools they need to chart new theoretical territory and discover aspects of the human experience we have yet to even wonder about. Onward and upward.

References

Chaiken, S. (1987). The heuristic model of persuasion. In M. P. Zanna, J. M. Olson & C. P. Herman (Eds.), *Social Influence: The Ontario Symposium* (Vol. 5, pp. 3–39). Hillsdale, NJ: Erlbaum.

Eagly, A. H., & Chaiken, S. (1984). Cognitive theories of persuasion. *Advances in experimental social psychology*, *17*, 267–359.

Fiske, S. T., & Taylor, S. E. (1984). *Social cognition* (1st ed.). Reading, MA: Addison-Wesley.

Hastie, R,. Ostrom, T. M., Ebbesen, E. B., Wyer, R. S., Jr., Hamilton, D. L., & Carlston, D. E. (1980). *Person memory: The cognitive basis of social perception* (pp. 121–153). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hastorf, A. H., Schneider, D. J., & Polefka, J. (1970). *Person perception*. Reading, MA: Addison-Wesley.

Higgins, E. T., & Sorrentino, R. M. (1986). *Handbook of motivation and cognition: Foundations of social behavior*. New York: Guilford.

Jones, E. E., & Davis, K. E. (1965). From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 219–266). New York: Academic Press.

Jones, E. E., & Thibaut, J. W. (1958). Interaction goals as bases of inference in interpersonal perception. In L. Petrullo & R. Tagiuri (Eds.), *Person Perception and Interpersonal Behavior* (pp. 151–178). Stanford, CA: Stanford University Press.

Uleman, J. S., & Bargh, J. A. (1989). *Unintended thought*. New York: Guilford.

Wegner, D. M., & Vallacher, R. R. (1977). *Implicit psychology: An introduction to social cognition*. London: Oxford University Press.

Wyer, R. S., & Srull, T. K. (1984). *Handbook of social cognition*. Hillsdale, NJ: Erlbaum.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Part I

Source of Input to Impression Formation: When Features of the External Physical World Meet Internal Mental Representations

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

1 Social Categorizations as Decisions Made under Uncertainty

Grace S. R. Gillespie[1], *Jessica L. Shropshire*[2], *and Kerri L. Johnson*[3]

[1]*Department of Psychology, University of California, Los Angeles, Los Angeles, California, U.S.A.*
[2]*College of Engineering, Cornell University, Ithaca, New York, U.S.A.*
[3]*Department of Communication, University of California, Los Angeles, Los Angeles, California, U.S.A.*

Encountering unknown individuals is a routine part of daily life. From merely a glimpse or a sound, observers spontaneously and effortlessly distinguish others according to their social category memberships. Doing so is efficient and (according to many) inevitable. Indeed, the flexible and effortless nature of social perception is widespread, allowing observers to arrive at spontaneous inferences about others (Uleman et al., 2008; Uleman et al., 1996). Similarly, the speed and ease of the social categorization process is also flexible and spontaneous, leading many to conclude that both the process of social categorization and subsequent consequences is part of a highly dynamic process.

In this chapter, we begin by summarizing overarching characteristics of social perception by emphasizing how observers make social decisions about other people, generally, rather than on the visual systems that inform such decisions, specifically. Thus, we focus primarily on the process of social categorization: making decisions about the groups to which others belong. In doing so, we assert that while social categorization is systematic, it is highly dynamic and subject to change depending on factors that originate in both the target and the perceiver.

We then apply a judgment and decision-making framework to characterize social perceptions—the process by which individuals form impressions and make inferences about others, generally, and social categorizations—the process through which we group individuals based upon social information, specifically, as decisions made under uncertainty. As such, predictable heuristics that lead to systematic biases provide a framework to understand fundamental processes in social categorization. Moreover, we contend that just as with other decisions, various social categorizations are prone to be biased by utility concerns. Some utility concerns can be characterized as self-focused and will tend

DOI: 10.4324/9781003045687-2

to bias social categorizations in a way that favors personal safety and security. In contrast, some utility concerns can be construed as other-focused and will tend to bias social categorizations in a way that shields the target of perception from negative outcomes. We call our perspective the Heuristic Decision Model of Social Categorization, and we lay out the explanatory power of this approach in providing an overarching theoretical framework for understanding how common social perception mechanisms produce distinct patterns in social categorization biases. We end by highlighting extensions of this work to a broader array of social categorizations and their implications in daily life.

Social Categorization: Systematic but Not Static

Social categorization research has historically focused either on the perceptual mechanisms by which categorization occurs or on the consequences of categorizing others into groups. Rarely has research focused on both components simultaneously. Here we review the existing literature that integrates both the mechanisms and consequences of categorizations.

First, research shows that social categorizations are *dynamic and probabilistic*. Although early models characterized social categorizations as yielding discrete cognitive representations, current evidence suggests that these perceptions gradually accrue by integrating information from visible cues in the face and body (e.g., Johnson & Tassinary, 2005, 2007a, 2007b; Lippa, 1983; Pollick et al., 2005; Singh, 1993). For example, we, along with our colleagues, have shown that sex categorizations integrate available information that tends to culminate in a binary decision about others. Although the decisions themselves are categorical, they reflect a probabilistic assessment based on available, albeit imperfect, information. For instance, sexually dimorphic cues such as the shape of the body and face are highly diagnostic of sex categorization, yet they are not fully reliable because the distributions of men and women overlap. Additionally, both the number and clarity of cues available to observers varies dramatically depending on the context of perception, including lighting, occlusion, distance, and visual perspective (e.g., frontal or rear facing). Thus, although categorizations are often binary in nature and highly accurate, they are the product of a dynamic and continuous decision process that yields a probabilistic judgment based on the strength of the observed cues (Freeman et al., 2008). Consequently, gender-typical faces tend to compel more efficient and fluent categorizations than gender-atypical faces even when they culminated in identical social decisions. Moreover, behavioral evidence from mouse tracking and reaction time paradigms confirms that visual cues dynamically informed observers' judgments (see also, Freeman & Ambady, 2009, 2011; Freeman, Pauker, et al., 2010).

Second, social categorizations are prone *to systematic biases* that stem from the incidental perception of orthogonal social information, even when such information is not directly pertinent for the decision being made. For

example, perceiving a target's sex biases judgments of attractiveness and sexual orientation because it contextualizes perceptions of gendered cues (Freeman, Johnson, et al., 2010; Johnson & Ghavami, 2011; Johnson et al., 2007; Johnson & Tassinary, 2007a, 2007b). The reverse is also true—cues to sexual orientation bias sex categorizations (Lick, Johnson, et al., 2013). Finally, face and body cues that convey racial information also systematically bias sex categorizations (Johnson, Freeman, et al., 2012; Lick, Johnson, et al., 2013), and vice versa (Carpinella et al., 2015), due to overlaps in facial cues and stereotypes (e.g., Asian/female and Black/male).

Third, while social categorizations are systematic, *features of the target and the perceiver* impact categorization. For example, in addition to the mutual influence of orthogonal social categories described above, some features of the targets of social categorization tend to reinforce and facilitate one another because of their routine co-occurrence. For example, hairstyles and emotions facilitate specific race categorizations (Hugenberg & Bodenhausen, 2003, 2004; Hutchings & Haddock, 2008; MacLin & Malpass, 2001, 2003). Furthermore, facial cues to race bias basic color perceptions to align with the expectations of the perceiver (Levin & Banaji, 2006). Additionally, assigning race labels to targets with race-ambiguous faces influence the depth of perceivers' processing (Corneille et al., 2004; Michel et al., 2007, 2010) and perceivers' accuracy of recognition (Pauker & Ambady, 2009; Pauker et al., 2009). Finally, the motivational states of the perceiver (e.g., self-protection) change their perceptual thresholds for making sex and race judgments (Johnson, Iida, et al., 2012; Miller et al., 2010).

Fourth and finally, social categorizations are *consequential*. Some research, for example, has established that social categorization is sufficient to elicit stereotypes that inform impressions, elicit attitudes, and shape interpersonal behaviors (Allport, 1954; Bargh, 1999; Brewer, 1988; Devine, 1989; Dovidio et al., 1986; Fazio & Dunton, 1997; Gilbert & Hixon, 1991; Grant & Holmes, 1981). Once social categorization occurs, the application of social categories activates knowledge structures that alter social evaluations (see e.g., Bodenhausen & Peery, 2009; Brewer, 1988; Fiske & Neuberg, 1990; Lick & Johnson, 2013). The consequences of categorization include physical embodiment of stereotyped characteristics (Bargh, 1999) and even mental and physical health disparities (Lick, Durso, et al., 2013).

Collectively, these findings characterize social categorization as a dynamic decision process that is prone to systematic biases. The dynamic nature of social categorization means that features of the social environment are likely to have an impact. More specifically, social categorizations happen in an uncertain environment that gives way for features of those being categorized (the target) and those doing the categorizing (the perceiver) to play a role in decisions that are made.

A Judgment and Decision-Making Framework for Social Categorizations

Decisions, in general, reflect an assessment of probability, corrected for biases that stem from the perceived utility (Gilovich & Dale, 2002). In existing models of decision making, utility is considered to reflect self-relevant consequences. When negative consequences for the self are extreme, decisions exhibit a conservative bias that aims to mitigate potential costs (e.g., loss aversion); when consequences for the self are minimal, they do not (Kahneman & Tversky, 1984). These tendencies vary with the magnitude of the consequences, producing some judgments that occur rapidly and others that are more contemplative.

Perceived utility impacts a range of judgments, showing distinct outcome patterns when the consequences are self- versus other-relevant. This distinction is crucial; self-relevant decisions foster self-protective and congenial biases. When estimating the likelihood of engaging in philanthropic activities, for example, people provide favorable but biased estimates for themselves, but accurate and unbiased estimates for others, in part because they consult different evidence to inform their predictions (Epley & Dunning, 2000). When considering whether they exhibit a desirable personality characteristic, people tend to define the characteristic in self-serving ways. Doing so allows them to privilege confirmatory evidence but ignore disconfirming evidence (Dunning et al., 1989). When deciding whether (or not) a negative stereotype applies to one's own ingroup, people engage in more logical reasoning that allows them to refute the possibility, relative to when they render similar judgments for others (Dawson et al., 2002). Finally, the same criteria that people use to assess the probability of an event are more stringent when considering non-preferred, relative to preferred outcomes (Ditto & Lopez, 1992; Ditto et al., 1998). Thus, self-relevant decisions engage decision strategies that protect the self and promote self-regard.

As described previously, social categorizations are probabilistic insofar as they rely on cues that are imperfectly diagnostic and perceived under variable conditions. Put another way, social categorizations are akin to social decisions that are made under some degree of uncertainty. Characterizing social categorizations in this way is consistent with the extant literature in social perception, and it also provides a unique perspective that provides novel predictions about the determinants and consequences of social categorization biases. Just as they impact other decisions made under uncertainty, utility concerns are likely to correspond to distinct social categorization biases. We propose that when the self-relevant utility associated with a decision is predominant, social categorizations will be *conservative and quick*; when other-relevant utility concerns are paramount, in contrast, social categorizations will be more *cautious and contemplative*.

Social Categorizations as Decisions under Uncertainty

While decision patterns associated with social categorizations are highly consistent *within* a specific judgment domain (e.g., *within* studies of race or

within studies of sexual orientation), they tend to be quite distinct when considered *between* judgment domains. Thus, whereas some social categorizations are made even for scant visual evidence, others require overwhelming visual evidence. For instance, race categorizations tend to occur readily and rapidly even when the visual evidence to support them is minimal, particularly when they involve minority "Black" categorizations (e.g., Peery & Bodenhausen, 2008). In contrast, categorizations of other minority and stigmatized groups show the opposite pattern. Sexual orientation categorizations, for example, demand a preponderance of visual evidence to support a "gay" categorization (e.g., Johnson et al., 2007). To date, such between-category discrepancies in social categorization biases have received minimal attention. Yet, these distinctions also provide insights into the motivations underlying social categorizations and the biases that accompany them. As such, rather than reflecting distinct social categorization mechanisms, we instead focus on how construing social categorizations as decisions made under uncertainty provides an integrative framework for understanding patterns of social categorizations. Here we develop the **Heuristic Decision Model**, characterizing social categorizations as decisions that are made under varying degrees of uncertainty, ideally providing a theoretical basis for understanding these distinctions and providing a foundation for making novel predictions.

Based on visual cues alone, observers can achieve varying degrees of accuracy for judgments of social categories that range from those described as perceptually obvious (e.g., sex, race, and age) to those historically considered to be perceptually opaque, often referred to as concealable (e.g., sexual orientation, political party affiliation, religious ideology). Not surprisingly, the accuracy of categorizations varies from nearly perfect (e.g., sex) to merely above chance (e.g., religious affiliation).

Despite a high degree of accuracy, social categorizations are also prone to systematic biases, and unique biases accompany specific judgments. Considering these decisions using a signal detection framework, both accuracy and bias can occur via different means. Accurate judgments, not surprisingly, correspond to correct categorizations involving both *hits* and *correct rejections*; erroneous judgments correspond to *misses* and *false alarms*. Interestingly, distinct social categorizations appear to differently privilege each type of accuracy and error, leading to systematic differences in bias. For instance, some categorizations tend to be more *conservative* (i.e., requiring a low threshold of evidence on criterion measures in signal detection) and *quick* (i.e., made more rapidly than the category alternative). In such instances, decision biases tend to favor the category that is stereotypically aligned with threat. Race and sex categorizations tend to follow this pattern. Specifically, race-ambiguous faces tend to be more readily categorized as Black (Freeman, Pauker, et al., 2010; Ho et al., 2011; Peery & Bodenhausen, 2008), and Black categorizations are made more rapidly than other race categorizations, particularly in predominantly White observers (Stroessner, 1996; Zárate & Smith, 1990).

Similarly, both bodies and faces are more readily and rapidly categorized as male instead of female by both male and female perceivers (e.g., Johnson, Freeman, et al., 2012; Johnson, Iida, et al., 2012; Johnson & Ghavami, 2011; Lick, Johnson, et al., 2013). Thus, observers appear to eagerly exploit information that a target might be Black or male. This occurs despite the fact that the two domains of judgment differ dramatically in their base rates in society.

Other categorizations, in contrast, tend to be more *cautious* (i.e., requiring a high threshold of evidence on criterion measures in signal detection) and *contemplative* (i.e., made less rapidly and more deliberatively than the alternative). For instance, sexual orientation judgments tend to achieve above-chance levels of accuracy, but they also show a sizable bias for "straight" categorizations (e.g., Freeman, Johnson, et al., 2010; Johnson et al., 2007; Johnson & Ghavami, 2011; Lick, Johnson, et al., 2013). Religious categorizations reveal a similar pattern—observers accurately categorize religious identity but are biased toward non-minority judgments (Rule et al., 2010). Thus, observers appear reluctant to utilize information that a target might be gay or belong to a religious minority. Of course, in these instances, the base rates are quite low, yet this does not seem to account for the pattern of results. Importantly, these patterns persist even when base rates are made explicit to observers (Lick & Johnson, 2016). Moreover, they exhibit a pattern that distinguishes them from other decision contexts in which the target is part of a numerical minority (e.g., Black categorizations).

Thus, different domains of judgment show distinct decision biases. Although each of the judgments described thus far can achieve sensitivity at above chance levels, they require relatively different thresholds of information in criterion measures, reflecting distinct biases in observers' decision strategies. Some categorizations appear to follow a "conservative and quick" decision pattern in which minimal visual information is sufficient to render a categorization; others reflect a more "cautious and contemplative" decision pattern in which abundant visual information is required to compel a categorization. Why might this occur? Decisions made under uncertainty, including social categorizations, involve trade-offs, and these are informative for understanding distinctions between types of social categorizations. Decisions that require a relatively low threshold of evidence are also likely to tolerate a relatively higher false alarm rate; decisions that demand a high threshold of evidence, in contrast, are also likely to allow more misses.

Several factors might determine which patterns will occur for social categorizations. For instance, stereotype valence fosters distinct decision strategies, with negatively stereotyped groups inviting more conservative/quick decisions but positively stereotyped groups inviting more cautious/contemplative decisions (e.g., Baumeister et al., 2001). This possibility falls short, however, because it cannot account for distinct patterns across groups with negative stereotypes.

Another possibility is that the relative diagnosticity of visual cues determines the decision patterns, with categorizations that are informed by

more diagnostic visual cues being associated with conservative/quick judgments but those informed by less diagnostic cues being more associated with cautious/contemplative judgments. This possibility also fails to account for existing patterns of data. Indeed, both anger-proneness and chronic disease correspond to few diagnostic visual cues, yet categorization of these characteristics reliably exhibits a conservative/quick rather than a cautious/contemplative decision strategy (Galperin et al., 2013; Holbrook et al., 2014; Schaller & Neuberg, 2012; Schaller & Park, 2011).

A third possibility is that decision biases track the relative population base rates such that low prevalence categories inspire more cautious/contemplative judgments. Once again, this seems unlikely insofar both the categories "gay" and "Black" are in the minority, yet such categorizations exhibit different biases. Thus, existing patterns of bias in social categorizations are not easily accounted for by stereotype valence, cue diagnosticity, or base-rate explanations.

A final possibility, and the basis of the current chapter, is that as decisions that are made under varying degrees of uncertainty, social categorization patterns shift as a function of utility concerns for the perceiver. This possibility provides clarity for predicting whether categorizations will be biased to be conservative/quick or cautious/contemplative, and it provides a single theoretical framework for understanding the relative shifts across different social categorizations.

Self-Relevant Utility Concerns Compel Conservative and Quick Categorizations

As described previously, self-relevant utility concerns, in general, tend to privilege information in a way that protects the decision maker. In the same way that a smoke detector is programmed to sound its alarm based on minimal cues to threat (even when doing so produces false alarms), decisions that protect oneself are similarly prone to require a low threshold of evidence. The possibility that self-relevant utility concerns bias social categorizations is consistent with the decision literature and enjoys a theoretical precedent in ecological approaches to social perception (e.g., McArthur & Baron, 1983; Neuberg & Sng, 2013; Schaller & Neuberg, 2012; Zebrowitz & Collins, 1997). In this work, interpersonal affordances are akin to assessments of self-relevant utility insofar as they calculate the risks of interpersonal contact. When potential risks of contact are high, perceivers tend to "err on the side of caution," adopting a risk-averse decision strategy that favors false-positive over false-negative categorizations (e.g., Johnson et al., 2013). Work from other domains shows a similar pattern, when resources are scarce, observers visual processing of racial outgroups is disrupted (Krosch & Amodio, 2019), which has the potential to exacerbate discrimination of racial outgroups.

Existing evidence supports the possibility that these decision patterns reflect self-relevant utility concerns. Sex and race categorizations are biased toward "male" and "Black," in part because members of these categories are

perceived to be physically formidable. Consequently, male categorizations occur more readily and rapidly than female categorizations (Johnson, Iida, et al., 2012). Threats to personal safety (e.g., a target is approaching rapidly or is Black) amplify these tendencies (Johnson, Iida, et al., 2012; Johnson, Freeman, et al., 2012). Similarly, Black categorizations also occur readily and rapidly, even when visual evidence is minimal (Peery & Bodenhausen, 2008; Ho et al., 2011). Once again, personal safety concerns exacerbate this bias (e.g., fearful perceivers, Miller et al., 2010; male targets, Carpinella & Johnson, 2013; or angry targets, Hugenberg & Bodenhausen, 2004). Finally, anger is more readily detected in the faces of young Black men (Hutchings & Haddock, 2008; Kang & Chasteen, 2009). Thus, self-relevant utility concerns bias social categorizations in a self-protective manner.

Other-Relevant Utility Concerns Compel Cautious and Slow Categorizations

Importantly, not all social categorizations are likely to reflect self-relevant utility concerns. Instead, some social categorizations have decidedly greater consequences for the target of perception than for the self. When might this occur? Instances in which a decision about another person might expose them to stigma are likely to arouse other-relevant utility concerns, at least among many perceivers. Current models of social perception have yet to fully consider the impact of these considerations on social categorizations, yet some evidence suggests that other-relevant concerns might be an important factor in social decision biases. For instance, spontaneous trait inferences tend to be positively biased when observers held affiliation goals (Rim et al., 2013), indicating that a prosocial or other-focused orientation is likely to favor more benevolent percepts of others. Additionally, when people hold more other-focused political affiliations, they are reluctant to utilize gendered cues for making sexual orientation judgments, in spite of their diagnosticity in the research context (Stern et al., 2013).

We propose that other-relevant utility concerns bias social categorizations, more specifically, to be cautious and contemplative, especially when these categorizations produce negative consequences for targets of perception. Specifically, when other-relevant utility concerns are pronounced, perceivers' categorizations tend to favor non-stigmatized judgments. From this perspective, other-relevant utility concerns might be sufficient to inspire benevolent motivations that bias social categorizations to be cautious and contemplative.

Although few current models of social perception account for other-focused motivations, existing evidence supports this prediction. Indeed, categorizing someone as a sexual or religious minority, for example, exposes the target of perception, but not the perceiver, to interpersonal animus and stigma (see, e.g., Lick, Durso, et al., 2013). Interestingly, sexual orientation judgments, although generally accurate, show a systematic bias for "straight" categorizations (e.g., Johnson et al., 2007; Lick, Johnson, et al., 2013). Similarly, judgments of

religious groups also show biases toward non-minority categorization (Quanty et al., 1975). Factors that moderate these biases also vary in their concern for oneself or others. Political conservatives, for example, exhibit *less* of a straight-categorization bias than liberals (Stern et al., 2013), and anti-Semitic observers also show less bias (Quanty et al., 1975). Thus, other-relevant utility concerns bias social decisions in a benevolent manner.

The Heuristic Decision Model of Social Categorization

In light of existing evidence, we characterize social categorizations as probabilistic decisions that are biased by perceived self- and other-relevant utilities consistent with prior evidence (e.g., Freeman et al., 2008), visual cues inform preliminary perceptions that differentially arouse self- and other-relevant utility concerns. Self-relevant concerns allow categorizations to occur at a low perceptual threshold (i.e., favoring self-protective false-positives); other-relevant concerns, in contrast, permit categorizations only at a high perceptual threshold (i.e., favoring benevolent false-negatives). This model has extensive implications for social categorizations, social reasoning, and evaluations. (Figure 1.1).

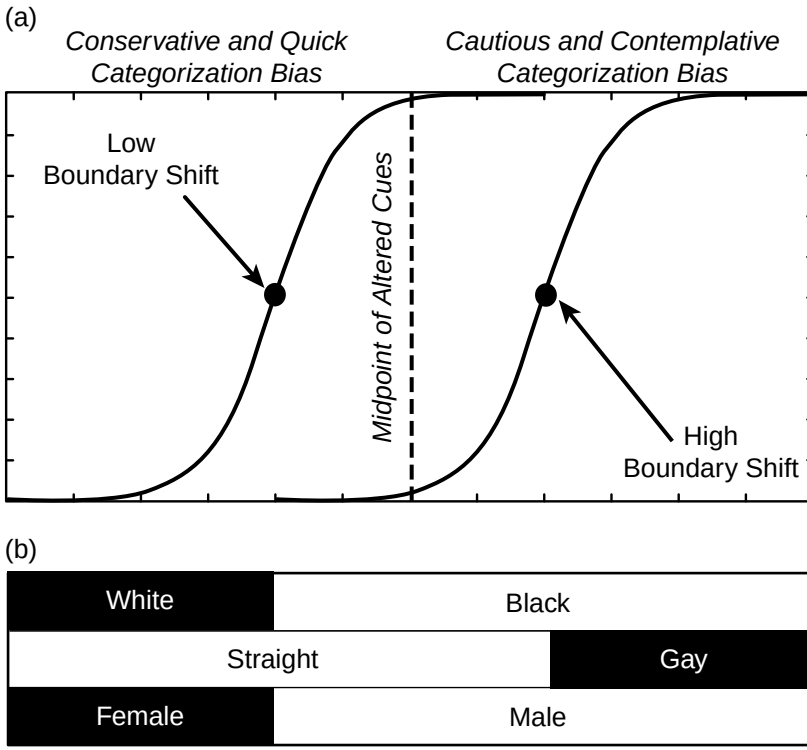

Figure 1.1 The Heuristic Decision Model of Social Categorization.

Implications for Categorization Biases

This model implies that decision biases will vary between social categorizations. When one's initial perception specifies a group that arouses self-relevant concerns (e.g., Black or male), categorizations are likely to occur at a *low* perceptual threshold and exhibit a conservative/quick decision strategy. In contrast, when one's initial perception specifies a group that arouses other-relevant concerns (e.g., gay), categorizations are likely to occur reluctantly at a *high* perceptual threshold and exhibit a cautious/contemplative strategy. In recent work (Alt et al., 2020), we found evidence consistent with this possibility. When categorizing the sexual orientation of faces that varied continuously in gendered appearance, observers' judgments showed a strong and consistent bias that favored the straight category alternative and shifted from gay to straight categorizations further along the gendered continuum.

It is important to note that although the existing social categorization literature is consistent with the theoretical model proposed herein, there remains much to be done. Our own lab's work has focused heavily on categorizations of sexual orientation and sex. Doing so has provided numerous key insights that inform the current perspective, yet there remain many exciting avenues to pursue. For instance, a broader interrogation of the more spontaneous and intentional aspects of early impression formation is warranted, particularly given its impact on trait inferences of others (Ferreira et al., 2012; Uleman et al., 2012).

Implications for Social Reasoning

This perspective implies that social reasoning—drawing inferences about others' intentions, dispositions, and actions—will vary with utility concerns. When reasoning about another person, perceivers integrate all available information, including stereotypes (e.g., Srull & Wyer, 1989), and they correct for these generalizations only when the motivation for accuracy is strong (Clary & Tesser, 1983; Hastie, 1984). Consequently, utility concerns govern whether observers seek additional information that could correct initial impressions. Specifically, when categorizations arouse self-relevant concerns (e.g., Black or male), observers are prone to readily accept initial stereotyped impressions. In contrast, when categorizations inspire other-relevant concerns (e.g., gay), observers tend to reconsider their initial stereotyped impressions and seek potentially exculpatory evidence. For instance, Stroessner et al. (2015) found that when perceivers are motivated to prevent harm to themselves, they willingly apply stereotypes that curtail others' civil liberties. Additionally, people scrutinize putative evidence for the truth of stereotypes more thoughtfully when it violates their attitudes than when it corroborates them (Munro & Ditto, 1997). In our work, we have found that observers readily attribute anger-prone dispositions to a person holding an

incidentally dangerous object (e.g., garden shears), but not to a person holding an innocuous object (e.g., a watering can; Holbrook et al., 2014). Additionally, we have also found that other-relevant focus (whether measured or manipulated) is associated with a stronger straight-categorization bias for judgments of sexual orientation that also allowed observers to avoid confirmation biases in decision tasks (Alt et al., 2020).

Implications for Evaluative Judgments

Finally, this perspective implies that the relation between social categorization processes and social evaluation varies as a function of utility concerns. In general, it has been assumed that the ease with which a target is perceived determines how favorably it is evaluated (Lick & Johnson, 2013; Winkielman et al., 2006; Winkielman et al., 2002). In contrast, we have found that the perceptual fluency associated with conservative/quick versus cautious/contemplative social categorizations produces distinct social evaluation tendencies. Specifically, fluency can lead perceivers to embrace the valence of an existing stereotype in their social evaluations for categories that arouse self-relevant concerns. In contrast, fluency can foster more thoughtful and deliberative processing for categories that inspire other-relevant concerns, allowing perceivers to overcome stereotype valence in evaluations.

Although evaluative judgments tend to be more favorable for objects and people who are more easily or fluently processed (Winkielman et al., 2002), this is not always the case. For instance, faces that exhibit more prototypical Black features compel fluent race categorizations, yet also tend to elicit negative evaluations (Maddox, 2004); faces and bodies that exhibit more prototypical gendered features, in contrast, compel fluent sex categorizations and also tend to elicit favorable evaluations (Johnson & Tassinary, 2007a, 2007b; Lick & Johnson, 2013). The current model therefore predicts that the relation between fluent processing of social category membership and evaluative judgments will be attenuated or even reversed for categories that arouse self-relevant concerns but strengthened for categories that arouse other-relevant concerns, over and above the stigma associated with a category. Testing this possibility entails comparing the relation between perceptual fluency and social evaluations between social categories and then relating these patterns to self- and other-relevant utility concerns.

Some evidence supports this characterization. For instance, when another person's sexual orientation is unclear, the uncertainty compromises observers' cognitive functioning as they strive to resolve the ambiguity (Everly et al., 2012). This finding suggests that perceptual uncertainty is cognitively taxing, carrying the potential to impact evaluations. In our research, we have related fluency to evaluations (Lick & Johnson, 2013). We defined fluency as the latency for participants to evaluate either gender- or racial-typicality and then related these measures to social evaluations. This approach revealed

that more fluent gender judgments corresponded to more favorable global evaluations, largely explaining differences in evaluations between gay and straight targets. In contrast, the latency of race typicality judgments was unrelated to differences in evaluations between Black and White targets.

Conclusion

In this chapter, we provided a novel framework for understanding biases in social categorizations. The Heuristic Decision Model of Social Categorizations situates social categorizations within a broader decision-making framework in an effort to reconcile seemingly contradictory patterns of bias that occur across domains. When self-relevant concerns, such as safety, are paramount, social categorizations tend to be conservative and quick. Such categorizations include both sex, which is biased toward male judgments, and race, which is biased toward Black judgments. In contrast, when other-relevant concerns are prioritized, social categorizations are prone to be more cautious and contemplative. Such categorizations include, but are not likely limited to, sexual orientation, which is biased toward straight judgments. We believe that this model holds promise for both reconciling seemingly inconsistent patterns of bias in the extant literature and generating novel predictions moving forward.

References

Allport, G. W. (1954). *The Nature of Prejudice*. Reading, MA: Addison Wesley.

Alt, N. P., Lick, D. J., & Johnson, K. L. (2020). The straight categorization bias: A motivated and altruistic reasoning account. *Journal of Personality and Social Psychology*, *119*, 1266–1289.

Bargh, J. A. (1999). The cognitive monster: The case against controllability of automatic stereotype effects. In S. Chaiken & Y. Trope (Eds.), *Dual Process Theories in Social Psychology* (pp. 361–382). New York: Guilford.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*, 323–370.

Bodenhausen, G., & Peery, D. (2009). Social categorization and stereotyping in vivo: The VUCA challenge. *Social and Personality Psychology Compass*, *3*, 133–151.

Brewer, M. B. (1988). A dual process model of impression formation. In R. S. Wyer, Jr. & T. K. Srull (Eds.), *Advances in Social Cognition* (Vol. 1, pp. 1–36). Hillsdale, NJ: Erlbaum.

Carpinella, C., & Johnson, K. L. (2013). Appearance-based politics: Sex-typed facial cues communicate political party affiliation. *Journal of Experimental Social Psychology*, *49*, 156–160.

Carpinella, C. M., Chen, J., Hamilton, D., & Johnson, K. L. (2015). Gendered facial cues influence race categorizations. *Personality and Social Psychology Bulletin*, *41*(3), 405–419.

Clary, E. G., & Tesser, A. (1983). Reactions to unexpected events: The naive scientist and interpretive activity. *Personality and Social Psychology Bulletin*, *9*, 609–620.

Corneille, O., Huart, J., Becquart, E., & Brédart, S. (2004). When memory shifts towards more typical category exemplars: Accentuation effects in the recollection of ethnically ambiguous faces. *Journal of Personality and Social Psychology*, *41*, 431–437.

Dawson, E., Gilovich, T., & Regan, D. T. (2002). Motivated reasoning and performance on the Wason selection task. *Personality and Social Psychology Bulletin*, *28*, 1379–1387.

Devine, P. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*, 5–18.

Ditto, P. H., & Lopez, D. F. (1992). Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology*, *63*, 568–584.

Ditto, P. H., Scepansky, J. A., Munro, G. D., Apanovich, A. M., & Lockhart, L. K. (1998). Motivated sensitivity to preference-inconsistent information. *Journal of Personality and Social Psychology*, *75*, 53–69.

Dovidio, J. F., Evans, N., & Tyler, R. B. (1986). Racial stereotypes: The contents of their cognitive representations. *Journal of Experimental Social Psychology*, *22*, 22–37.

Dunning, D., Meyerowitz, J. A., & Holzberg, A. D. (1989). Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of Personality and Social Psychology*, *57*, 1082–1090.

Epley, N., & Dunning, D. (2000). Feeling "holier than thou": Are self-serving assessments produced by error in self or social prediction? *Journal of Personality and Social Psychology*, *79*, 861–875.

Everly, B., Shih, M. J., & Ho, G. C. (2012). Don't ask, don't tell? Does disclosure of gay identity affect partner performance? *Journal of Experimental Social Psychology*, *48*, 407–410.

Fazio, R. H., & Dunton, B. C. (1997). Categorization by race: The impact of automatic and controlled components of racial prejudice. *Journal of Experimental Social Psychology*, *33*, 451–470.

Ferreira, M. B., Garcia-Marques, L., Hamilton, D., Ramos, T., Uleman, J. S., & Jerónimo, R. (2012). On the relation between spontaneous trait inferences and intentional inferences: An inference monitoring hypothesis. *Journal of Experimental Social Psychology*, *48*, 1–12.

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 23, pp. 1–74). New York, Academic Press.

Freeman, J. B., & Ambady, N. (2009). Motions of the hand expose the partial and parallel activation of stereotypes. *Journal of Experimental Social Psychology*, *46*, 179–185.

Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review*, *118*, 247–279.

Freeman, J. B., Ambady, N., Rule, N. O., & Johnson, K. L. (2008). Will a category cue attract you? Motor output reveals dynamic competition across person construal. *Journal of Experimental Psychology: General*, *137*, 673–690.

Freeman, J. B., Johnson, K. L., Ambady, N., & Rule, N. O. (2010). Sexual orientation perception involves gendered facial cues. *Personality and Social Psychology Bulletin*, *36*, 1318–1331.

Freeman, J., Pauker, K., Apfelbaum, E., & Ambady, N. (2010). Continuous dynamics in the real-time perception of race. *Journal of Experimental Social Psychology*, *46*, 179–185.

Galperin, A., Fessler, D. M. T., Johnson, K. L., & Haselton, M. G. (2013). Seeing storms behind the clouds: Biases in the attribution of anger. *Evolution and Human Behavior*, *34*, 358–365.

Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs *Journal of Personality and Social Psychology*, *60*, 509–517.

Gilovich, T. D., & Dale, D. (2002). Heuristics and biases: Then and now. In T. Gilovich, D. Griffin & D. Kahneman, (Eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment* (pp. 1–18). New York: Cambridge University Press.

Grant, P. R., & Holmes, J. G. (1981). The integration of implicit personality theory schemas and stereotype images. *Social Psychology Quarterly*, *44*, 107–115.

Hastie, R. (1984). Causes and effects of causal attribution. *Journal of Personality and Social Psychology*, *46*, 44–56.

Ho, A. K., Sidanius, J., Levin, D. T., & Banaji, M. R. (2011). Evidence for hypodescent and racial hierarchy in the categorization and perception of biracial individuals. *Journal of Personality and Social Psychology*, *100*, 492–506.

Holbrook, C., Galperin, A., Fessler, D. M. T., Johnson, K. L., Bryant, G. A., & Haselton, M. (2014). If looks could kill: Anger attributions are intensified by affordances for doing harm. *Emotion*, *14*(3), 455–461.

Hugenberg, K., & Bodenhausen, G. V. (2003). Facing prejudice: Implicit prejudice and the perception of facial threat. *Psychological Science*, *15*, 640–643.

Hugenberg, K., & Bodenhausen, G. V. (2004). Ambiguity in social categorization: The role of prejudice and facial affect in race categorization. *Psychological Science*, *15*, 342–345.

Hutchings, P. B., & Haddock, G. (2008). Look Black in anger: The role of implicit prejudice in the categorization and perceived emotional intensity of racially ambiguous faces. *Journal of Experimental Social Psychology*, *44*, 1418–1420.

Johnson, D. D. P., Blumstein, D. T., Fowler, J. H., & Haselton, M. G. (2013). The evolution of error: Error management, cognitive constraints, and adaptive decision-making biases. *Trends in Ecology & Evolution*, *28*, 474–481.

Johnson, K. L., Freeman, J. B., & Pauker, K. (2012). Race is gendered: How covarying phenotypes and stereotypes bias sex categorization. *Journal of Personality and Social Psychology*, *102*, 116–131.

Johnson, K. L., & Ghavami, N. (2011). At the crossroads of conspicuous and concealable: What race categories communicate about sexual orientation. *PLoS One*, *6*, e18025, doi: 10.1371/journalpone.0018025.

Johnson, K. L., Gill, S., Reichman, V., & Tassinary, L. G. (2007). Swagger, sway, and sexuality: Perceiving sexual orientation from the body's shape and motion. *Journal of Personality and Social Psychology*, *93*, 321–334.

Johnson, K. L., Iida, M., & Tassinary, L. G. (2012). Person (mis)perception: Functionally biased sex categorization of bodies. *Proceedings of the Royal Society, Biological Sciences*, *279*, 4982–4989.

Johnson, K. L., & Tassinary, L. G. (2005). Perceiving sex directly and indirectly: Meaning in motion and morphology. *Psychological Science*, *3*, 890–897.

Johnson, K. L., & Tassinary, L. G. (2007a). Compatibility of basic social perceptions determines perceived attractiveness. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 5246–5251.

Johnson, K. L., & Tassinary, L. G. (2007b). The functional significance of the WHR in judgments of attractiveness. In V. Swami & A. Furnham (Eds.), *Body Beautiful: Evolutionary and Socio-cultural Perspectives* (pp. 159–184). New York: Palgrave Macmillian.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, *39*, 341–350.

Kang, S. K., & Chasteen, A. L. (2009). Beyond the double-jeopardy hypothesis: Assessing emotion in the faces of multipy-categorizable targets of prejudice. *Journal of Experimental Social Psychology*, *45*, 1281–1285.

Krosch, A. R., & Amodio, D. M. (2019). Scarcity disrupts the neural encoding of black faces: A socioperceptual pathway to discrimination. *Journal of Personality and Social Psychology*, *117*, 859–875.

Levin, D. T., & Banaji, M. R. (2006). Distortions in the perceived lightness of faces: The role of race categories. *Journal of Experimental Psychology: General*, *135*, 501–512.

Lick, D. J., Durso, L. E., & Johnson, K. L. (2013). Minority stress and physical health in sexual minority communities. *Perspectives on Psychological Science*, *8*, 521–548.

Lick, D. J., & Johnson, K. L. (2013). Fluency of visual processing explains prejudiced evaluations following categorization of concealable stigmas. *Journal of Experimental Social Psychology*, *49*, 419–425.

Lick, D. J., & Johnson, K. L. (2016). Straight until proven gay: A systematic bias toward straight categorizations in sexual orientation judgments *Journal of Personality and Social Psychology*, *110*, 801–817.

Lick, D. J., Johnson, K. L., & Gill, S. V. (2013). Deliberate changes to gendered body motion influence basic social perceptions. *Social Cognition*, *31*, 656–671.

Lippa, R. (1983). Sex typing and the perception of body outlines. *Journal of Personality*, *51*, 667–682.

MacLin, O. H., & Malpass, R. S. (2001). Racial categorization of faces: The ambiguous race face effect. *Psychology, Public Policy, and Law*, *7*, 98–118.

MacLin, O. H., & Malpass, R. S. (2003). The ambiguous race face illusion. *Perception*, *32*, 249–252.

Maddox, K. B. (2004). Perspectives on racial phenotypicality bias. *Personality and Social Psychology Review*, *8*, 383–401.

McArthur, L. Z., & Baron, R. M. (1983) Toward an ecological theory of social perception. *Psychological Review*, *90*, 215–238.

Michel, C., Corneille, O., & Rossion, B. (2007). Race categorization modulates holistic face encoding. *Cognitive Science*, *31*, 911–924.

Michel, C., Corneille, O., & Rossion, B. (2010). Holistic face encoding is modulated by perceived face race: Evidence from perceptual adaptation. *Visual Cognition*, *18*, 434–455.

Miller, S. L., Maner, J. K., & Becker, D. V. (2010). Self-protective biases in group categorization: Threat cues shape the psychological boundary between "us" and "them.". *Journal of Personality and Social Psychology*, *99*, 62–77.

Munro, G. D., & Ditto, P. H. (1997). Biased assimilation, attitude polarization, and affect in the processing of stereotype-relevant scientific information. *Personality and Social Psychology Bulletin*, *23*, 636–653.

Neuberg, S. L., & Sng, O. (2013). A life history theory of social perception: Stereotyping at the intersections of age, sex, ecology, (and race). *Social Cognition*, *31*, 696–711.

Pauker, K. B., & Ambady, N. (2009). Multiracial faces: The boundaries of race. *Journal of Social Issues*, *65*, 69–86.

Pauker, K., Weisbuch, M., Ambady, N., Sommers, S. R., Adams, R. B., Jr., & Ivcevic, Z. (2009). Not so black and white: Memory for ambiguous group members. *Journal of Personality and Social Psychology*, *96*, 795–810.

Peery, D., & Bodenhausen, G. V. (2008). Black + white = black: Hypodescent in reflexive categorization of racially ambiguous face. *Psychological Science*, *19*, 973–977.

Pollick, F. E., Kay, J. W., Heim, K., & Stringer, R. (2005). Gender recognition from point-light walkers. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 1247–1265.

Quanty, M. B., Keats, J. A., & Harkins, S. G. (1975). Prejudice and criteria for identification of ethnic photographs. *Journal of Personality and Social Psychology*, *32*, 449–454.

Rim, S. Y., Min, K. E., Uleman, J. S., Chartrand, T. L., & Carlston, D. E. (2013). Seeing others through rose-colored glasses: An affiliation goal and positivity bias in implicit trait impressions. *Journal of Experimental Social Psychology*, *49*, 1204–1209.

Rule, N. O., Garrett, J. V., & Ambady, N. (2010). On the perception of religious group membership from faces. *PLoS ONE*, *5*, e14241.

Schaller, M., & Neuberg, S. L. (2012). Danger, disease, and the nature of prejudice (s). *Advances in Experimental Social Psychology*, *46*, 1–54.

Schaller, M., & Park, J. H. (2011). The behavioral immune system (and why it matters). *Current Directions in Psychological Science*, *20*, 99–103.

Singh, D. (1993). Adaptive significance of female physical attractiveness: Role of waist-to-hip ratio. *Journal of Personality and Social Psychology*, *65*, 293–307.

Srull, T. K., & Wyer, R. S., Jr. (1989). Person memory and judgment. *Psychological Review*, *96*, 58–83.

Stern, C., West, T. V., Jost, J. T., & Rule, N. O. (2013). The politics of gaydar: Ideological differences in the use of gendered cues in categorizing sexual orientation. *Journal of Personality and Social Psychology*, *104*, 520–541.

Stroessner, S. J. (1996). Social categorization by race or sex: Effects of perceived non-normalcy on response times. *Social Cognition*, *14*, 247–276.

Stroessner, S. J., Scholer, A. A., Marx, D. M., & Weisz, B. M. (2015). When threat matters: Self-regulation, threat salience, and stereotyping. *Journal of Experimental Social Psychology*, *59*, 77–89.

Uleman, J. S., Newman, L. S., & Moskowitz, G. B. (1996). People as flexible interpreters: Evidence and issues from spontaneous trait inference. In M. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 28, pp. 211–280). San Diego, CA: Academic Press.

Uleman, J. S., Rim, S. Y., Saribay, S. A., Kressel, L. M. (2012). Controversies, questions, and prospects for spontantaneous social inferences. *Social and Personality Psychology Compass*, *6*, 657–673.

Uleman, J. S., Saribay, S. A., & Gonzalez, C. M. (2008). Spontaneous inferences, implicit impressions, and implicit theories. *Annual Review of Psychology*, *59*, 329–360.

Winkielman, P., Halberstadt, J., Fazendeiro, T., & Catty, S. (2006). Prototypes are attractive because they are easy on the mind. *Psychological Science*, *17*, 799–806.

Winkielman, P., Schwarz, N., & Nowak, A. (2002). Affect and processing dynamics: Perceptual fluency enhances evaluations. In S. Moore & M. Oaksford (Eds.), *Emotional Cognition: From Brain to Behaviour* (pp. 111–136). Amsterdam, NL: John Benjamins.

Zárate, M. A., & Smith, E. R. (1990). Person categorization and stereotyping. *Social Cognition*, *8*, 161–185.

Zebrowitz, L. A., & Collins, M. A. (1997). Accurate social perception at zero acquaintance: The affordances of a Gibsonian approach. *Personality and Social Psychology Review*, *1*, 204–223.

2 From Spontaneous Trait Inferences to Spontaneous Person Impressions

Alexander Todorov
The University of Chicago Booth School of Business

I have great memories from my graduate student days at New York University. Our student offices didn't have windows and I could clearly hear the answering machines in the adjacent offices, but we were happy and productive. The social psychology area was highly collaborative and most of us worked with multiple faculty members. I worked with John Bargh, Shelly Chaiken, Yaacov Trope, and Jim Uleman. The lab groups were collaborative, with some weekly meetings being highly argumentative, while others calm. Regardless, they all were fun.

At the time, I was fascinated by subliminal priming effects and models of assimilation and contrast effects in judgments. However, my fascination met with mixed empirical success. Subliminal priming effects were generally weak and that made design of complex experiments with multiple factors challenging. In response to these normal quirks of experimental exploration, I started what was as a side project at the time. I started a series of studies with Jim Uleman on spontaneous trait inferences. This side project turned into my dissertation and shaped my research for the next 20 years.

Spontaneous Person Inferences

Jim was a pioneer in research on unintentional higher-level inferences (Uleman et al., 1996; Winter & Uleman, 1984), and we set out to work on an unresolved question in the area of spontaneous trait inferences. We asked how people form trait inferences (e.g., "honest") from behavioral statements (e.g., "Bob returned the lost wallet."). It was well established that such inferences are made spontaneously upon reading a trait-implying behavioral statement, but it was not clear whether such spontaneous inferences are free-floating inferences that are simply temporary accessible in memory or are bound to the representation of the agent who performed the behavior. This question was important, because the two possibilities have radically different implications for person perception in particular and social cognition in general. According to the first possibility of free-floating inferences, spontaneous inferences are important in the immediate situation but have no long-term implications for person representations. According to the second

DOI: 10.4324/9781003045687-3

possibility of agent-bound representations, these inferences modify person representations. The evidence from cued recall paradigms in which the inference ("honest") cues the recall of the agent and the behavior was mixed with respect to these two possibilities (Winter & Uleman, 1984). Whereas this inference did facilitate the recall of the behavior, it did not seem to facilitate the recall of the agent. Or at least it only seemed to do so in perceivers with motives that encouraged the encoding of links to the agent (Moskowitz, 1993).

To test whether trait inferences are bound to the representation of the agents, we designed a false recognition paradigm. In this paradigm, participants study faces with trait-diagnostic behaviors for a subsequent memory task. Later in a recognition test, they see face-trait pairs (e.g., Bob's photo and "honest") and decide whether they saw the trait in the sentence presented with Bob's face. In our first paper (Todorov & Uleman, 2002), we showed that participants were more likely to falsely recognize implied traits when presented with the agent's photo than when presented with other familiar photos. This effect was robust and large (the average effect size across experiments in terms of Pearson's *r* was 0.66)—an effect size in stark contrast to what I was observing with my attempts to create subliminal priming paradigms. The effect did not seem to depend on the number of faces and behaviors participants were exposed to. Across experiments, this number varied from 36 to 120. Moreover, the effect was not dependent on explicit memory for the behaviors. Even when participants did not recall or recognize the specific behavior, they were more likely to associate the implied trait with the agent's face. Analyses at the level of the stimuli (behavioral statements) showed that false recognition rates of implied traits were predicted by the strength of the trait implications of the behavioral statements (as measured by explicit judgments of a separate group of participants), showing that spontaneous inferences are highly specific and their strength varies as a function of the behavioral evidence. These findings, coupled with findings from the savings in relearning paradigm (Carlston & Skowronski, 1994; Carlston et al., 1995), demonstrated that spontaneous trait inferences are bound to the representation of the person enacting the behavior.

Encouraged by the robustness of the evidence for links between inferred traits and agents' faces, in our second paper (Todorov & Uleman, 2003), we studied to what extent the processes leading to these links are relatively independent of attentional resources. In earlier experiments, we presented the faces and behaviors for 5 or 10 seconds (if self-paced, participants typically spend a little over 6 seconds per face and behavior). In our first experiment (Todorov & Uleman, 2003), we included a condition, in which each face-behavior pair was presented for only 2 seconds. Nonetheless, participants were more likely to falsely recognize implied traits in the context of the agent's face than in the context of another familiar face. In our second experiment, we induced shallow processing of the information by asking participants to count the number of nouns in each sentence. Although this

manipulation reduced the false recognition effect, it did not eliminate it. In the third experiment, we introduced cognitive load. Participants were asked to rehearse six-digit numbers while reading the behavioral statements. Once again, the false recognition effect was present and large in size. In a final experiment, we collected person and behavior judgments of the behavioral statements. When asked to make a judgment about a person from the statement "Bob returned the lost wallet," participants considered the question, "Is Bob an honest person?" In contrast, when asked to make a judgment about a behavior, they considered the question, "Is this an honest behavior?" We used these two types of judgments to predict false recognition rates across experiments, including our initial experiments in Todorov and Uleman (2002). Person judgments, but not behavior judgments, predicted the false recognition rates, showing that people infer and associate traits with agents' faces rather than simply associate the meaning of behaviors with faces. These findings clearly supported the hypothesis that spontaneous trait inferences modify specific person representations.

Yet it was not clear from our previous studies whether trait associations are specifically bound to the representation of the face of the agent who performed the behavior rather than to any face that happened to be co-present with the behavior. In our final paper (Todorov & Uleman, 2004), we modified the learning trials of the false recognition paradigm to include two faces and a behavior referring to one of the faces. Participants were more likely to associate the traits with the face of the person who performed the behavior than with the control face. This effect, though reduced, persisted after a week. Interestingly, after a week the hit rate of correct recognition of presented traits was indistinguishable from the false recognition rate of implied traits. But in both cases, these traits were more likely to be associated with the right face. We also ruled out that our findings could be explained by differential attention to the faces. In the final two experiments, on each learning trial participants were presented with two faces and two behaviors, each referring to one of the faces. This paradigm forced participants to pay attention to both of the faces and behaviors. Nonetheless, we again found that participants were more likely to associate the implied traits with the faces of the actors who performed the trait-implying behaviors. Finally, we obtained the same results when we used different images of the same face identity during learning and testing, showing that spontaneously inferred traits are associated with abstract person representations rather than with specific image representations of faces.

Findings from the two most prominent paradigms for detecting spontaneous trait inferences—false recognition (Todorov & Uleman, 2002) and savings in relearning (Carlston & Skowronski, 1994)—clearly demonstrate that such inferences are bound to the representations of the agents who enacted the behavior. There are differences between these paradigms (see Crawford et al., 2007; Goren & Todorov, 2009), but the similarities are more important. Both rely on the retrieval of traits implied by behaviors, and these

traits are cued by photos of the agents initially presented with the behaviors. The key to success of both of these paradigms is not so much the specific measures they use, but the presence of faces. In contrast to names or other labels such as occupations, faces are highly distinctive, memorable, and the natural stimuli around which to organize person memories.

The Importance of Faces

The work with Jim led me to conduct systematic studies of the importance of the face in social cognition (Todorov, 2017; Todorov et al., 2015). This work, as well as the training in John Bargh's and Yaacov Trope's labs, also introduced persistent themes in my research: the efficiency and the importance of social judgments.

In experiments conducted with Jim, we were not interested in facial appearance per se. Typically, we randomly assigned behaviors to faces, as well as counterbalanced faces and behaviors, to make sure that the observed effects are due to the behaviors paired with the faces. But faces are a rich source of social inferences. Already in the 1950s, Paul Secord conducted a number of studies demonstrating that people infer traits from faces (e.g., Secord, 1958). Leslie Zebrowitz conducted seminal studies in the 1980s showing how specific facial characteristics such as baby-faced features trigger specific trait inferences such as "naïve" (e.g., Berry & Zebrowitz McArthur, 1985, 1986; Montepare & Zebrowitz McArthur, 1986; Zebrowitz McArthur & Apatow, 1984).

Following in their steps, I started systematic studies on inferences from faces in my newly formed lab at Princeton Univesity (Oosterhof & Todorov, 2008; Todorov et al., 2008). Two sets of initial findings demonstrated the importance and efficiency of social judgments from faces. In the first set of findings, we showed that naïve judgments of competence based solely on the facial appearance of politicians predicted electoral success (Todorov et al., 2005). The findings were surprising, but replicated in many different contexts (Antonakis & Dalgas, 2009; Lawson et al., 2010; Olivola & Todorov, 2010a; Poutvaara et al., 2009; Sussman et al., 2013). Studies by political scientists showed that the effects of appearance on voting decisions are limited to those voters who know next to nothing about politics and are exposed to images of the politicians (Ahler et al., 2017; Lenz & Lawson, 2011), a great example of heuristic processing where shallow, rapid inferences substitute more cognitively demanding inferences from substantive information (Hall et al., 2009).

The second set of findings was that people need minimal exposure to faces to form specific trait inferences such as trustworthiness (Willis & Todorov, 2006). In our initial studies, we presented faces for 100, 500, or 1,000 ms. Contrary to our expectations, judgments did not differ as a function of the length of exposure. The only effect of the latter was to increase confidence in judgments. Subsequent studies used better masking procedures and presented faces for even shorter exposures (Ballew & Todorov, 2007; Bar et al., 2006; Borkenau et al., 2009; Porter et al., 2008; Rule et al., 2009; Todorov et al., 2010; Todorov et al., 2009).

Generally, as little as 34 ms exposure is sufficient for people to form a judgment that is correlated with judgments made in the absence of time constraints, and this correlation doesn't increase in magnitude with exposures longer than about 200 ms (Todorov et al., 2009; 2010). Trait inferences from faces are literally single glance impressions.

Although there is little evidence that trait inferences from facial appearance are accurate (Hassin & Trope, 2000; Olivola & Todorov, 2010b; Todorov, 2017; Todorov et al., 2015), these initial findings showed that these inferences are highly efficient and matter for important social outcomes. In terms of the construction of social judgments, the findings also showed that people agree on these judgments. This agreement formed the basis of one of the questions that has guided much of the research in my lab for more than a decade (Oosterhof & Todorov, 2008; Todorov & Oh, 2021). The question was, given the agreement in judgments, how can we identify the perceptual basis or the configurations of facial features that lead to specific trait inferences.

To answer this question, we developed data-driven computational methods, which do not depend on prior hunches of what facial features are important for judgments (Oosterhof & Todorov, 2008; Todorov et al., 2011; Todorov & Oh, 2021; Todorov & Oosterhof, 2011). These methods were necessary, because it was practically impossible to discover the configurations of features that matter for judgments in the standard hypothesis-driven framework. In the latter framework, one posits that a set of features (e.g., shape of mouth, shape of eyebrows) influences judgments (e.g., friendliness) and then manipulates these features to test their effects on judgments. But manipulating just 10 binary facial features in a factorial design results in over 1,000 combinations; and manipulating 20 binary features results in over a million. Moreover, features are not binary and we don't even know what constitutes a feature (e.g., mouth vs. lips vs. corner of lips). Finally, features would not even be manipulated, if the experimenter doesn't think that they are important for judgments.

In our data-driven framework, we used a statistical model of face representation, in which each face is represented as a 100-dimensional vector. The appearance of each face is perfectly determined by its coordinates in this multi-dimensional face space. Rather than manipulating features, we simply randomly sampled faces from the multi-dimensional face space, and asked participants to judge the faces on various trait dimensions. Given the average trait judgment, we can then build a model of this judgment that captures the variation in appearance that is important for the judgment. The process is akin to finding the regression line, predicted from 100 orthogonal predictors (the coordinates of the faces), that accounts for most variance in judgments (for a detailed review of the methods, see Todorov & Oh, 2021).

Over the years, we have generated dozens of models of trait judgments (Funk et al., 2016; Oh, Buck, et al., 2019; Oh, Dotsh, et al., 2019; Said & Todorov, 2011; Todorov et al., 2013). These models can manipulate the

appearance of novel faces parametrically, increasing or decreasing their perceived value on trait dimensions such as trustworthiness and competence. Based on the models, we have created many databases of faces parametrically manipulated on trait dimensions and made those available for academic research. More than 4,000 users from over 900 institutions have used these databases for research, addressing a variety of questions: from studying infants' sensitivity to facial signals of trustworthiness and dominance (Jessen & Grossmann, 2016) to the effects of appearance on economic decisions (Rezlescu et al., 2012) and voting preferences (Laustsen & Petersen, 2016).

Although the models described previously are models of explicit judgments, they are easily extendable to implicit measures of judgments. Moreover, models of implicit measures could be immediately related to models of explicit judgments, because both are in the same statistical multidimensional space. That is, the similarity of models (whether based on explicit or implicit measures) is immediately given; it is simply captured by the correlation of the models (e.g., each model is a vector in the same multidimensional face space).

In recent work, harking back to my days at New York University, Ran Hassin and I collaborated to build a model of faces that break faster into consciousness (Abir et al., 2018). At New York University, Ran and I spent a lot of time arguing with each other; and unconscious processes were a core interest of the social cognition group back then (Hassin et al., 2005). Using continuous flash suppression, which suppresses visual input from one of the eyes, we measured the response times to detecting faces breaking into consciousness (e.g., being seen by the suppressed eye). We built a model of these response times, capturing the variation in facial appearance that emerges faster in consciousness. This model was highly correlated with a model of judgments of dominance. Recently, we have also built models of neural measures to faces (Cao et al., 2020). Such models of implicit measures can capture the content of truly spontaneous impressions.

My work with Jim was about associating trait inferences with person representations. As it turned out, faces were the critical stimuli to detect these associations. The importance of faces led me to studying trait inferences based solely on facial appearance, but I never abandoned the original question we studied with Jim.

The Robustness of Associating Affective Inferences with Faces

The models of judgments from faces are extremely powerful, but they also mask individual differences in trait inferences from facial appearance, simply because these are models of aggregated judgments. The typical measure of agreement in judgments is Cronbach's alpha. A high alpha of 0.90 simply indicates the expected correlation between the aggregated judgments of two groups (with the same size) of raters. But the average correlation between

raters within a group would be much smaller, typically of the order 0.30. In fact, partitioning the reliable variance in judgments from faces to shared variance with others and to idiosyncratic (individually stable) variance shows that the only judgment, in which these are relatively equal, is attractiveness. For any other social judgment, such as approachability, the idiosyncratic variance is much larger than the shared variance (Martinez et al., 2020).

What determines idiosyncratic contributions to trait inferences from facial appearance? One possibility is similarity to the faces of significant others, another theme that has its origins at New York University (Andersen & Cole, 1990; Chen & Andersen, 1999; Kraus & Chen, 2010). To the extent that different individuals have different-looking significant others, friends, and foes and these individuals use the similarity of strangers to their familiar others to make trait inferences based on this resemblance, there should be systematic individual differences in trait inferences. To experimentally test this hypothesis, we followed the logic of our studies with Jim on inducing trait inferences from behavioral statements (Verosky & Todorov, 2010). In the first stage of our experiments, we had participants associate faces with positive, negative, or neutral behavioral statements. Then, we asked them to make judgments of novel faces, which were subtly morphed with the familiar faces. Participants judged novel faces more positively when they were morphed with faces associated with positive information and more negatively when they were morphed with faces associated with negative information. In a subsequent study, we showed that this learning generalization from familiar others occurred even when participants were explicitly asked to disregard facial similarity information and made their judgments under cognitive load (Verosky & Todorov, 2013). Such processes of learning generalization based on similarity to familiar others are one of the mechanisms underlying learning to trust (FeldmanHall et al., 2018).

The studies described previously led me to a series of studies, which are a direct descendant of my work with Jim. In these studies (Falvello et al., 2015; Ferrari et al., 2020; Verosky et al., 2018), we did not use a false recognition paradigm, but we studied highly related questions about the nature of the associations between faces and the evaluative trait implications of behaviors. In the experiments, participants were first presented with faces and behaviors, which varied in valence, and then evaluated the faces without the presence of the behaviors.

As described in the previous section, people need minimal exposure to faces to form trait inferences. The mechanisms underlying this finding are straightforward to explain, given the computational work on models of judgments. The trait inferences are triggered by specific configurations of facial features. But in the case of trait associations with faces, there is nothing in the physical appearance of the face that "codes" the association. For the association to be retrieved, one needs to access a specific representation of the person who performed the behavior, perhaps requiring extra cognitive resources. To explore this question, we contrasted the effects of inferences

from facial appearance and the effects of inferences from behavioral information (Verosky et al., 2018). Rather surprisingly, the effect of inferences from behaviors was detectable after 35 ms exposure to the face: participants evaluated more positively faces associated with positive behaviors than faces associated with negative behaviors. If anything, this effect was larger than the effect of appearance (evaluating "trustworthy-looking" faces more positively than "untrustworthy-looking"). In a second study, we introduced a response deadline procedure forcing participants to make rapid judgments. Nonetheless, the effect of inferences from behaviors was detectable after 35 ms exposure to the face, although the effect was reduced in size. Finally, we measured the recognition of the faces. Not surprisingly, as face recognition increased, so did the effect of inferences from behaviors: the difference between the evaluation of faces associated with positive behaviors and the evaluation of faces associated with negative information increased. But the effect of inferences from behaviors was detectable at exceedingly low levels of face recognition. This effect emerged when participants reported recognizing the faces at a recognition value of three (on a nine-point scale), which was below the average value of recognition for novel faces. This was also the case for face exposures as short as 27 ms. These findings show how powerful social learning is in modifying person representations.

In our studies with Jim on spontaneous trait inferences, we used as many as 120 face-behavior pairs and observed that the effect size of the false recognition effect did not seem to vary as a function of the number of face-behavior pairs. To test whether there are limits on the ability to form affective associations with faces, we presented participants with as many as 500 face and behavior pairs (Falvello et al., 2015). We expected that as the number of faces and behaviors increases, the effect of inferences from behaviors on evaluation of faces would decrease. Surprisingly, we found that this effect was as strong after seeing 400 faces and behaviors as after seeing 100. A post-hoc analysis across three experiments suggested that the effect might start decreasing after seeing 300 faces and behaviors. But given the post-hoc nature of the analysis, it remains to be determined when affective associations with faces start breaking down.

Another surprising finding of the previous study was that we found similar effects for scenes. Participants were able to form affective associations with scenes paired with positive or negative descriptions, and the strength of the effect was similar to the effect for faces. This finding suggested that both kinds of affective associations (with faces and scenes) are driven by the same affect-based mechanisms and that perhaps the rich person-attribution processes, which we posited in the case of spontaneous trait inferences, are not necessary. To test this possibility, we contrasted learning of associations with faces and learning of associations with places (e.g., scenes and houses) (Ferrari et al., 2020). The key manipulation was whether the statements were relevant (e.g., a behavior paired with a face; a positive scene description with a scene) or irrelevant (e.g., a behavior paired with a scene). We found that

when statements were repeated, participants formed affective associations with places irrespective of the relevance of the source of these affective associations. This finding is consistent with a simple associative affect-based mechanism. In contrast, affective associations with faces were much stronger when the source of associations was relevant (e.g., behaviors).

Taken together, our findings show that people are remarkably good at forming affective associations with faces from relevant behavioral information, that these associations are specific to the person who performed the behavior, and that they are rapidly triggered by the mere presence of the person's face. All these findings were foreshadowed by my early work with Jim on spontaneous trait inferences and find a new expression in the recent research of Melissa Ferguson, a peer from NYU and a member of the lab groups of Bargh and Trope. Her recent work shows that implicit impressions can be rapidly updated (just like explicit impressions) in light of relevant behavioral information (Ferguson et al., 2019; Shen et al., 2020).

Beyond Inferences from Faces and Behaviors

All of the inferences described previously had to do either with inferences from facial appearance or behavioral statements, but they need not be limited to these two sources of information. People would use whatever information is available to rapidly form coherent person impressions. Two recent research examples are on inferences from bodily information and clothing cues.

Indeed, bodily information informs inferences of emotional expressions (Aviezer et al., 2012a, 2012b; Aviezer et al., 2015; Hassin et al., 2013). The driving force behind these studies was Hillel Aviezer, who was a post-doc with me and Yaacov Trope. Before joining our labs, Aviezer had already shown that people cannot ignore bodily information when inferring facial expressions of emotions (Aviezer et al., 2011; Aviezer et al., 2008). An expression of disgust is instantaneously perceived as anger, if the face expressing disgust is perched on a body about to hit someone. We studied extreme real-life emotions (e.g., winning, losing, pain, pleasure) and found that when people were only shown faces, they could not discriminate between positive and negative emotions (Aviezer et al., 2012a). In contrast, when shown bodies, they were pretty good at discriminating the valence of emotions. Yet, when asked what is the main source of their emotion inferences, the majority of participants believed that it was the face rather than the body. When provided with the intact images (e.g., faces and bodies), participants rapidly disambiguated the emotional expressions without ever occurring to them that the expressions were ambiguous.

The second example is about inferences of competence from clothing cues indicating economic status (Oh et al., 2020). In this work, we asked participants to make judgments of competence from faces. The critical manipulation was that the faces were presented with upper body clothing that was either perceived as "richer" or "poorer," though none of the clothing indicated poverty. The same face was evaluated as more competent when paired with "richer" than

with "poorer" clothing. Moreover, in nine experiments, we failed to eliminate this effect. We presented the faces for brief time, we told participants to ignore the clothing, we told them that the people depicted in the photos worked similar jobs and earned similar salaries, and we told them that the clothing was completely undiagnostic for real competence. In one study, we introduced large incentives (the participant who made judgments most similar to judgments of the faces alone was paid $100). None of these manipulations eradicated the effect of clothing on inferences of competence.

Conclusion

Inferences about people are powerful and many of them have the characteristics of automatic processes (Bargh, 1994): they are efficient, often unintentional, often uncontrollable, and often we are not aware of the cues that really influence our judgments. When encountering other people, we grab on whatever information is available at the moment to rapidly form spontaneous person impressions. Spontaneous trait inferences are part of this process. They are not just trait inferences; they are trait inferences that become integrated into the representation of the person. This is precisely their functional significance. After all, information about past actions is a more reliable source of person inferences than facial appearance or clothing.

References

Abir, Y., Sklar, A. Y., Dotsch, R., Todorov, A., & Hassin, R. R. (2018). The determinants of consciousness of human faces. *Nature Human Behaviour*, *2*, 194–199.

Andersen, S. M., & Cole, S. W. (1990). "Do I know you?": The role of significant others in general social perception. *Journal of Personality and Social Psychology*, *59*, 384–399.

Antonakis, J., & Dalgas, O. (2009). Predicting elections: Child's play! *Science*, *323*, 1183–1183.

Ahler, D. J., Citrin, J., Dougal, M. C., & Lenz, G. S. (2017). Face value? Experimental evidence that candidate appearance influences electoral choice. *Political Behavior*, *39*, 77–102.

Aviezer, H., Bentin, S., Dudarev, V., & Hassin, R. R. (2011). The automaticity of emotional face-context integration. *Emotion*, *11*, 1406–1414.

Aviezer, H., Hassin, R. R., Ryan, J., Grady, C., Susskind, J., Anderson, A., Moscowitch, M., & Bentin, S. (2008). Angry, disgusted, or afraid? Studies on the malleability of emotion perception. *Psychological Science*, *19*, 724–732.

Aviezer, H., Messinger, D. S., Zangvil, S., Mattson, W. I., Gangi, D. N., & Todorov, A. (2015). Thrill of victory or agony of defeat? Perceivers fail to utilize information in facial movements. *Emotion*, *15*, 791–797.

Aviezer, H., Trope, Y., & Todorov, A. (2012a). Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science*, *338*, 1225–1229.

Aviezer, H., Trope, Y., & Todorov, A. (2012b). Holistic person processing: Faces with bodies tell the whole story. *Journal of Personality and Social Psychology*, *103*, 20–37.

Ballew, C. C., & Todorov, A. (2007). Predicting political elections from rapid and unreflective face judgments. *Proceedings of the National Academy of Sciences*, *104*, 17948–17953.

Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion*, *6*, 269–278.

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In J. R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (2nd ed., Vol. 1, pp. 1–40). Hillsdale, NJ: Erlbaum.

Berry, D. S., & Zebrowitz McArthur, L. (1985). Some components and consequences of a babyface. *Journal of Personality and Social Psychology*, *48*, 312–323.

Berry, D. S., & Zebrowitz McArthur, L. (1986). Perceiving character in faces: The impact of age-related craniofacial changes on social perception. *Psychological Bulletin*, *100*, 3–18.

Borkenau, P., Brecke, S., Möttig, C., & Paelecke, P. (2009). Extraversion is accurately perceived after a 50-ms exposure to a face. *Journal of Research in Personality*, *43*, 703–706.

Cao, R., Li, X., Todorov, A., & Wang, S. (2020). A flexible neural representation of faces in the human brain. *Cerebral Cortex Communications*, *1*, 1–12.

Carlston, D. E., & Skowronski, J. J. (1994). Saving in the relearning of trait information as evidence for spontaneous inference generation. *Journal of Personality and Social Psychology*, *66*, 840–856.

Carlston, D. E., Skowronski, J. J., & Sparks, C. (1995). Savings in relearning: II. On the formation of behavior-based trait associations and inferences. *Journal of Personality and Social Psychology*, *69*, 420–436.

Chen, S., & Andersen, S. M. (1999). Relationships from the past in the present: Significant-other representations and transference in inter- personal life. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 31, pp. 123–190). San Diego, CA: Academic Press.

Crawford, M. T., Skowronski, J. J., & Stiff, C. (2007). Limiting the spread of spontaneous trait transference. *Journal of Experimental Social Psychology*, *43*, 466–472.

Crawford, M. T., Skowronski, J. J., Stiff, C., & Leonards, U. (2008). Seeing, but not thinking: Limiting the spread of spontaneous trait transference II. *Journal of Experimental Social Psychology*, *44*, 840–847.

Falvello, V., Vinson, M., Ferrari, C., & Todorov, A. (2015). The robustness of learning about the trustworthiness of other people. *Social Cognition*, *33*, 368–386.

FeldmanHall, O., Dunsmoor, J. E., Tompary, A., Hunter, L. E., Todorov, A. T., & Phelps, E. A. (2018). Stimulus generalization as a mechanism for learning to trust. *Proceedings of the National Academy of Sciences*, *115*, E1690–E1697.

Ferguson, M. J., Mann, T. C., Cone, J., & Shen, X. (2019). When and how implicit impressions can be updated. *Current Directions in Psychological Science*, *28*, 331–336.

Ferrari, C., Oh, D., Labbree, B., & Todorov, A. (2020). Learning the affective value of people: More than affect-based mechanisms. *Acta Psychologica*, *203*, 103011.

Funk, F., Walker, M., & Todorov, A. (2016). Modelling perceptions of criminality and remorse from faces using a data-driven computational approach. *Cognition and Emotion*, *31*, 1431–1443.

Goren, A., & Todorov, A. (2009). Two faces are better than one: Eliminating false trait associations with faces. *Social Cognition*, *27*, 222–248.

Hall, C. C., Goren, A., Chaiken, S., & Todorov, A. (2009). Shallow cues with deep effects: Trait judgments from faces and voting decisions. In E. Borgida, J. L. Sullivan,

& C. M. Federico (Eds.), *The political psychology of democratic citizenship* (pp. 73–99). Oxford University Press.

Hassin, R. R., Aviezer, H., & Bentin, S. (2013). Inherently ambiguous: Facial expressions of emotion in context. *Emotion Review*, *5*, 60–65.

Hassin, R., & Trope, Y. (2000). Facing faces: Studies on the cognitive aspects of physiognomy. *Journal of Personality and Social Psychology*, *78*, 837–852.

Hassin, R., Uleman, J. S., & Bargh, J. A. (Eds.) (2005). *The New Unconscious*. New York: Oxford University Press.

Jessen, S., & Grossmann, T. (2016). Neural and behavioral evidence for infants' sensitivity to the trustworthiness of faces. *Journal of Cognitive Neuroscience*, *28*, 1728–1736.

Kraus, M. W., & Chen, S. (2010). Facial-feature resemblance elicits the transference effect. *Psychological Science*, *21*, 518–522.

Laustsen, L., & Petersen, M. B. (2016). Winning faces vary by ideology: How nonverbal source cues influence election and communication success in politics. *Political Communication*, *33*, 188–211.

Lawson, C., Lenz, G. S., Baker, A., & Myers, M. (2010). Looking like a winner: Candidate appearance and electoral success in new democracies. *World Politics*, *62*, 561–593.

Lenz, G. S., & Lawson, C. (2011). Looking the part: Television leads less informed citizens to vote based on candidates' appearance. *American Journal of Political Science*, *55*, 574–589.

Martinez, J. E., Funk, F., & Todorov, A. (2020). Quantifying idiosyncratic and shared contributions to judgment. *Behavior Research Methods*, *52*, 1428–1444.

Montepare, J. M., & Zebrowitz McArthur, L. (1986). The influence of facial characteristics on children's age perceptions. *Journal of Experimental Child Psychology*, *42*, 303–314.

Moskowitz, G. B. (1993). Individual differences in social categorization: The influence of personal need for structure on spontaneous trait inferences. *Journal of Perosnality and Social Psychology*, *65*, 132–142.

Oh, D., Buck, E. A., & Todorov, A. (2019). Revealing hidden gender biases in competence impressions of faces. *Psychological Science*, *30*, 65–79.

Oh, D., Dotsch, R., Porter, J., & Todorov, A. (2019). Gender biases in impressions from faces: Empirical studies and computational models. *Journal of Experimental Psychology: General*, *149*, 323–342.

Oh, D., Shafir, E., & Todorov, A. (2020). Economic status cues from clothes affect perceived competence from faces. *Nature Human Behavior*, *4*, 287–293.

Olivola, C. Y., & Todorov, A. (2010a). Elected in 100 milliseconds: Appearance-based trait inferences and voting. *Journal of Nonverbal Behavior*, *34*, 83–110.

Olivola, C. Y., & Todorov, A. (2010b). Fooled by first impressions? Reexamining the diagnostic value of appearance-based inferences. *Journal of Experimental Social Psychology*, *46*, 315–324.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, *105*, 11087–11092.

Porter, S., England, L., Juodis, M., Ten Brinke, L. & Wilson, K. (2008). Is the face the window to the soul?: Investigation of the accuracy of intuitive judgments of the trustworthiness of human faces. *Canadian Journal of Behavioural Science*, *40*, 171–177.

Poutvaara, P., Jordahl, H., & Berggren, N. (2009). Faces of politicians: Babyfacedness predicts inferred competence but not electoral success. *Journal of Experimental Social Psychology*, *45*, 1132–1135.

Rezlescu, C., Duchaine, B., Olivola, C. Y., & Chater, N. (2012). Unfakeable facial configurations affect strategic choices in trust games with or without information about past behavior. *PLoS ONE*, *7*, e34293.

Rule, N. O., Ambady, N., & Adams, R. B. (2009). Personality in perspective: Judgmental consistency across orientations of the face. *Perception*, *38*(11), 1688–1699.

Said, C. P., & Todorov, A. (2011). A statistical model of facial attractiveness. *Psychological Science*, *22*, 1183–1190.

Secord, P. F. (1958). Facial features and inference processes in interpersonal perception. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior* (pp. 300–315). Stanford University Press.

Shen, X., Mann, T. C., & Ferguson, M. J. (2020). Beware a dishonest face? How we update our implicit impressions of untrustworthy faces. *Journal of Experimental Social Psychology*, *86*, 103888.

Sussman, A. B., Petkova, K., & Todorov, A. (2013). Competence ratings in US predict presidential election outcomes in Bulgaria. *Journal of Experimental Social Psychology*, *49*, 771–775.

Todorov, A. (2017). *Face value: The irresistible influence of first impressions*. Princeton University Press.

Todorov, A., Dotsch, R., Porter, J. M., Oosterhof, N. N., & Falvello, V. B. (2013). Validation of data-driven computational models of social perception of faces. *Emotion*, *13*, 724–738.

Todorov, A., Dotsch, R., Wigboldus, D. H. J., & Said, C. P. (2011). Data-driven methods for modeling social perception: Modeling social perception. *Social and Personality Psychology Compass*, *5*, 775–791.

Todorov, A., Loehr, V., & Oosterhof, N. N. (2010). The obligatory nature of holistic processing of faces in social judgments. *Perception*, *39*, 514–532.

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science*, *308*, 1623–1626.

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology*, *66*, 519–545.

Todorov, A., & Oosterhof, N. (2011). Modeling social perception of faces. *IEEE Signal Processing Magazine*, *28*, 117–122.

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition*, *27*, 813–833.

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, *12*, 455–460.

Todorov, A., & Uleman, J. S. (2002). Spontaneous trait inferences are bound to actors' faces: Evidence from a false recognition paradigm. *Journal of Personality and Social Psychology*, *83*, 1051–1065.

Todorov, A., & Uleman, J. S. (2003). The efficiency of binding spontaneous trait inferences to actors' faces. *Journal of Experimental Social Psychology*, *39*, 549–562.

Todorov, A., & Uleman, J. S. (2004). The person reference process in spontaneous trait inferences. *Journal of Personality and Social Psychology*, *87*, 482–493.

Todorov, A., & Oh, D. (2021). The structure and perceptual basis of social judgments from faces. In *Advances in experimental social psychology* (Vol. 63, pp. 189–245). Academic Press.

Uleman, J. S., Newman, L. S., & Moskowitz, G. B. (1996). People as flexible interpreters: Evidence and issues from spontaneous trait inference. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 28, pp. 211–279). San Diego, CA: Academic Press.

Verosky, S. C., Porter, J. M., Martinez, J. E., & Todorov, A. (2018). Robust effects of affective person learning on evaluation of faces. *Journal of Personality and Social Psychology*, *114*, 516–528.

Verosky, S. C., & Todorov, A. (2010). Generalization of affective learning about faces to perceptually similar faces. *Psychological Science*, *21*, 779–785.

Verosky, S. C., & Todorov, A. (2013). When physical similarity matters: Mechanisms underlying affective learning generalization to the evaluation of novel faces. *Journal of Experimental Social Psychology*, *49*, 661–669.

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, *17*, 592–598.

Winter, L., & Uleman, J. S. (1984). When are social judgments made? Evidence for the spontaneousness of trait inferences. *Journal of Personality and Social Psychology*, *47*, 237–252. Also see correction in *Journal of Personality and Social Psychology* (1986), *50*, 355.

Zebrowitz McArthur, L., & Apatow, K. (1984). Impressions of baby-faced adults. *Social Cognition*, *2*, 315–342.

3 Expressed Accuracy: Spontaneous Trait Production and Inference from Voice

Emily Sands and Lasana T. Harris
University College London, Experimental Psychology

The English idiom "don't judge a book by its cover" implores one not to determine value by outward appearance. This idiom also highlights social cognition research that demonstrates spontaneous inferences from relatively limited information over a wide range of socially relevant characteristics, including emotional states, personality traits, values, beliefs, and attitudes (Uleman & Bargh, 1989; Mohammadi & Vinciarelli, 2015). Such *spontaneous trait inferences* (STI; Newman & Uleman, 1990) are cognitive processes predicated on behavior or *spontaneous trait production* (STP). In this chapter, we discuss both spontaneous processes in the auditory domain: trait inference and production from the voice.

Like non-verbal behaviors, our ability to produce and infer traits from voice is socially derived, adaptive, and acquired across the lifespan (Funder, 1995; Harris, 2010). Social experiences lead to adaptive behaviors based on expectations of social outcomes (Frühholz & Schweinberger, 2021). Consequently, adults spontaneously describe behaviors in trait specific terms, relying on an index of prototypical behaviors when interacting with an unknown person (Newman & Uleman, 1990). Similarly, we develop and store an index of prototypical vocal cues or acoustic parameters (i.e., prosody) that informs trait inferences from voice, generating predictions about traits, emotions, and mental states when listening to an unknown person (McAleer et al., 2014; Mohammadi & Vinciarelli, 2015; Scherer, 1972).

Prosody describes a combination of acoustic parameters (e.g., pitch, decibels, shimmer) that helps the listener infer the mental state of the speaker. Stated differently, it is not only what someone says (words), but how someone says it (prosody) that provides the information listeners need to decide whether to approach or avoid another person. Thus, STP and STI are important social cognitive processes because the way we perceive others initiates approach and avoidant motivational tendencies and subsequent behaviour (Mohammadi & Vinciarelli, 2015). People modulate prosody when attempting to regulate social interactions, facilitate communication goals, express connections or social bonds, and manage impressions (Bänziger et al., 2015).

Approximately 38% of non-lexical communication occurs through prosodic features (Mehrabian, 2007), yet research on prosody and trait inferences has

DOI: 10.4324/9781003045687-4

been limited due to the complexity of phonetic variation, subjectivity concerns, and lack of experimental control. Prosodic features are often difficult to isolate and measure because they appear in more than one dimension and are gradient in nature (see Figure 3.1). For example, paralinguistic features such as age and health continuously change throughout one's life span, and drastically affect how acoustic parameters are expressed and perceived, making such features hard to isolate and measure (Scholtz, 2002). Moreover, prosody is modulated by cultural, situational, and individual motivations that inhibit and or dictate STP and STI. Due to this volatility and the potential to confound acoustic parameters with numerous social unknowns, the voice as a behavioral signal is contextually dependent (Chapman & Allport, 1938; Gray, 1982;

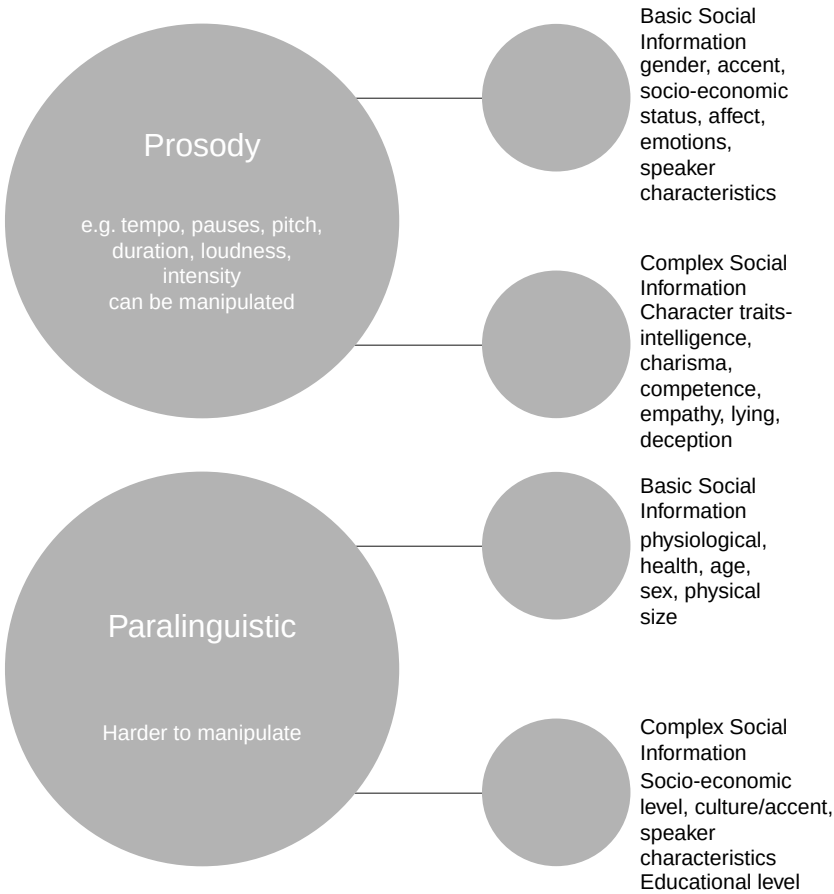

Figure 3.1 Prosody contains nonverbal information about an individual's physical state, emotions, and intentions towards others. We have less control over paralinguistic features.

Mcadams & Pals, 2006; Mcaleer et al., 2014). Nonetheless, scientists studying STP from voice typically measure brief vocal segments; this tactic fails to capture how contextual and paralinguistic factors interact with acoustic parameters during a social interaction, minimizing the most important aspect of prosody—its use as a socially motivated communication tool (Casasanto & Lupyan, 2015; Deyoung, 2014).

Due to the previous complexities, research tends to focus more on vocal *emotive* expression. In particular, research focuses on (a) inferring emotions from voice, and (b) understanding how specific emotions are expressed through the voice. Such studies explore a small number of basic emotions such as anger, sadness, joy, and disgust; the research shies away from complex more emotions or traits. The consequence of this approach is that research minimizes the importance of modelling STP and STI processes as a unified theoretical and functional approach (Bezooijen, 1984; Scherer et al., 1991), creating disjointed and nondescript outcomes that do not address the interactions between social context, motivations, and acoustic parameters.

The Problem of the Big Five

When researchers decide to study traits rather than emotions from voice, they define traits as predefined and stable. Where an individual's internal state (personality traits) will create contextually independent and outwardly predictable behaviors (Matthews et al., 2008), rather than traits being situationally dependent and directly determined by the social context, individual goals and motivations. This limitation of STI and STP from voice research occurs in part because of the unchallenged use of the Big Five Personality Assessment Model (Big5)—the most widely researched personality taxonomy—as a framework for traits. It is organized across broad traits as a five factor-analytically derived taxonomy, most commonly labeled extraversion, neuroticism, conscientiousness, agreeableness, and openness to experience (Costa & McCrae, 1994; Goldberg, 1993; John & Srivastava, 1999; Wood et al., 2015). Though its taxonomy is well established and offers a dimensional system that can be used to summarize traits as fixed representations of the self, the Big5 is not designed to measure underlying motivational and goal-directed processes that occur during social interactions. Such processes modulate STP and STI, limiting the Big5's ability to serve as a trait a measurment for the transitory nature of vocal trait expression in dynamic real-world interactions.

For example, imagine someone who is considered an introvert, but who decides at there place of work to strategically express traits more commonly associated with extroverts such as charisma in order to receive a promotion. Similarly, a doctor expressing empathy with their voice to gain a patient's trust without truly feeling empathetic when treating the patient is utilizing a trait for a goal directed outcome. Since the Big5 defines traits as enduring over time and context, it is irrelevant as a framework in these examples due to the complexity of such dynamic behavior. These examples highlight how

STP and STI extend beyond fixed representations of our personality. Therefore, use of the Big5 minimizes the importance of social judgment, and misses the matched component between the desired trait expressed and the trait inferred by eliminating the analysis of the speaker's intentions, motives, and goals from the communication process (Brunswik, 1956).

AI and Synthetic Voices

A decade ago, computer scientists began to recognize the importance of personality and social psychology for integrating voice with artificial intelligence (AI). This led them to the Big5, making it the foundational method for measuring and integrating automatic personality recognition (APR) and automatic personality perception (APP; Vinciarelli & Mohammadi, 2014). While computer scientists quote the validity and consistency of the Big5 as an accurate framework for measuring traits, rarely do they supply evidence of its validity and functional application for vocal communication. However, the need for convenient operational methods for binary computational functions motivates the continued use of the Big5, which offers ridged representations of trait perception that can be easily identified and categorized.

However, diverse computational applications of ARP and APP as well as varying experimental methods has led to less than satisfactory trait-accuracy outcomes. High accuracy is necessary for any classifier or synthetic voice. APP and APR researchers have thus examined the utility and feasibility of studying the encoding, transmission, and decoding of traits from voice with the goal of identifying classifiers for computational approaches. The first substantial findings explored whether it was possible to infer traits from dialogue derived from random conversations (Mairesse et al., 2007). They also systematically examined acoustic parameters and word-use, and compared them to inferences about traits made by listeners. Accuracy required performance above chance or 50% in these binary choice paradigms. Extroversion was the only trait that was identified by listeners at above chance levels, with 65% accuracy. Similarly, Polzehl et al. (2010) supplied 220 vocal samples of a professional actor conveying 10 traits from the Big5 to listeners and found that listener accuracy across traits was at 60%. Mohammadi and Vinciarelli (2012) used a large corpus of vocal clips randomly extracted from 96 French speakers reading Swiss national news bulletins. A corpus of 11 French-speaking judges produced low inter-rater reliability when making trait inferences about the speaker that ranged between $r = 0.12$ and 0.28 depending on the trait. Outcomes for trait accuracy were slightly higher than previous findings, which were between 60% to 70% depending on the trait, with extroversion and conscientiousness as traits that were easiest to accurately detect. Additionally, Valente et al. (2012) integrated a social component into their study design by using a corpus of real-world conversations. Accuracies for trait perception ranged from 50% to 68%. Finally, the inclusion of more acoustic parameters only resulted agreeableness and openness at around 55% accuracy (Mairesse et al., 2007; Mohammadi

et al., 2012; Polzehl et al., 2010; Valente et al., 2012). Ultimately, researchers have been largely unsuccessful getting listeners to accurately infer Big5 traits with high accuracy.

Similar failings have occurred when algorithms instead of people attempt to accurately infer traits. The 2012 INTERSPEECH Speaker Trait Challenge (Schuller et al., 2012) was the first rigorous comparison of the accuracy with which different algorithms using the same data set, experimental controls, could identify the Big5 traits in speakers. 640 emotionally neutral vocal clips were randomly selected from French news bulletins. These clips did not include any words associated with well-known places or people, and consisted of clips from professional (N = 307) and nonprofessional speakers (N = 333). In addition to algorithms, 11 French-speaking listeners also categorized Big5 traits from voice. Once again inter-rater reliability was extremely low for trait inferences, with the average between 0.12 and 0.28 depending on the particular trait (Schuller et al., 2012). Listeners were most accurate when identifying extroversion. Further, no algorithm outperformed another, and none could reliably predict traits from voice, suggesting no optimal solution (Schuller et al., 2012; Vinciarelli & Mohammadi, 2014). Once again, there was little mention in the overview of findings as to the consideration and or misuse of the Big5 as a reliable instrument in measuring trait expression from the voice.

In all of the previous studies, researchers identified, tested, and analyzed inferences made about Big5 traits based on vocal cues. Researchers varied prosodic elements, vocal-social interactions, and computational methods to probe APP and APR with limited success. Accuracy and reliability were low. All experiments used the same questionnaire and rating system with no variation in application. All experiments also minimized or eliminated altogether the social context, social display rules, and their potential effects on the transmission process from speaker to listener. While all researchers questioned the Big5's validity and functionality to varying degrees, none offered a solution nor considered an alternative method of classification. To date, the Big5 continues to be the most widely used procedural element in all APP and APR research with very little challenge to its validity during social communication. Instead, computational models have become more complex in an attempt to organize and classify data outputs under the Big5 rubric.

Challenges and Solutions When Inferring Traits from Voice

Agreement Among Listeners

We argue that underlying or enduring personality traits are irrelevant factors when producing speech or when listeners are making spontaneous trait inferences from voice; more accurate trait inference from voice requires a probabilistically formulated integrated process model that accounts for speaker's and listeners' goals, intentions, motives, and the social context.

A trait inferred in a specific social interaction is dependent on and modulated by the social environment. Therefore, accuracy requires consensus among listeners and speakers who share the same social and display rules that determine the appropriate traits in a given social context. We must also consider behavioral predictions based on motivations (e.g., cooperation, competition, relationship building), relationship stratification within the social interaction (e.g., peer, parent, superior), as well as the potential for these variables to interact with paralinguistic features. By considering the social context, motivations, and the roles of the speaker and listener, we might be able to address accuracy as dependent on the speaker's and listener's shared social expectations and motivations.

An example is the approach/avoidance behavior, and or cooperation and competition. If someone communicates friendliness through the voice, one should assume that it is okay to approach. *Perceptual accuracy should thus be defined as a matched component; when the appropriate response ensues, the speaker can conclude that their intended expressed trait was accurately inferred by a listener.* By encouraging researchers to map accuracy in such a way, we are considering the quick statistical computations the brain is making at every moment when interacting with a person. Theoretically, such inferences can be identified with AI through regression techniques and ordinal scales, though further research is needed to determine how this can be achieved.

Acknowledgment of Developmental Stages, Motivations, and Culture

Listeners belonging to different cultures tend to assign different traits to the same speaker. Thus, it is not possible to know if a particular trait was inferred given the Big5 criterion that assumes trait perception is nonconditional and decontextualized (McAdams & Pals, 2006). To measure trait inferences from voice, it is imperative to understand the social and display rules that moderate prosody and its perception (Burgoon & Bacue, 2003). These are developmentally dependent, culturally specific, and interconnected. Therefore, STP to STI should be measured as a unit; an integrated model that has coevolved where voice perception largely relies on voice production, and meaning lies in the matching of the encoding and decoding cues that are shared, rather than an actual representation of a speaker's authentic self and a listener's ability to accurately perceive it. Researchers therefore need to measure both the speakers intent and the listener's inferences in order to determine accuracy.

A Lack of a Neutral Communication Signal

"Neutral" vocals as a control forgoes the most fundamental element of vocal expression—its use as a socially derived behavioral signal. For instance, a "neutral" news segment (such as those used as the source of vocal clips in the INTERSPEECH Challenge) is hardly neutral, as the story itself creates a

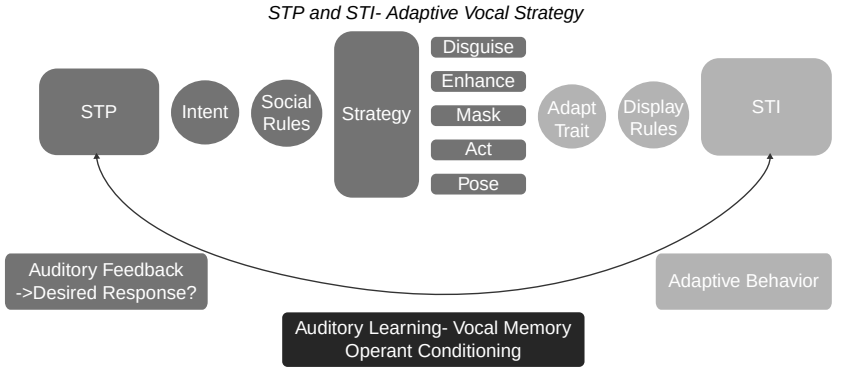

Figure 3.2 Expressed accuracy is an advanced adaptive strategic use of STP to enhance personal goals and motives. It is moderated situationally by the appropriate social display rules. Consequently, from the vantage point of self-interest, STP can be used to promote a favorable impression that determines STI. This cognitive process functions as a feedback loop. Did the desired response ensue? If so, does it inform future use in similar contexts?

social context. Consider how a professional reporter uses their voice actively to signal and elicit reactions from a listener when divulging information about a tragic event compared to when interviewing a politician. Now consider how such intentional vocal modulation might or might not affect the accuracy of the non-professional voices stating the same material, where their expressive intention and experiences are not within the same context as the professionals. Consequently, the vocal idiosyncrasies create different impressions (see Figure 3.2). Therefore, we must consider speaker intent and impression management since speakers can express or suppress traits to fulfill personal goals or motives.

To summarize, current personality trait theories consider persistent patterns of emotion, motivation, cognition, and behavior as the main factors influencing trait expression, focusing on each human's idiosyncrasies (Matthews, 2008). The Big5 is based on these assumptions, where an individual's internal state will create outwardly predictable behaviors that are context independent. By limiting the definition of accuracy to individual differences in STP through inert representations of the speaker's self, a fundamental component of STI from voice is missing: the *why*. Why is a particular acoustic parameter being utilized and why do people infer it similarly? We believe accuracy of STI from voice should be measured *as the matched component between a speaker's and listener's intention and a listener's matched perception*. As such, accuracy occurs when the speaker's intended trait is inferred by the listener, not when a listener hears a voice without context and guesses at a trait inference. This functionalist perspective of

accuracy provides a comprehensive assessment of the mechanisms underlying both human-human and human-AI vocal interactions. Stated differently, it is not whether you are accurately inferring someone's internally consistent state, but whether the person's desired expressed state is being accurately inferred.

Evolution and Development: Understanding the Voice as a Behavioral Signal

Here we postulate a new model of measurement and analysis of STP and STI from voice where acoustic parameters are used to define accuracy as *targeted expressivity*—an index that captures the pattern of acoustic parameters that predict the expressed trait. We outline our theory on measuring expressed accuracy of STP and STI from voice as follows:

a STP and STI depend on adaptive, developmental signaling cues that can be goal-oriented and recurrent;
b Accuracy is defined as *expressed accuracy* (Rosenthal et al., 1979) where STP is only successful to the extent that acoustic parameters are accurately decoded and result the intended STI; and,
c STP and STI should be measured via a multistep functional model based on shared social and display rules that are context dependent.

Evolutionary personality theorists argue that evolved mechanisms that serve as social signals (such as the voice) have proved useful over time, particularly in relation to an increase in social status or social relationship maintenance (Harris, 2010). Such mechanisms influence long-term patterns of social behavior within overlapping systems, which can be hierarchically organized to represent the social world. Each system in this hierarchical organisation of the social world is predicated on learned behaviors. The first learned system is the relationship system (forming and maintaining useful relationships), followed by the socialization system (forming and maintaining group membership), then finally the status system (competing successfully with rivals; Harris, 2010). Each system makes use of previously learned signaling devices such as the voice. This suggests that as the brain learns and social intricacies increase, more attuned signaling is required to achieve STP for strategic expression.

Our vocal development model considers how the voice evolved as a social signaling method, particularly in relation to STP and STI through an integrated functional approach (see Figure 3.3). STP requires social development and learning; as humans develop, and our social worlds become more dynamic, our ability and need to communicate grows in complexity. We rely on multiple biological and psychological mechanisms to interpret, respond, and learn advanced interpersonal techniques in the pursuit of individual goals (Fridja, 1986; Juslin, 1997; De Young, 2014). We thus build on basic

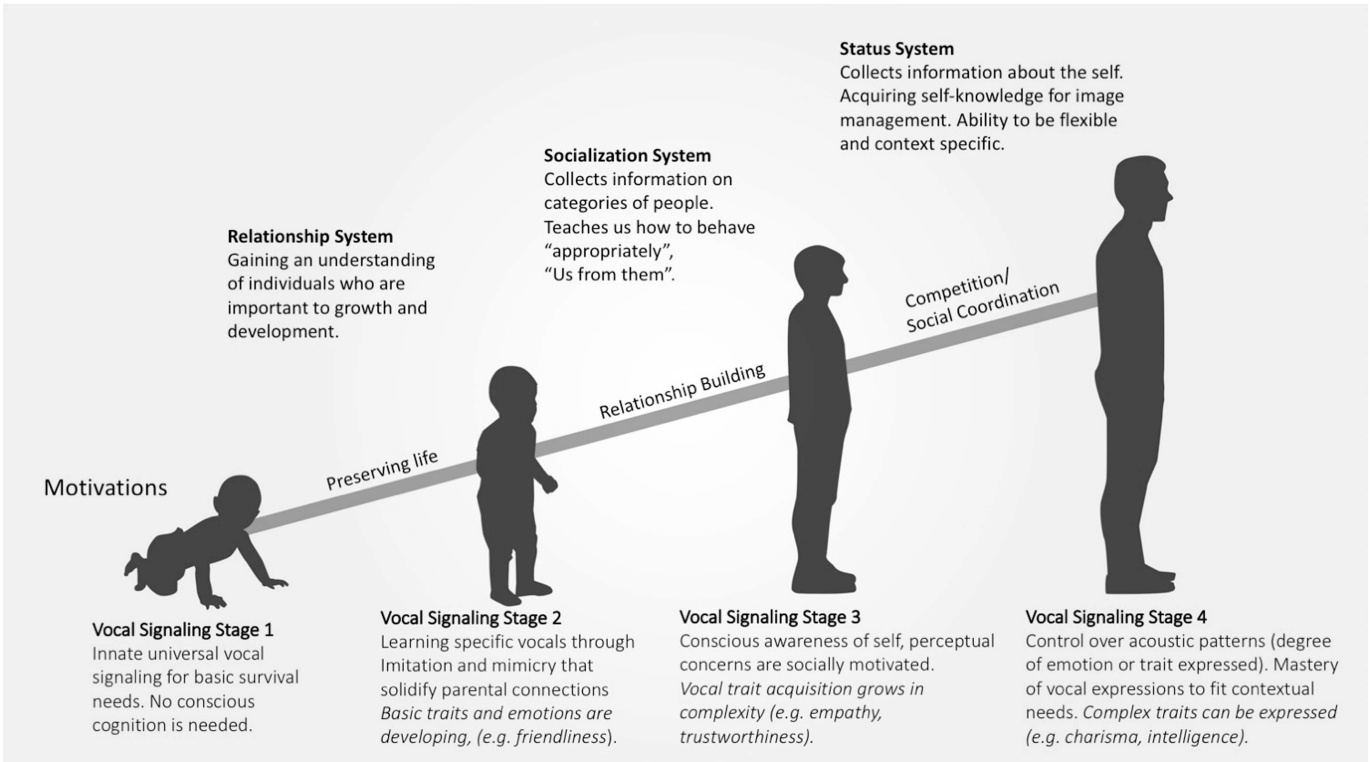

Figure 3.3 Our model of vocal development. Each stage builds on previous stages leading to an increased capacity in learned behavior that is stored as trait expressive memories, leading to complex STP production that is self-regulatory.

vocal skills, creating a vocal signaling hierarchy where certain traits are easier to express (STP) and to infer (STI), and less susceptible to contextual effects. These base-level traits (e.g., friendliness, trustworthiness) facilitate approach-avoidance behavior, enabling cooperation and building crucial social relationships. Other more complex traits such as charisma could encourage both approach and avoidance, but are more context dependent; someone acting charismatically at a funeral may be deemed inappropriate and avoided, rather than the same behavior at an ice-cream social. Base-level acoustic parameters are thus easier to express and perceive since they require less sophisticated cognitive skills and serve as a spontaneous communicative baseline. Thus, less precision is needed to express these base-level traits with the voice.

As social developmental stages increase, vocal learning abilities increase, and social needs become more complex, allowing for more strategic use of vocal prosody for complex traits (e.g., charisma, intelligence, competence, empathy). Each developmental level makes use of lower-level signaling devices (see Figure 3.3). As people develop and the complexity of social hierarchies and relationships increase, more contextually attuned signaling is required. Stated differently, people need to better understand how acoustic parameters interact with the social context. Such acoustic parameters emerge through an auditory feedback loop where the speaker must register and evaluate their own vocal performance based on social outcomes (e.g., imitation, mimicry, auditory feedback; see Figure 3.3—vocal stage 0–1; Frühholz & Schweinberger, 2021). Mimicry and imitation come from parental and close family unit instruction, allowing children to learn socially acceptable responses or display rules—a pattern of behavior, an expectation, or an emotional response. Further vocal development during adolescence (see Figure 3.3—vocal signaling stage 2–3) includes the ability to monitor vocal errors, fine-tune STP and expand learning (including the ability to decode, recognize, and index acoustic parameters) to create vocal trait templates that have desirable social effects.

The highest developmental level of our model describes strategic and volitional expression fine-tuned based on experience (see Figure 3.3—vocal signaling stage 4). Experience creates an index of probable outcomes that lead to predictions informed by the auditory feedback loop. This prepares the speaker for adaptable responses such as STP to enhance or suppress trait expression, as well as mastery of STI, including the capacity to create vocal patterns that result in positive social interactions. The speaker also learns to master the ability to be flexible, fluctuating acoustic parameters to fit varying contexts, social obligations, relationships, and prejudicial perturbations, all unique to human communication.

Here, we argue STP and STI function as a feedback loop, where the speaker considers their own goals, motives, or intentions and the best trait expression strategy (STP) to meet them (see Figure 3.4). The listener functions via similar processes, including making predictions for which traits

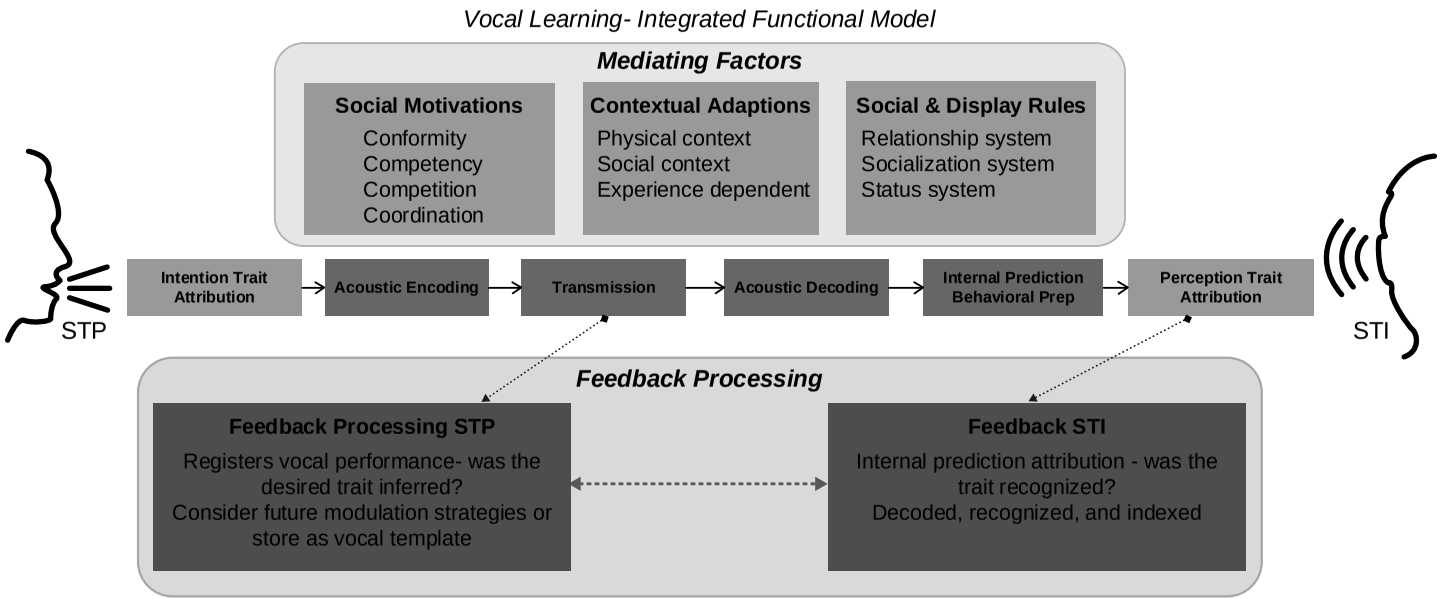

Figure 3.4 An integrated functional perspective where STI relies on STP. These processes function on a feedback loop where the speaker registers and evaluates their STP based on the appropriate behavioral response from the listener. The speaker and listener store these adaptions as vocal templates that are experience dependent and triggered for future use in expression or inference.

should be expressed in a content (STI; Uleman et al., 2005). If the expression and the inference align, the desired social effect ensues (Frühholz & Schweinberger, 2021; Rosenthal et al., 1979).

Further support for a matching approach to STI accuracy from voice comes from the brain; brain systems responsible for voice production and perception overlap, and include the basal ganglia (BG), which plays an essential role in vocal learning, particularly vocal modulation mediated by developmental changes and experience acquired through imitation and vocal practice (Frühholz & Schweinberger, 2021). These socialization processes require additional psychological processes including auditory memory, vocal error monitoring, and fine-tuning, as well as learning vocal tone modulations for accurate signaling. This neural overlap highlights the intrinsic connection between STP and STI as an integrated model to measure *expressed accuracy* (Frühholz & Schweinberger, 2021; Van Lanckersidtis et al., 2006).

To conclude, developmental stages of vocal signaling coincide with personality developmental stages hierarchically (see Figure 3.3). STP and STI have coevolved as psychological processes that mediate our ability to communicate and infer traits accurately throughout our lives (see Figure 3.4). STP and STI are based on conditioned interpretations of social display rules learned through developmental stages and individual experiences. In other words, a person may learn to attend to certain acoustic parameters, while not attending to others, and such attention can be situationally specific such that some parameters are attended to only under certain situations. However, individual motivations are governed by shared cultural and social display rules solidified as we identity with a larger cultural and social sphere. Social display rules ultimately dictate how and when we display certain traits regardless of an individual's idiosyncrasies. This allows us to make certain assumptions about which traits are likely to be displayed or not in given situations (Burgoon et al., 2017).

Further, our ability to progress to a level of strategic expression is based on a progressive vocal and auditory learning process where some traits are fundamental baseline markers necessary to establish and maintain relationships in an effort to enhance social connectivity associated with survival value. Hypothetically, baseline traits require less contextual information for interpretation and are therefore easier to imitate (Frijda, 1986; Juslin, 1997). As our abilities develop and social needs become more complex, we rely on our vocal communication skill level to decipher a variety of behavioral patterns. For example, we theorize that compared to friendliness, charisma is a complex trait. Specifically, charisma may have evolved as a nuanced but diverse expression of friendliness, trust, competence and extroversion, is often intentional, goal-directed, and strategically expressed to achieve tripartite instrumental, relational, and identity goals. Essentially, you must be a competent communicator to display charisma with your voice. Charisma is therefore more difficult to express, but not necessarily more difficult to perceive since the acoustic parameters may be more contextual derived, informing expectations. This suggests that if the speaker's STP matches the listener's expectations (e.g., a politician speaking to constituents), then

charisma is accurately inferred. Thus, prosody alone is not sufficient for interpreting and analysing STI from the voice, and as social lives become more complex, our ability to engage STI becomes more complex, requiring one to consider social motivations, rules, roles, and obligations associated with the specific social interaction to differentiate between traits that share similar prosodic elements.

Expressed Accuracy: An Integrated Approach to Modeling STP and STI

Communicating and inferring traits from voice requires consideration of a substantial range of data to manage the variety of variables that affect STP and STI in any given social interaction. We assert that it is impossible to accurately detect and or judge STI and STP simply through trait correlates of human judgment alone. Hence, we define *expressed accuracy* as the matched component between a speaker's expressed intention, and a listener's matched perception. Thus, STP and STI should be evaluated based on conditionally expected occurrence (social and cultural display rules), which rely on diverse data-driven methods that allow for contextual variation, and models of associations between STP and STI based on the likelihood that social context allows for a more plausible set of predictors for a given STP in a similar context (Crivelli & Fridlund, 2018). This method of conditional probability allows us to understand how acoustic parameters are influenced by the social context and motives to confirm the expected likelihood of the occurrence of an acoustic parameter's use under contextually relevant conditions. Ultimately, the goal of expressed accuracy is to operationalize the process that the brain quickly computes when STP and STI occur naturally in social interactions (see Figure 3.4). Here we theorize that vocal trait expression and inference are not contingent on long-term patterns of personality, but are instead a strategic method of signaling to meet immediate needs and obligations within the social interaction (Funder, 1995). A reflection of the ability to judge transitory states such as the state of "being empathetic" or being charismatic rather than measurements that pertain to someone's intrinsic longitudinal personality patterns or our "true selves".

Next, we outline a causal path for a classification system that uses a variety of conditional probability models such as frequentist and Bayesian interpretations that depend on situational and goal complexity. Our objective is not to prescribe meaning to each trait, but rather to approach design methodologies through a cause-and-effect lens where we focus on how STP and STI function within a social interaction (Funder, 1995; Crivelli & Fridlund, 2018). This allows us to simultaneously model how acoustic parameters are exploited to differentiate one trait from another, how acoustic parameters diverge or converge, and are used instantaneously or interchangeably, and how we can account for situational expectations that affect these processes.

To model expressed accuracy, participants should share the same social display rules. Stated differently, participants should be aware of the social

norms operating in specific social contexts, as well as their partner's interaction goals. It is essential that the design methodology selects a specific situational experience where social display rules can be easily monitored and considered (e.g., relationship system, status system, socialization system), as well as the intent of the speaker, and display expectations of the listener.

The Brunswik Model

Brunswik (1956) argued for more focus on the connections between perception and real-life settings. He proposed that successful adjustment to an unpredictable world requires an organism to rely on probabilistic inferences using tentative information (proximal cues) when making decisions about behavioral intentions (distal objects). These probabilistic references bypass the accuracy notion prevalent in fixed trait taxonomies such as the Big5. The best way to determine accuracy is to consider the variety of possible responses and decide which is best suited to the interaction. Stated differently, it is best to consider how social display rules and expectations allow for the likelihood of a certain trait being expressed in a specific context.

The Brunswik Lens Model (BLM) is enhanced through computational statistics including the lens model equation (LME) and the tripartite emotion expression perception model (TEEP) designed to compute communication achievement (Hammond et al., 1964; Juslin & Scherer, 2008). Although the Brunswik model was originally designed to focus on visual perception, it has been applied to interpersonal perception and nonverbal communication (Gifford & Hine, 1994), including the acoustic parameters that inform trait perception of status (Ko et al., 2014), emotion perception in music performance (Juslin, 1997), and vocal emotion communication of anger (Bänziger et al., 2015). Similarly, our model of *expressed accuracy* utilizes Brunswik's theory of person perception and the LME statistical modeling. By mapping perceptual accuracy in such a way, we are considering the quick statistical computations the brain is taking when interacting with a person.

We implemented the BLM in a study where we examined how social context, social roles and intent affect STP, and how social context, social roles, and acoustic cues affect STI from voice (Sands & Harris, under review). Specifically, we created a doctor-patient paradigm where participants (patients) listened to speakers (doctors) communicating information about prescription or bad medical news. We relied on the doctor-patient paradigm because social display rules are widely shared among English speakers from the United States and the United Kingdom. The speakers were native English-speaking professional actors who were asked to communicate five traits with their voice: charisma, intelligence, empathy, friendliness, and trustworthiness. To understand intent and strategic trait expression, the actors were instructed to remember an experience associated with the scenario (experience dependent), or to portray the trait spontaneously (experience independent). We also examined the implications of paralinguistic

features such as accent and gender relating to social display rules. To analyze our data, we completed a spectrogram analysis to measure the acoustic parameters comprising prosody for each trait using Pratt and ProsdyPro software. We used a large corpus of acoustic parameters in a data-driven attempt to have a vast catchment of acoustics without assumptions about which are used in STP and STI (see Figure 3.5).

The inference component of our study relied on ratings from the general public, and were culturally dependent (American and English). We used the BLM to identify expressed accuracy—the extent to which STP and STI corresponded—a necessary step in understanding the entirety of the voice in relation to the scope of its influence on trait perception (Ko et al., 2014). Specifically, within the framework of the BLM, we computed coefficients that described the strength of the match between expressed and perceived trait (see Figure 3.5).

Overall, our results show STI and STP are affected by context, gender, and speaker intent. For example, speakers showed convergent acoustic parameter patterns when expressing charisma and giving a patient a prescription, but speakers' acoustic parameter patterns were divergent when they were asked to give a patient bad news, suggesting that if the trait being expressed was inappropriately matched to a context (did not follow social display rules), acoustic parameters did not converge. Traits that were inappropriately matched also showed more volatility in inference ratings.

Conclusion and Future Directions

Jim Uleman's pivotal work on STI allowed researchers to study how people store and process trait information, clearing a path to re-examining STP as an evolved social signaling tool (Uleman et al., 1996). These findings paved the way for interdisciplinary researchers who created modular frameworks that considered adaptive models in unison with situational, cultural, and environmental factors (Uleman et al., 2012). One area of Uleman's work that is of particular interest to STI from voice is his consideration of the effects of social and cultural cues such as gender stereotyping (Ko et al., 2006, 2014; O'connor & Barclay, 2017; Suire et al., 2019; Uleman et al., 2005). Another area of significance was the definitive declaration by Uleman and colleagues that self-reporting methods do not accurately measure the person-environment variables that occur at the encoding stage (Uleman & Bargh, 1989), an assertion that grounded our theoretical approach particularly when considering the misuse of the Big5.

Similar to Uleman's findings across STI modalities, we found that the voice is inherently complex, often leading to inconclusive scientific results and more questions rather than answers. The Big5 as a catch-all measurement standard for STI from voice should be reassessed as it lacks the fundamental operationalization needed to analyze trait expression from voice as a contextually contingent communicative signal that facilitates social motivations. By marginalizing the situation-specific component of the voice, research on STI and

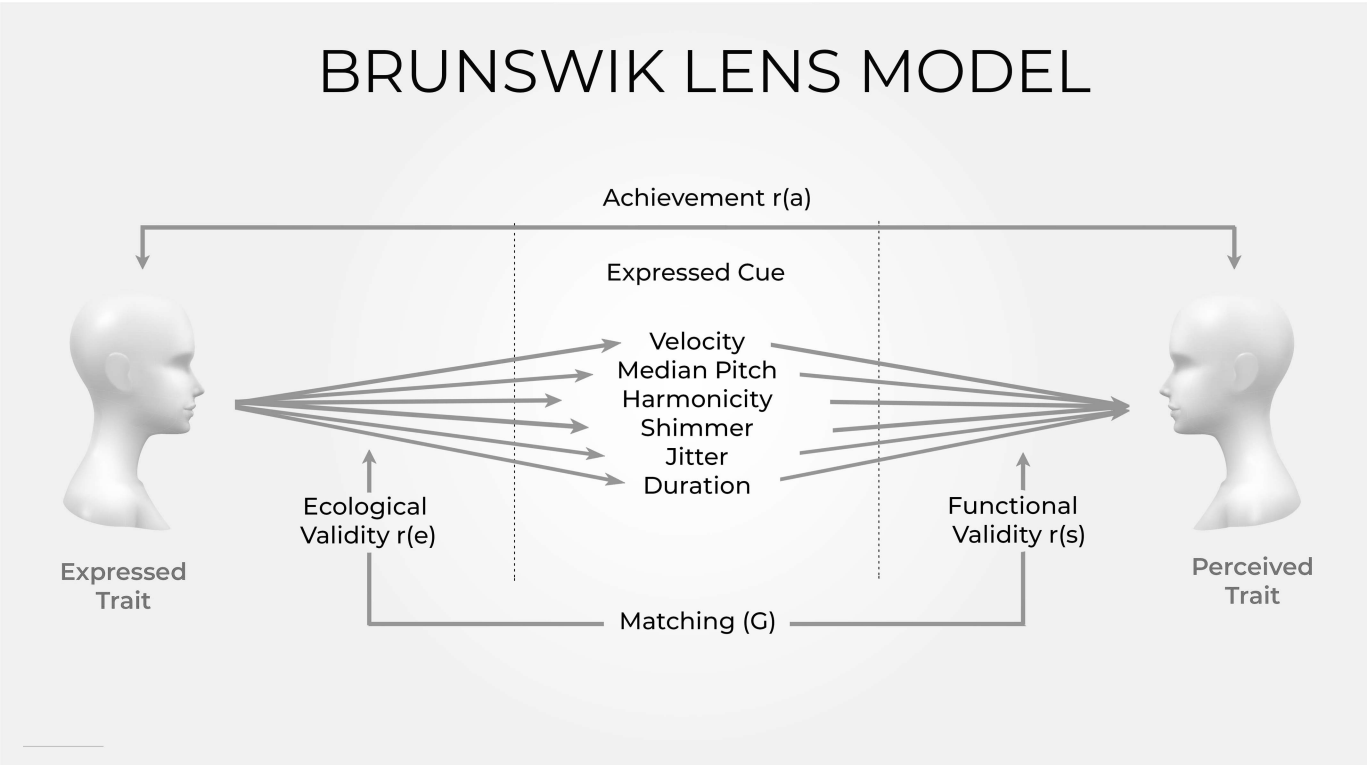

Figure 3.5 We applied the BLM to investigate the link between acoustic parameters and STP and STI. Unlike past models, we considered gender and accent as well as social context.

trait communication from voice is confined, lacking in the fundamental properties needed to form a functionalist perspective of accuracy.

During perception, what matters is not whether you are *accurately inferring someone's internally consistent state, but whether the person's desired expressed state is being accurately inferred*. If we want to re-create human interactions synthetically, we must consider the attribution of personality based on observable behavior (Vinciarelli & Mohammadi, 2014). This intermediary is the vocal social signal that governs social interactions. To further our understanding for computer science, particularly AI (Matthews et al., 2021), we must continue to use an interdisciplinary approach, one that considers social cognition, neuroscience, developmental and evolutionary psychology, as well as functionalism to investigate the complex multimodal system that governs trait expression from voice.

In conclusion, STI from voice is predicated on a basic developmentally and socially derived cultural and display rules. It is therefore essential that we consider STI from voice not as a singular experience, but rather as a collaboration between a speaker and a listener. To further research in this area researchers should: (1) Take the objective view that vocal trait expression is adaptive, goal-oriented, and learned; (2) consider that STP from voice is used to transmit social codes; (3) trait expression depends on flexible cognition and development, where culture, individual experiences, environmental influences, and their functionality determine adaptive significance; (4) there are basic traits whose acoustic parameters are easier to imitate and to use as a means to drive cooperation, conflict resolution, and competition; and (5) STI through the voice is based on a shared social code where accuracy is derived from matching (Juslin & Laukka, 2003). Future research should build on these principles as we have built off the work of Jim Uleman.

References

Bezooijen, R. V. (1984). *Characteristics and Recognizability of Vocal Expressions of Emotion*. Dordrecht, Holland : Foris Publications. doi:10.1515/9783110850390

Bänziger, T., Hosoya, G., & Scherer, K. R. (2015). Path Models of Vocal Emotion Communication. *Plos One*, *10*(9), e0136675. doi:10.1371/journal.pone.0136675

Brunswik, E. (1956). *Perception and the Representative Design of Psychological Experiments*. doi:10.1525/9780520350519

Burgoon, J. K., & Bacue, A. E. (2003). Nonverbal Communication Skills. In *Handbook of Communication and Social Interaction Skills* (pp. 179–221). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Burgoon, J. K., Dunbar, N. E., & Giles, H. (2017). Interaction Coordination and Adaptation. *Social Signal Processing* (pp. 78–96). doi:10.1017/9781316676202.008

Casasanto, D., & Lupyan, G. (2015). All Concepts Are Ad Hoc Concepts. In E. Margolis, & S. Laurence (Eds.), *The Conceptual Mind: New Directions in the Study of Concepts*, (pp. 543–566). Cambridge: MIT Press. doi:10.7551/mitpress/9383.003.0031

Chapman, D., & Allport, G. W. (1938). Personality: A Psychological Interpretation. *Sociometry*, *1*(3/4), 420. doi:10.2307/2785590

Costa Jr, P. T., & McCrae, R. R. (1994). Stability and Change in Personality from Adolescence through Adulthood. (pp. 139–150). Lawrence Erlbaum Associates.

Crivelli, C., & Fridlund, A. J. (2018). Facial Displays Are Tools for Social Influence. *Trends in Cognitive Sciences*, *22*(5), 388–399.

Deyoung, C. G. (2014). Cybernetic Big Five Theory. *Journal of Research in Personality*, *56*, 33–58. doi:10.1016/j.jrp.2014.07.004

Frijda N. H. (1986). *The Emotions*. Cambridge, UK: Cambridge University Press.

Frühholz, S., & Schweinberger, S. R. (2021). Nonverbal Auditory Communication–evidence for Integrated Neural Systems for Voice Signal Production and Perception. *Progress in Neurobiology*, 199, 101948.

Funder, D. C. (1995). On the Accuracy of Personality Judgment: A Realistic Approach. *Psychological Review*, *102*(4), 652–670. doi:10.1037/0033-295X.102.4.652

Gifford, R., & Hine, D. W. (1994). The role of verbal behavior in the encoding and decoding of interpersonal dispositions. *Journal of Research in Personality*, *28*(2), 115–132. doi:10.1006/jrpe.1994.1010

Goldberg, L. R. (1993). The Structure of Phenotypic Personality Traits. *American Psychologist*, *48*(1), 26.

Gray, J. A. (1982). Précis of the Neuropsychology of Anxiety: An Enquiry into the Functions of the Septo-hippocampal System. *Behavioral and Brain Sciences*, *5*(3), 469–484. doi:10.1017/s0140525x00013066

Harris, J. R. (2010). Explaining Individual Differences in Personality: Why We Need a Modular Theory. *The Evolution of Personality and Individual Differences*, (pp. 121–153). Oxford University Press. doi:10.1093/acprof:oso/9780195372090.003.0005

Hammond, K. R., Hursch, C. J., & Todd, F. J. (1964). Analyzing the components of clinical inference. *Psychological Review*, *71*(6), 438–456. doi:10.1037/h0040736

John, O., & Srivastava, S. (1999). The Big Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives. In *Handbook of personality: Theory and research*. 2nd ed. (pp. 102–138). New York: Gilford Press.

Juslin, P. N. (1997). Emotional Communication in Music Performance: A Functionalist Perspective and Some Data. *Music Perception*, *14*(4), 383–418. doi:10.2307/40285731

Juslin, P. N., & Laukka, P. (2003). Communication of Emotions in Vocal Expression and Music Performance: Different Channels, Same Code? *Psychological Bulletin*, *129*(5), 770–814. doi:10.1037/0033-2909.129.5.770

Juslin, P. N., & Scherer, K. R. (2008). *Scholarpedia*, 3(10), 4240.

Ko, S. J., Judd, C. M., & Blair, I. V. (2006). What the Voice Reveals: Within- and between-Category Stereotyping on the Basis of Voice. *Personality and Social Psychology Bulletin*, *32*(6), 806–819. doi:10.1177/0146167206286627

Ko, S. J., Sadler, M. S., & Galinsky, A. D. (2014). The Sound of Power. *Psychological Science*, *26*(1), 3–14. doi:10.1177/0956797614553009

Mairesse, F., Walker, M. A., Mehl, M. R., & Moore, R. K. (2007). Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text. *Journal of Artificial Intelligence Research*, *30*, 457–500. doi:10.1613/jair.2349

Matthews (2008). Personality and Information Processing: A Cognitive-adaptive Theory. *The SAGE handbook of personality theory and assessment: Volume 1—Personality theories and models*. (pp. 56–79). doi:10.4135/9781849200462.n3

Matthews, G., Hancock, P. A., Lin, J., Panganiban, A. R., Reinerman-Jones, L. E., Szalma, J. L., & Wohleber, R. W. (2021). Evolution and Revolution: Personality

Research for the Coming World of Robots, Artificial Intelligence, and Autonomous Systems. *Personality and Individual Differences*, *169*, 109969. doi:10.1016/j.paid.2020.109969

Mcadams, D. P., & Pals, J. L. (2006). A New Big Five: Fundamental Principles for an Integrative Science of Personality. *American Psychologist*, *61*(3), 204–217. doi:10.1037/0003-066x.61.3.204

Mcaleer, P., Todorov, A., & Belin, P. (2014). How Do You Say 'Hello'? Personality Impressions from Brief Novel Voices. *PLoS ONE*, *9*(3), e90779. doi:10.1371/journal.pone.0090779

Mehrabian, A. (2007). *Nonverbal communication*. New Brunswick, NJ.: Aldine Transaction.

Mohammadi, G., & Vinciarelli, A. (2012). Automatic Personality Perception: Prediction of Trait Attribution Based on Prosodic Features. *IEEE Transactions on Affective Computing*, *3*(3), 273–284. doi:10.1109/t-affc.2012.5

Mohammadi, G., Origlia, A., Filippone, M., & Vinciarelli, A. (2012). From Speech to Personality: Mapping Voice Quality and Intonation into Personality Differences. *Proceedings of the 20th ACM International Conference on Multimedia - MM 12*, 789–792. doi:10.1145/2393347.2396313

Mohammadi, G., & Vinciarelli, A. (2015). Automatic Personality Perception: Prediction of Trait Attribution based on Prosodic Features Extended Abstract. *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*. doi:10.1109/acii.2015.7344614

Newman L. S., Uleman J. S. (1990). Assimilation and Contrast Effects in Spontaneous Trait Inference. *Personality and Social Psychology Bulletin*, *16*(2), 224–240. doi:10.1177/0146167290162004

O'Connor, J. J., & Barclay, P. (2017). The Influence of Voice Pitch on Perceptions of Trustworthiness across Social Contexts. *Evolution and Human Behavior*, *38*(4), 506–512. doi:10.1016/j.evolhumbehav.2017.03.001

Polzehl, T., Moller, S., & Metze, F. (2010). Automatically Assessing Personality from Speech. *2010 IEEE Fourth International Conference on Semantic Computing*. doi:10.1109/icsc.2010.41

Rosenthal, R., Archer, D., Hall, J. A., DiMatteo, M. R., & Rogers, P. L. (1979). Measuring sensitivity to nonverbal communication: The Pons test, Nonverbal Behavior, 67–98. doi: 10.1016/b978-0-12-761350-5.50012-4

Scherer, K. R. (1972). Judging Personality from Voice: A Cross-cultural Approach to an Old Issue in Interpersonal Perception1. *Journal of Personality*, *40*(2), 191–210. doi:10.1111/j.1467-6494.1972.tb00998.x

Scherer, K. R., Banse, R., Wallbott, H. G., & Goldbeck, T. (1991). Vocal Cues in Emotion Encoding and Decoding. *Motivation and Emotion*, *15*(2), 123–148. doi:10.1007/bf00995674

Scholtz, S. (2002). Prosody in Relation to Paralinguistic Phonetics. *Speech Technology*, 43, 9.

Schuller, B., Steidl, S., Batliner, A., Noeth, E., Vinciarelli, A., Burkhardt, F., van Son, R., Weninger, F., Eyben, F., Bocklet, T., Mohammadi, G., & Weiss, B. (2012). The INTERSPEECH 2012 Speaker Trait Challenge. *Conference Paper*.

Suire, A., Raymond, M., & Barkat-Defradas, M. (2019). Male Vocal Quality and Its Relation to Females' Preferences. *Evolutionary Psychology*, *17*(3), 147470491987467. doi:10.1177/1474704919874675

Uleman, J. S., & Bargh, J. A. (Eds.). (1989). *Unintended thought*. New York: Guilford Press.

Uleman, J. S., Newman, L. S., & Moskowitz, G. B. (1996). People as flexible interpreters: Evidence and issues from spontaneous trait inference. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 29, pp. 211–279). San Diego: Academic Press.

Uleman, J. S., Blader, S. L., & Todorov, A. (2005). Implicit impressions. In R. R. Hassin, J. S. Uleman & J. A. Bargh (Eds.), *The new unconscious* (pp. 362–392). New York: Oxford University Press.

Uleman, J., Rim, S., Saribay, S., & Kressel, L. (2012). Controversies, Questions, and Prospects for Spontaneous Social Inferences. *Social and Personality Psychology Compass*, 6(9), 657–673. doi:10.1111/j.1751-9004.2012.00452.x

Valente, F., Kim, S., & Motlick, P. (2012). Annotation and Recognition of Personality Traits in Spoken Conversations from the Ami Meetings Corpus. *Proceedings of Interspeech*. 1183–1186. doi:10.21437/Interspeech.2012-125

Vinciarelli, A., & Mohammadi, G. (2014). A Survey of Personality Computing. *IEEE Transactions on Affective Computing*, 5(3), 273–291. doi:10.1109/taffc.2014.2330816

Van Lanckersidtis, D., Pachana, N., Cummings, J., & Sidtis, J. (2006). Dysprosodic speech following basal ganglia insult: Toward a conceptual framework for the study of the cerebral representation of prosody. *Brain and Language*, 97(2), 135–153. doi: 10.1016/j.bandl.2005.09.001

Wood, D., Gardner, M. H., & Harms, P. D. (2015). How Functionalist and Process Approaches to Behavior Can Explain Trait Covariation. *Psychological Review*, *122*(1), 84.

4 O Brother, O Sister, Who Art Thou? Inferring the Gender of Others in Ambiguous Situations

Amy Arndt and Marlone Henderson
University of Texas

Consider the following riddle:

> A father and his son are driving together in a car. They get into a serious accident and the man dies at the scene. When the child is taken to the hospital and rushed into the emergency room, the surgeon pulls away and says: "I can't operate on this boy, he's my son." How can this be?

Over two-thirds of people will answer that the surgeon is the boy's other father. However, only 30% of people report the most probable solution: the surgeon is the boy's mother (Belle et al., 2021). Whether people answer that the surgeon is the boy's other father, stepfather, mother, or even the boy's hallucination, there is one aspect these answers have in common—they require people make an inference about the surgeon's gender.

Goals

Categories about identity variables, such as gender or race, are often perceived as obvious and readily apparent (Cosmides et al., 2003; Fazio & Dunton, 1997). Initial research on categorization assumed that drawing inferences about a person's gender meant automatically matching a target's physical features to a gender category to infer one's gender quickly and confidently (see Bruner, 1957, for a general model on perceptual categorization). In this chapter, we demonstrate how this is frequently not the case in making gender inferences. Indeed, many common, everyday interactions involve making a gender inference under some degree of ambiguity or uncertainty. This chapter explores how this process occurs, including the components of a gender inference, the predictors of an inference, and the subsequent consequences of a gender inference.

Ultimately, the exploration of gender inferences serves to demonstrate how other categorical or identity variables, such as race, age, or sexuality, frequently can also be ambiguous, and require more complex inferences. Additionally, exploring how gender inferences are made opens a new avenue of research to explore how gender itself is conceptualized and how gender

DOI: 10.4324/9781003045687-5

inferences and subsequent gender stereotypes may influence communication in new and unexpected ways.

Gender Inferences

We use the term "gender" as a multi-component variable, including, but not limited to, aspects of gender identity, gender expression, and biological sex (Tate et al., 2014). When making a gender inference, a person is making a judgment about one or more of these components. Gender inferences can occur in two forms. First, a target can be inferred to either possess a gender or be genderless. Second, inferences can be made regarding a specific gender, such as inferring a target to be a man or a woman.

Studying gender inferences exposes not only how gender is understood, but how this understanding guides our future behaviors. Often, gender is assumed to be readily apparent. As such, it would not need to be inferred, but simply quickly identified. However, research demonstrates that these inferences are frequently incorrect (Herring & Stoerger, 2014). Many behavioral differences typically ascribed to gender may instead be a product of gender inferences. For example, are men better negotiators than women, or will people have an easier time in a negotiation if they are inferred to be male? By studying the underpinning of gender inferences—how they form and what they affect—we can not only predict future behavior, but perhaps influence it.

Formation of Gender Inferences

Categorizations about one's group have been considered "primitive categorizations," or categorizations that are innate, habitual, automatic, and "go beyond a certain level of certainty," (Bruner, 1957, p. 149). Gender inferences are thought to occur through an automatic system of feature-matching processes, sorting a person into a gender group based on which physical traits are present (Brewer, 1988; Fiske & Neuberg, 1990). These initial categories could then be expanded or recategorized, depending on a person's goal or motivation (Fiske & Neuberg, 1990). Instead, we argue that gender inferences can arise from multiple types of stimuli, and do not require physical features such as a body or a face to form. In ambiguous situations, gender inferences occur automatically, are pervasive across cultures and stimuli, and are generative of other inferences.

Automaticity

In psychology, when a thought occurs *automatically* it means it occurs without conscious effort. However, there is some debate over the exact criteria of automaticity. For many, a thought that is formed truly automatically must be

done without awareness, intention, and control and must formed efficiently (Bargh, 1994). However, others have suggested automaticity is rarely "process pure," and the process of producing a thought may not exhibit all four traits. Instead, many cognitive responses can vary in the extent to which they are the result of these four traits but still can be considered automatic (Jacoby, 1991).

Notably, perceptions are rarely fully automatic or fully controlled, and are instead a combination of the two. For example, dual process models of impression formation highlight how impressions are formed through a combination of automatic, or implicit, and explicit inferences, and how these two types of inferences can influence each other throughout impression formation (Wyer & Srull, 1988). To this end, many researchers have designed assessment tools to capture distinct implicit and explicit components, including the implicit association test, or IAT (Conrey et al., 2005), and process dissociation procedures, or PDP (Jacoby, 1991). In studying inferences, the term "automatic" encompasses inferences that are both uncontrollable and unconscious, often captured by these assessment tools and uncoupled from any explicit components (Uleman et al., 2008; Winter et al., 1985).

Researchers have tested if gender is inferred automatically, and the evidence is mixed. Gender may be perceived efficiently, in as little as 130 ms (Hügelschäfer et al., 2016), and gender categorization can still occur while the brain is engaged in other cognitive tasks (Jung et al., 2019). Additionally, gender perceptions show evidence of unintentionality (Ito & Urland, 2003, 2005). However, some level of conscious awareness may be required for processing gender (Amihai et al., 2011).

Whether or not gender is a purely processed automatic trait, gender inferences possess many characteristics of automatic inferences. In particular, people tend to sort others into gender categories without gender being explicitly mentioned or alluded to. A study on gender-fair language found that people form gender inferences from short vignettes, even when the subjects of the vignettes are intentionally written in gender-neutral ways (Arndt & Henderson, 2021a). That is, when asked to summarize the vignette, participants described the characters using gendered pronouns (e.g., "he") not found in the text. When later explicitly asked what gender the character was, participants explicitly gendered the character using the same gender as their original gender inference. Additionally, early work on implicit gender inferences found that gender is processed during encoding and is recalled faster than other trait inferences during retrieval (Smith & Miller, 1983). Finally, effects of gender inferences are found when gender is not given time to be explicitly processed. Priming short statements with a split-second image of a gendered face promoted trait inferences for stereotypically gendered traits (Yan et al., 2012). Even though participants could not consciously process the gender of the face due to the short exposure time, the implicit gender perception of the image promoted further inferences of gendered traits (i.e., masculine or feminine traits). Together, these results demonstrate that

gender inferences form when gender is not explicitly mentioned, form very quickly, and are formed when gender is not consciously processed; such evidence aligns with the major components of automaticity.

Pervasiveness

In addition to forming automatically, inferences are thought to be pervasive and occur across a variety of contexts and stimuli (Uleman et al., 2008). Inferences can occur when reading written text (Winter et al., 1985), observing people's actions in daily life (Elsbach et al., 2010), and while browsing online social media profiles (Levordashka & Utz, 2017).

Gender inferences, specifically, are particularly pervasive and occur across multiple mediums, stimuli, and perceiver cultures. While it may be apparent that gender inferences would quickly form from stimuli such as a person's face, gender inferences also form from less apparent stimuli, such as handwriting. Instead of gender categorization relying on individuals identifying aspects of a body or face to form inferences about gender (Pendry & Macrae, 1996), stimuli that do not contain a physical representation of a person are equally capable of prompting gender inferences. When teachers lament over the sloppy handwriting of their first graders, or FBI analysts comb a handwritten note from a crime scene, perceivers draw inferences about the gender of the creator (Burr, 2002; Lewis, 2014). Gender is one of the earliest variables people infer from handwriting (Burr, 2006; Sprouse & Webb, 1994), and gender inferences form automatically even from small amounts of stimuli. Most handwriting analysis studies rely on a few sentences for a handwriting sample (Bouadjenek et al., 2014; Hartley, 1991), but others show that gender inferences are formed when viewing one word or even a single letter (Burr, 2006; Hayes, 1996). Similarly, gender inferences based on handwriting have been shown to occur not only in English, but in other languages as well, including French, Urdu, Turkish, and Arabic (Akbari et al., 2017; Morera et al., 2018; Topaloglu & Ekmekci, 2017). Together, research on handwriting demonstrates that gender inferences can occur separate from a specific individual, on very little stimuli, and occur cross-culturally.

As with research on handwriting, research on online interactions also finds gender inferences also occur separate from a human form and with little initial stimuli present. Seeing an ambiguous username or email address (e.g., "puffyfish19") alone is enough to elicit a gender inference (Danet, 1988; Heisler & Crabill, 2006; Pelletier, 2009). In fact, gender is the most prevalent factor inferred about a person when seeing a username, with studies reporting between 74% and 80% of participants inferring a person's gender when viewing a gender-ambiguous username alone (Heisler & Crabill, 2006; Pelletier, 2009). In comparison, 65% of participants inferred a person's age and 55% inferred a person's race based on username alone, implying that gender-based social categorizations are stronger than categorizations based on other demographics (Hagström, 2012).

Binding and Generative Inferences

Lastly, gender inferences influence subsequent inferences about that person. When reading a behavioral statement such as "Joan aced the test," there are two possible inferences that could form. The first is that Joan is smart. However, it is also possible to infer that the test was easy. Research on trait inferences finds that people can form both inferences at once (Ham & Vonk, 2003; Todd et al., 2011). However, inferences that are made about individuals bind to the person (Todorov & Uleman, 2002, 2003) and operate in a generative manner (Chen et al., 2014; Frankenstein et al., 2020). That is, when people make an inference about a person's trait, they assume the person's future behavior will embody that trait. If a participant infers that a person is "kind," they predict that a person will act kindly towards others in the future or will avoid actions that are unkind (Frankenstein et al., 2020). Inferences about individuals can then generate new inferences. For example, a person first inferred to be a "leader" is more likely to be inferred as "confident" (Chen et al., 2014).

Gender inferences demonstrate the pattern of binding gender traits to individuals and creating further evaluative judgments about them. Once a target is perceived as male or female, a perceiver forms subsequent inferences about the target, such as the target's emotions or personality based on that gender inference (Fong & Mar, 2015). For example, perceptions of trustworthiness depend upon targets' perceived gender. Both children and adults report higher levels of trust for voices perceived as gender-congruent to the topic they are discussing (Lee et al., 2007). Similarly, traits about a target stemming from the target's perceived gender can result in additional downstream consequences. Gender inferences can invoke inferences of a target's competency or ability, in which men are normally perceived as more competent than women (Kuchenbrandt et al., 2014, Parks, 2004; Salvaggio et al., 2009). These competency judgments can then go on to affect future judgments or behaviors. For example, people are more likely to accept help from a target they perceive as male, versus female (Kuchenbrandt et al., 2014). The generative effects of gender inferences are explored further in the Effects of Gender Inferences section of this chapter.

Accuracy of Gender Inferences

Gender inferences may be automatic, prevalent, and generative of other trait inferences, but are they correct? Gender is often thought to be an explicit and objective trait, in that it is easily perceptible and unambiguous (Zosuls et al., 2011), two claims that modern gender research suggests questions (Hyde et al., 2019). Additionally, remote communication creates additional ambiguity regarding one's identity, allowing gender to be easily concealed, whether intentionally or unintentionally. As a result, ambiguity surrounding gender inferences creates situations in which gender inferences may not always be accurate.

Studies on electronic communication reveal a range in accuracy of gender inferences. Research looking at single messages and longer, online chat exchanges finds that people perform above chance in accurately inferring the gender of others, though a sizeable percentage of gender inferences are inaccurate (Cornetto & Nowak, 2006; Herring & Stoerger, 2014; Koch et al., 2005; Mou et al., 2019; Savicki et al., 1999). Overall, studies report accuracy ratings as low as 57% (Savicki et al., 1999) to a high of 95% when based on an individual message (Thomson & Murachver, 2001), though average accuracy ratings tend to fall between 60 and 70% (Cornetto & Nowak, 2006; Thomson & Murachver, 2001). When reading transcripts of communications between humans and machines (i.e., chatbots), accuracy ratings of the perceived gender of the human speaker drop to below chance (43%). This suggests that people's gender inferences do not adapt for any compensatory speaking patterns a person might make when conversing with nonhuman technology (Mou et al., 2019). Notably, even when accuracy levels are at their absolute highest, gender inferences are still incorrect around one out of every 10 inferences. Given how pervasive gender inferences are, this level of accuracy means that people will make many incorrect gender inferences in their lifetime.

Predictors of Gender Inferences

If gender inferences are not always accurate, factors outside of a target's gender identity must be influencing their formation. Most likely, there are an abundant number of individual, cultural, and situational variables that predict the outcome of any given gender inference. However, the literature on gender inferences converges on two variables that predict gender inferences across multiple mediums: gender stereotypes and androcentric bias.

Gender Stereotypes

Gender inferences are largely affected by gender stereotypes. Stereotypes surrounding the traits of a given target can prime an individual into inferring one gender over another. Additionally, the stereotypes of the overall context in which an inference takes place can also influence gender inferences. Both types of gender stereotypes work together to influence the formation of gender inferences.

Gender stereotypes surrounding target traits can promote stereotype-consistent gender inferences. For example, in online communication, gender stereotypes about a message's content can be used to infer gender. Participants report using stereotypes about the topics people talk about in their messages (e.g., football = male) as predictors of a person's gender. However, research indicates this is a poor predictor of accuracy of a gender inference (Koch et al., 2005). Even when participants were aware that gender deception may be taking place in electronic chats, people made gender inferences on easily

manipulated phrases that evoked gender stereotypes, instead of less easily manipulated characteristics, such as linguistic patterns (Herring & Martinson, 2004). Additionally, targets demonstrating high warmth characteristics, a stereotypically feminine trait, were frequently inferred as female. Researchers found that the more engaged a person was in an online chat, the more likely they were perceived as female (Savicki et al., 1999). In a similar vein, the more person-centered anonymous online commenters seemed to be, the more likely they were assumed to be a female (Spottswood et al., 2013). Overall, gender inferences in online communication come largely from gender stereotypes, whether it be in the direct content of the online messages or in the way messages are delivered.

In addition to making stereotype-consistent inferences from a target's characteristics, people form gender inferences based on the context in which the inference takes place. If a context is judged as feminine, a person is more likely to be inferred as female, whereas more masculine contexts are more likely to evoke male inferences. On social media sites, the type of site or blog can impact gender inferences, in which diary blogs are seen as feminine and more report-style blogs are seen as masculine (Herring & Paolillo, 2006). As such, people tend to be susceptible to gender biases based on location or setting, in which otherwise ambiguous users are ascribed the gender of the site's stereotype (Bivens & Haimson, 2016; Herring & Paolillo, 2006). Literature from gaming also shows that the setting can serve as a major predictor of inferred gender, as the genre of a game predicts the inferred gender of the game's players (Eden et al., 2010). For example, players in shooter games (stereotypically masculine) are much more likely to be assumed to be male than those in puzzle games or massive multiplayer online role-playing games (stereotypically less masculine).

The reliance on gender stereotypes in online settings is particularly notable, as adults, especially women, tend to present themselves as less gendered in online environments than offline environments (Oberst et al., 2016; van Doorn et al., 2007). In online chatrooms, only 33% of participants chose gendered human avatars (photos or cartoon representations accompanying a message), with men more likely to choose human avatars than women (Nowak & Fox, 2018; Nowak & Gomes, 2014). Women are also less prone to revealing their gender on social media profile images than men (Zheng et al., 2016). So, while many people intentionally avoid gender stereotypes in online communication, and gender stereotypes remain a poor predictor of accurate gender inferences, gender stereotypes are nonetheless highly influential in promoting gender inferences.

Androcentric Bias

People tend to possess an androcentric, or male-as-default, bias, in which men are viewed as the norm and women are viewed as an exception or

deviation (Bem, 1993; Stahlberg et al., 2007). Androcentric biases largely predict gender inferences, in which people are more inclined to perceive an unknown other as male rather than female. The clearest examples of this come from research in digital communication, which show that social media profiles with neutral or default icons are perceived as male at roughly twice the rate as female perceptions (Bailey & Lafrance, 2016). Additionally, when usernames contain no stereotypically gendered information, users are typically inferred as male (Karniol et al., 2016; Lambdin et al., 2003).

Androcentric biases also predict the gender inferences of people with gender-ambiguous names. Job applications with gender-ambiguous names are evaluated more similarly to those with masculine rather than feminine names (McKelvie & Waterhouse, 2005; Salvaggio et al., 2009), while authors adopting gender-ambiguous pseudonyms are similarly more likely to be perceived as male (Denham, 2015; Laird, 2003). While men and women both display androcentric biases in gender inferences, these biases tend to be stronger for men (Bailey & LaFrance, 2016) and in people with higher endorsement of hostile sexism (Parks, 2004; Salvaggio et al., 2009).

The androcentrism of gender inferences likely stems from the automatic nature of how inferences form, and how implicit biases can impact automatic judgments. Work in implicit bias has demonstrated that gender-neutral words such as "humanity" or "person" are associated more with men than with women (Bailey et al., 2020). Likewise, intersectionality research demonstrates that if people are asked to picture a person from a subordinate group (e.g., an ethnic minority vs. an ethnic majority), they will default to prototypical features on other identity dimensions (Purdie-Vaughns & Eibach, 2008). So if asked to picture a person who is black, people are more inclined to picture a black man than a black woman. Interestingly, the tendency towards androcentric gender inferences is not innate, but seems to occur over time (Lei et al., 2021). Young children are more inclined to infer a person as their own gender, while girls become more likely to make male inferences as they age. This finding suggests that gender inferences may be biased towards a person's own gender identity but is able to shift if new biases are learned.

Effects of Gender Inferences

Gender inferences do not occur in a vacuum. Once made, these inferences prompt other events, including perceivers' behavior, interpretation of actions, and the use of gender stereotypes (e.g., Ellemers, 2018). This line of research offers a new avenue of exploring how stereotypes of different genders can originate from the same initial stimuli, and how effects of gender stereotypes can interact with effects of one's own gender and gendered experiences. Likewise, we explore how the resulting actions of gender inferences affect not only the target whose gender is inferred, but the one making the inference as well.

Promotion of Stereotypes

Once a gender inference is made, stereotypes associated with that gender become cognitively activated and bind to the target of the inference (Banaji & Hardin, 1996; Stangor, 1988). While the effects of gender stereotypes can be moderated by other traits (e.g., prejudice; Devine, 1989), the activation of stereotypes generally guide how a target is evaluated. For example, gender inferences can affect how much a person is liked. Past research demonstrates that agreeableness and likeability is stereotypically associated with women (Roberts & Norris, 2016). Research in electronic communication demonstrates that people perceived as women are often more likeable than those perceived as men (Eyssel & Hegel, 2012), an effect moderated by the gender of the perceiver (Spottswood et al., 2013). How men infer the gender of a speaker impacts how likable and effective men think that person is. When men are perceiving others in a platonic online chat environment, targets that are assumed to be male are judged as more likable and more effective than those assumed to be female. Researchers have speculated that men may see effective female speakers as a threat to masculinity (Willemsen et al., 2012). Researchers have also noted that men may view those most similar to them as more capable (Willemsen et al., 2012). However, findings show that women evaluate targets similarly regardless of whether they assume targets are male or female, supporting the possibility that masculinity threat may explain variation in men's evaluations (Spottswood et al., 2013). Regardless, both men and women tend to view others as more intelligent and more agentic if they perceive them to be men, in line with gender stereotypes surrounding competency and agency (Eyssel & Hegel, 2012). Likewise, when playing video games, players inferred as female are seen as less intelligent, particularly when their appearance is sexualized (Behm-Morawitz & Mastro, 2009). This effect is found for both men and women and can be internalized for female players. It is likely that the same androcentric bias that can lead to more male gender inferences also contributes to people who are perceived as men being judged as more competent (Parks, 2004; Salvaggio et al., 2009).

Furthermore, gender inferences can also activate stereotypes which impact a person's actions, not just one's judgments. For example, research in employment demonstrates that gender inferences can ultimately result in serious financial consequences for an employee through both the hiring process and through work interactions. Gender inferences are commonly made in the first steps of the hiring process. When an employer receives a resume, the gender of the applicant is inferred through their name (Salvaggio et al., 2009). When applicants with ambiguous names are perceived to be men, they are more likely to be rated favorably and receive an interview (McKelvie & Waterhouse, 2005; Salvaggio et al., 2009). However, when ambiguous names are perceived as female, applicants are likely to be rated less favorably, even less favorably than those with unambiguously female names (McKelvie & Waterhouse, 2005). In this case, the gender inference is priming gender

stereotypes about men and women in the workplace, in which men are typically seen as more competent than women (Foley & Williamson, 2018; Koch et al., 2015). This same stereotype is also activated by gender inferences made during an employee's day-to-day tasks. A case study conducted by author Catherine Nichols found that her male pseudonym was eight times more successful than her given name, even when submitting the same material to the same publishers (Denham, 2015). Wider reports show that across 62 different publishing platforms, male names are both published more often and are the recipient of more prestigious awards than female authorial names (King & Clark, 2019).

Since people are unlikely able to control a person's gender stereotypes, they may seek to instead influence one's initial gender inference. This behavior is very prevalent in online gaming environments (Chou et al., 2017; Yee et al., 2011). A common gender stereotype in online gaming is that women are less skilled players and in need of more help (Waddell & Ivory, 2015). Indeed, gaming research examining gender inferences demonstrates that female avatars are given more in-game goods and help than male avatars, even when controlling for the player's actual gender (Chou et al., 2017; Ducheneaut et al., 2009; Yee et al., 2011; Yee, 2014). While many players report being very aware of others intentionally manipulating how their gender is inferred from their avatar, awareness alone does not seem to prevent stereotypes forming from a player's gender inference (Yee, 2014). Additionally, male players using female avatars are given more items than female players using male avatars, suggesting that gender inferences and their resulting stereotypes may ultimately play a larger role in than a player's actual gender (Waddell & Ivory, 2015). Since it is difficult to change gender stereotypes, people may instead attempt to change gender inferences in order to receive their desired outcome or treatment.

Bidirectional of Effects

We have focused mostly on how causes gendered inference and evaluative consequences for the target of these inferences. However, effects of gender inferences are bidirectional, affecting not only the target of inference, but also the perceiver who made the inference. For example, research demonstrates that gender inferences can affect a perceiver's performance in an online negotiation task (Arndt & Henderson, 2021b). Using a dyadic study design, we found that participants who perceived their negotiation partner as female performed better in the negotiation, regardless of the target's actual gender. Furthermore, being perceived as either male or female did not affect personal negotiation performance, suggesting that any performance effects occurred only from the process of making the inference.

Additional perceiver effects can be found when examining gender inferences made of nonhumans. By definition, an object or robot cannot change based on external inferences. Perceiving a vehicle to be male or

female can do nothing to the properties of the vehicle itself. Instead, any significant effects rising from this type of gender inference can only affect the perceiver. For example, people's interactions with machines vary greatly by the gender inferences people form (Beldad et al., 2016; Kuchenbrandt et al., 2014; Tay et al., 2014). When taking instructions from robots, women perform equally as well whether they perceive a robot as male or female, while men perform much better when they perceive robots as male versus female (Kuchenbrandt et al., 2014). Meanwhile, people are more likely to buy more items from a virtual sales assistant (an artificial intelligence chatbot) when they perceived the sales assistant as gender-congruent to the product being sold, even when the nonhuman status of the sales assistant was made apparent (Beldad et al., 2016). This same pattern applies to physical objects, in which sales increase when brand logos are perceived as gender-congruent to the products they sell (Grohmann, 2009; Lieven et al., 2015).

On a more extreme measure, people's behavior based on the gender inferences of an object can have devastating consequences when it comes to the perceived gender of natural disasters. Research shows that storms with feminine names result in more fatalities than storms with masculine names, as people perceive them as less dangerous and take fewer safety precautions (Jung et al., 2014). While these findings have been debated (Malter, 2014), the gendering of natural disasters nonetheless remains a vivid demonstration of how an individual's behavior can change based on their gender inference of an inherently genderless thing.

Implications

Delving into how gender inferences are made in uncertain and ambiguous situations accomplishes two goals. First, this line of research addresses a commonly overlooked aspect of gender as a theoretical concept, raising new theoretical and practical questions about the cognitive formation of gender perceptions. Second, this work expands the scope of inference research on other external factors typically thought to be readily apparent. Together, we hope this work opens new avenues of future research on both topics.

For Gender

Modern understandings of gender conceptualize gender, at least in part, as a social, performative action. In 1949, Simone de Beauvoir famously penned the phrase, "One is not born, but rather becomes, a woman," suggesting that a person's attitudes and experiences are an essential component of gender. Since then, many researchers have adopted the idea that gender is formed and communicated through one's actions. In their work *Doing Gender*, West and Zimmerman lay the foundation for this very theory (1987, 2009). The two argue that gender exists within the framework of the actions people perform, from the way they dress to their posture, vocal tone, or way of

moving. More recent models of gender typically view the performative notions of gender as gender expression, which is well-regarded in psychology as an essential component of gender (Hyde et al., 2019; Tate et al., 2014). Research on remote communication in particular has explored how gender is a continual performance and can be performed in a wider variety of ways than traditional, face-to-face communication (Shapiro, 2015).

However, if gender is a performance, who then is the audience, and how are they reacting to the performance? Gender inferences serve as a measure of an audience's response to performative gender. When researchers measure the accuracy of gender inferences, they measure if the audience understands or agrees with the performance. When they measure predictors of gender inferences, they break down the nuances of the gender performance, seeing which behaviors communicate which information. Finally, when researchers measure the effects of gender inferences, they measure the interpretation and overall reaction to the gender performance.

While there is a long and rich history of viewing gender as a performance, the idea of researching the audience's reaction to that performance has long been understudied. It is our hope that in examining how gender is inferred, we can begin to understand the underpinnings of an audience's reaction to gender performances. While gender ambiguity is popular in online communication, gender ambiguity also occurs in face-to-face interactions through androgyny or unfamiliar gender identities. Assuming that gender inferences are always certain or accurate greatly limits our understandings of the effects of gender in communication. By expanding the scope of gender inferences, we can create a more holistic model of gender communication—one that looks at both performer and audience.

For External Categories

When it comes to making inferences, the majority of past research has focused on internal attributes, such as personality traits, beliefs, goals, and values (Uleman et al., 2008). More external traits, such as race or gender, are often thought to be too explicit to form unintentional, automatic impressions about. Instead, a person would automatically perform a feature-matching analysis from physical attributes to sort a person into a definite category (Bruner, 1957; Pendry & Macrae, 1996). However, research on gender inferences demonstrates that there are many situations in which external traits such as gender are not explicitly known or stated, be it through ambiguity, omission, or masking. This ambiguity creates an environment for more complicated inferences to form around external traits, which may then further influence other judgments or behaviors beyond a target's true group categorization.

If ambiguity breeds gender inferences, it stands to reason that ambiguity could breed inferences of other identity variables, such as race, age, or sexuality. While many people hold the belief that these variables are explicit

or obvious (Cosmides et al., 2003; Fazio & Dunton, 1997), research suggests that identity variables are more subjective than people believe (Young et al., 2013). Additionally, it is important to identify where inferences on identity are arising automatically and how they are affecting future inferences and behaviors. For example, will learning about a person's profession promote an inference about their race? If so, how does that affect future interaction? Examining how inferences form around identity or category variables under ambiguity, looking at automaticity, pervasiveness, and generativity, can reveal much about how these identities shape communication from both a perceiver and target's perspective.

In researching inferences about external categories, researchers uncover findings which can challenge and change our conceptualization of the very categories themselves. For example, research in race-based inferences found that traits associated with black ethnic groups can become bound to white faces (Blair et al., 2002). Both white and black faces that contained more Afrocentric features were rated higher in traits associated with stereotypes of black Americans. Furthermore, while people can typically become aware and even correct for inferences that arise due to a person's race, people are unaware of and not able to correct race-based inferences that arise due to a person's facial features (Blair et al., 2004). This line of research on racial inferences revealed that race and racial discrimination were much more complex concepts than previously believed. Examining the inference of external traits and the effects these inferences have expand our conceptual understandings and open new lines of research into both traits and the scope of inferences themselves.

Together, this work suggests that the trait categorization process is not as straightforward as once believed. External traits such as gender are frequently ambiguous and require individuals to infer a person's category or group membership. It is the inference itself that then promotes stereotypes which influence other judgments and actions. These stereotypes, judgments, and actions arise separately from the category or group to which a target actually belongs. Identifying where discrepancies can occur between inferences and category belonging and examining how these inferences come about can reveal new information about the traits themselves or change how certain traits have been conceptualized.

References

Akbari, Y., Nouri, K., Sadri, J., Djeddi, C., & Siddiqi, I. (2017). Wavelet-based gender detection on off-line handwritten documents using probabilistic finite state automata. *Image and Vision Computing*, *59*, 17–30.

Amihai, I., Deouell, L., & Bentin, S. (2011). Conscious awareness is necessary for processing race and gender information from faces. *Consciousness and Cognition*, *20*(2), 269–279.

Arndt, A., & Henderson, M. D. (2021a). Reading between the lines: Perceptions of gender in gender-fair language. Unpublished manuscript.

Arndt, A., & Henderson, M. D. (2021b). The role of perceived gender during integrative negotiations. Unpublished manuscript.

Bailey, A. H., & LaFrance, M. (2016). Anonymously male: Social media avatar icons are implicitly male and resistant to change. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, *10*(4). 10.5817/CP2016-4-8

Bailey, A. H., LaFrance, M., & Dovidio, J. F. (2020). Implicit androcentrism: Men are human, women are gendered. *Journal of Experimental Social Psychology*, *89*, 103980.

Banaji, M. R., & Hardin, C. D. (1996). Automatic stereotyping. *Psychological Science*, *7*(3), 136–141.

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. Wyer Jr., & T. K. Srull (Eds.), *Handbook of Social Cognition*. (pp. 1–40). Hillsdale, NJ: Erlbaum.

Behm-Morawitz, E., & Mastro, D. (2009). The effects of the sexualization of female video game characters on gender stereotyping and female self-concept. *Sex Roles*, *61*(11–12), 808–823.

Beldad, A., Hegner, S., & Hoppen, J. (2016). The effect of virtual sales agent (VSA) gender–product gender congruence on product advice credibility, trust in VSA and online vendor, and purchase intention. *Computers in Human Behavior*, *60*, 62–72.

Belle, D., Tartarilla, A. B., Wapman, M., Schlieber, M., & Mercurio, A. E. (2021). "I Can't Operate, that Boy Is my Son!": Gender schemas and a classic riddle. *Sex Roles*, *85*(3–4), 161–171. 10.1007/s11199-020-01211-4

Bem, S. L. (1993). *The lenses of gender: Transforming the debate on sexual inequality*. Yale University Press.

Bivens, R., & Haimson, O. L. (2016). Baking gender into social media design: How platforms shape categories for users and advertisers. *Social Media+ Society*, *2*(4), 1–12.

Blair, I. V., Judd, C. M., & Fallman, J. L. (2004). The automaticity of race and Afrocentric facial features in social judgments. *Journal of Personality and Social Psychology*, *87*(6), 763.

Blair, I. V., Judd, C. M., Sadler, M. S., & Jenkins, C. (2002). The role of Afrocentric features in person perception: Judging by features and categories. *Journal of Personality and Social Psychology*, *83*(1), 5.

Bouadjenek, N., Nemmour, H., & Chibani, Y. (2014). Local descriptors to improve off-line handwriting-based gender prediction. *2014 6th International Conference of Soft Computing and Pattern Recognition (SoCPaR)*, (pp. 43–47). IEEE.

Bruner, J. S. (1957). On perceptual readiness. *Psychological Review*, *64*(2), 123.

Brewer, M. B. (1988). A dual process model of impression formation. In T. K. Srull , & R. S. Wyer Jr. (Eds.), *A dual process model of impression formation*, (pp. 1–36). Lawrence Erlbaum Associates, Inc.

Burr, V. (2002). Judging gender from samples of adult handwriting: Accuracy and use of cues. *The Journal of Social Psychology*, *142*(6), 691–700.

Burr, V. (2006). The art of writing: Embodiment and pre–verbal construing. In P. Caputi , H. Foster , & L.L. Viney (Eds.), *Personal Construct Psychology: New Ideas*, (pp. 317–322). John Wiley & Sons Ltd. 10.1002/9780470713044.ch24

Chen, J. M., Banerji, I., Moons, W. G., & Sherman, J. W. (2014). Spontaneous social role inferences. *Journal of Experimental Social Psychology*, *55*, 146–153.

Chou, Y.-J., Lo, S.-K., & Teng, C.-I. (2017). Reasons for avatar gender swapping by online game players: A qualitative interview-based study. In Information Resources Management Association (Ed.), *Discrimination and diversity: Concepts, methodologies, tools, and applications*. (pp. 202–219). IGI Global.

Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. J. (2005). Separating multiple processes in implicit social cognition: The quad model of implicit task performance. *Journal of Personality and Social Psychology*, *89*(4), 469.

Cornetto, K. M., & Nowak, K. L. (2006). Utilizing usernames for sex categorization in computer-mediated communication: Examining perceptions and accuracy. *Cyber Psychology & Behavior*, *9*(4), 377–387.

Cosmides, L., Tooby, J., & Kurzban, R. (2003, April). Perceptions of race. *TRENDS in Cognitive Sciences*, *7*(4), 173–179.

Danet, B. (1988). Text as mask: Gender and identity on the internet. In S. Jones (Ed.), *Cybersociety 2.0*, (pp. 129–158). Thousand Oaks, CA: Sage.

Denham, J. (2015, August 6). *Writing under a male name makes you eight times more likely to get published, one female author finds | The Independent*. Independent. https://www.independent.co.uk/arts-entertainment/books/news/writing-under-a-male-name-makes-you-eight-times-more-likely-to-get-published-one-female-author-finds-10443351.html

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*(1), 5.

Ducheneaut, N., Wen, M.-H., Yee, N., & Wadley, G. (2009). Body and mind: A study of avatar personalization in three virtual worlds. In *Proceedings of the 27th International Conference on Human Factors in Computing Systems (CHI '09)*. New York: ACM Press, 1151–1160.

Eden, A., Maloney, E., & Bowman, N. D. (2010). Gender attribution in online video games. *Journal of Media Psychology: Theories, Methods & Applications*, *22*(3), 114–124.

Ellemers, N. (2018). Gender stereotypes. *Annual Review of Psychology*, *69*, 275–298.

Elsbach, K. D., Cable, D. M., & Sherman, J. W. (2010). How passive 'face time'affects perceptions of employees: Evidence of spontaneous trait inference. *Human Relations*, *63*(6), 735–760.

Eyssel, F., & Hegel, F. (2012). (S) He's got the look: Gender stereotyping of robots 1. *Journal of Applied Social Psychology*, *42*(9), 2213–2230.

Fazio, R. H., & Dunton, B. C. (1997). Categorization by race: The impact of automatic and controlled components of racial prejudice. *Journal of Experimental Social Psychology*, *33*(5), 451–470.

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In *Advances in experimental social psychology*, (Vol. 23, pp. 1–74). Academic Press.

Foley, M., & Williamson, S. (2018). Does anonymising job applications reduce gender bias? Understanding managers' perspectives. *Gender in Management: An International Journal*, *33*(8), 623–635.

Fong, K., & Mar, R. A. (2015). What does my avatar say about me? Inferring personality from avatars. *Personality and Social Psychology Bulletin*, *41*(2), 237–249.

Frankenstein, A. N., McCurdy, M. P., Sklenar, A. M., Pandya, R., Szpunar, K. K., & Leshikar, E. D. (2020). Future thinking about social targets: The influence of prediction outcome on memory. *Cognition*, *204*, 104390.

Grohmann, B. (2009). Gender dimensions of brand personality. *Journal of Marketing Research*, *46*(1), 105–119.

Hagström, C. (2012). Naming me, naming you. personal names, online signatures and cultural meaning. *Oslo Studies in Language*, *4*(2), Article 2.

Ham, J., & Vonk, R. (2003). Smart and easy: Co-occurring activation of spontaneous trait inferences and spontaneous situational inferences. *Journal of Experimental Social Psychology*, *39*(5), 434–447.

Hartley, J. (1991). Sex differences in handwriting: A comment on Spear. *British Educational Research Journal*, *17*(2), 141.

Hayes, W. N. (1996). Identifying sex from handwriting. *Perceptual and Motor Skills*, *83*(3), 791–800.

Heisler, J. M., & Crabill, S. L. (2006). Who are "stinkybug" and "Packerfan4"? Email pseudonyms and participants' perceptions of demography, productivity, and personality. *Journal of Computer-Mediated Communication*, *12*(1), 114–135.

Herring, S. C., & Martinson, A. (2004). Assessing gender authenticity in computer-mediated language use: Evidence from an identity game. *Journal of Language and Social Psychology*, *23*(4), 424–446.

Herring, S. C., & Paolillo, J. C. (2006). Gender and genre variation in weblogs. *Journal of Sociolinguistics*, *10*(4), 439–459.

Herring, S., & Stoerger, S. (2014). Gender and (a)nonymity in computer-mediated communication—The handbook of language, gender, and sexuality—Wiley online library. In *The Handbook of Language, Gender, and Sexuality* (2nd ed.). 10.1002/9781118584248.ch29

Hügelschäfer, S., Jaudas, A., & Achtziger, A. (2016). Detecting gender before you know it: How implementation intentions control early gender categorization. *Brain Research*, *1649*, 9–22.

Hyde, J. S., Bigler, R. S., Joel, D., Tate, C. C., & van Anders, S. M. (2019). The future of sex and gender in psychology: Five challenges to the gender binary. *American Psychologist*, *74*(2), 171.

Ito, T. A., & Urland, G. R. (2003). Race and gender on the brain: Electrocortical measures of attention to the race and gender of multiply categorizable individuals. *Journal of Personality and Social Psychology*, *85*(4), 616–626.

Ito, T. A., & Urland, G. R. (2005). The influence of processing objectives on the perception of faces: An ERP study of race and gender perception. *Cognitive, Affective, & Behavioral Neuroscience*, *5*(1), 21–36.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*(5), 513–541.

Jung, K., Shavitt, S., Viswanathan, M., & Hilbe, J. M. (2014). Female hurricanes are deadlier than male hurricanes. *Proceedings of the National Academy of Sciences*, *111*(24), 8782–8787.

Jung, K. H., White, K. R., & Powanda, S. J. (2019). Automaticity of gender categorization: A test of the efficiency feature. *Social Cognition*, *37*(2), 122–144.

Karniol, R., Artzi, S., & Ludmer, M. (2016). Children's production of subject–verb agreement in hebrew when gender and context are ambiguous. *Journal of Psycholinguistic Research*, *45*(6), 1515–1532. 10.1007/s10936-016-9419-1

King, A., & Clark, S. (2019). The 2018 VIDA count. *VIDA. Women in Literary Arts*, 17.

Koch, S. C., Mueller, B., Kruse, L., & Zumbach, J. (2005). Constructing gender in chat groups. *Sex Roles*, *53*(1–2), 29–41.

Koch, A. J., D'Mello, S. D., & Sackett, P. R. (2015). A meta-analysis of gender stereotypes and bias in experimental simulations of employment decision making. *Journal of Applied Psychology*, *100*(1), 128.

Kuchenbrandt, D., Häring, M., Eichberg, J., Eyssel, F., & André, E. (2014). Keep an eye on the task! How gender typicality of tasks influence human–robot interactions. *International Journal of Social Robotics*, *6*(3), 417–427.

Laird, H. A. (2003). The coauthored pseudonym: Two women named Michael Field. In R. J. Griffin (Ed.), *The Faces of Anonymity: Anonymous and Pseudonymous Publication from the Sixteenth to the Twentieth Century* (pp. 193–209). Palgrave Macmillan US. 10.1007/978-1-137-11109-8_9

Lambdin, J. R., Greer, K. M., Jibotian, K. S., Wood, K. R., & Hamilton, M. C. (2003). The animal = male hypothesis: Children's and adults' beliefs about the sex of non–sex-specific stuffed animals. *Sex Roles*, *48*(11–12), 471–482.

Lee, K. M., Liao, K., & Ryu, S. (2007). Children's responses to computer-synthesized speech in educational media: Gender consistency and gender similarity effects. *Human Communication Research*, *33*(3), 310–329.

Lei, R., Leshin, R., Moty, K., Foster-Hanson, E., & Rhodes, M. (2021). How race and gender shape the development of social prototypes in the United States.*Journal of Experimental Psychology: General*. Advance online publication. 10.1037/xge0001164

Levordashka, A., & Utz, S. (2017). Spontaneous Trait inferences on social media. *Social Psychological and Personality Science*, *8*(1), 93–101. 10.1177/1948550616663803

Lewis, J. (2014). *Forensic document examination: Fundamentals and current trends*. Elsevier.

Lieven, T., Grohmann, B., Herrmann, A., Landwehr, J. R., & Van Tilburg, M. (2015). The effect of brand design on brand gender perceptions and brand preference. *European Journal of Marketing*.

Malter, D. (2014). Female hurricanes are not deadlier than male hurricanes. *Proceedings of the National Academy of Sciences*, *111*(34), E3496–E3496.

McKelvie, S. J., & Waterhouse, K. (2005). Impressions of people with gender-ambiguous male or female first names. *Perceptual and Motor Skills*, *101*(2), 339–344.

Morera, Á., Sánchez, Á., Vélez, J. F., & Moreno, A. B. (2018). Gender and handedness prediction from offline handwriting using convolutional neural networks. *Complexity*, *2018*. 10.1155/2018/3891624

Mou, Y., Xu, K., & Xia, K. (2019). Unpacking the black box: Examining the (de)Gender categorization effect in human-machine communication. *Computers in Human Behavior*, *90*, 380–387. doi: 10.1016/j.chb.2018.08.049

Nowak, K. L., & Fox, J. (2018). Avatars and computer-mediated communication: A review of the definitions, uses, and effects of digital representations. *Review of Communication Research*, *6*, 30–53.

Nowak, K. L., & Gomes, S. B. (2014). The choices people make: The types of buddy icons people select for self-presentation online. *AI & Society*, *29*(4), 485–495.

Oberst, U., Renau, V., Chamarro, A., & Carbonell, X. (2016). Gender stereotypes in Facebook profiles: Are women more female online? *Computers in Human Behavior*, *60*, 559–564.

Parks, J. B. (2004). Attitudes toward women mediate the gender effect on attitudes toward sexist language. *Psychology of Women Quarterly*, *28*(3), 233–239.

Pelletier, L. (2009). You've got mail: Identity perceptions based on email usernames. *Journal of Undergraduate Research at Minnesota State University, Mankato*, 9(1), 13.

Pendry, L. F., & Macrae, C. N. (1996). What the disinterested perceiver overlooks: Goal-directed social categorization. *Personality and Social Psychology Bulletin*, *22*(3), 249–256.

Purdie-Vaughns, V., & Eibach, R. P. (2008). Intersectional invisibility: The distinctive advantages and disadvantages of multiple subordinate-group identities. *Sex Roles*, 59(5–6), 377–391.

Roberts, F., & Norris, A. (2016). Gendered expectations for "agreeableness" in response to requests and opinions. *Communication Research Reports*, *33*(1), 16–23.

Salvaggio, A. N., Streich, M., & Hopper, J. E. (2009). Ambivalent sexism and applicant evaluations: Effects on ambiguous applicants. *Sex Roles*, *61*(9–10), 621.

Savicki, V., Kelley, M., & Oesterreich, E. (1999). Judgments of gender in computer-mediated communication. *Computers in Human Behavior*, *15*(2), 185–194. 10.1016/S0747-5632(99)00017-5

Shapiro, E. (2015). *Gender circuits: Bodies and identities in a technological age*. Routledge.

Smith, E. R., & Miller, F. D. (1983). Mediation among attributional inferences and comprehension processes: Initial findings and a general method. *Journal of Personality and Social Psychology*, *44*(3), 492.

Spottswood, E. L., Walther, J. B., Holmstrom, A. J., & Ellison, N. B. (2013). Person-centered emotional support and gender attributions in computer-mediated communication. *Human Communication Research*, *39*(3), 295–316. 10.1111/hcre.12006

Sprouse, J. L., & Webb, J. E. (1994). *The Pygmalion Effect and Its Influence on the Grading and Gender Assignment on Spelling and Essay Assessments*.

Stahlberg, D., Braun, F., Irmen, L., & Sczesny, S. (2007). Representation of the sexes in language. In K. Fiedler (Ed.), *Social Communication*, (pp. 163–187). New York: Psychology.

Stangor, C. (1988). Stereotype accessibility and information processing. *Personality and Social Psychology Bulletin*, *14*(4), 694–708.

Tate, C. C., Youssef, C. P., & Bettergarcia, J. N. (2014). Integrating the study of transgender spectrum and cisgender experiences of self-categorization from a personality perspective. *Review of General Psychology*, *18*(4), 302–312.

Tay, B., Jung, Y., & Park, T. (2014). When stereotypes meet robots: the double-edge sword of robot gender and personality in human–robot interaction. *Computers in Human Behavior*, *38*, 75–84.

Thomson, R., & Murachver, T. (2001). Predicting gender from electronic discourse. *British Journal of Social Psychology*, *40*(2), 193–208.

Todd, A. R., Molden, D. C., Ham, J., & Vonk, R. (2011). The automatic and co-occurring activation of multiple social inferences. *Journal of Experimental Social Psychology*, *47*(1), 37–49.

Todorov, A., & Uleman, J. S. (2002). Spontaneous trait inferences are bound to actors' faces: Evidence from a false recognition paradigm. *Journal of Personality and Social Psychology*, *83*(5), 1051.

Todorov, A., & Uleman, J. S. (2003). The efficiency of binding spontaneous trait inferences to actors' faces. *Journal of Experimental Social Psychology*, *39*(6), 549–562. 10.1016/S0022-1031(03)00059-3

Topaloglu, M., & Ekmekci, S. (2017). Gender detection and identifying one's handwriting with handwriting analysis. *Expert Systems with Applications*, *79*, 236–243. 10.1016/j.eswa.2017.03.001

Uleman, J. S., Adil Saribay, S., & Gonzalez, C. M. (2008). Spontaneous inferences, implicit impressions, and implicit theories. *Annual Review of Psychology*, *59*, 329–360.

van Doorn, N., van Zoonen, L., & Wyatt, S. (2007). Writing from experience: Presentations of gender identity on weblogs. *European Journal of Women's Studies*, *14*(2), 143–158. 10.1177/1350506807075819

Waddell, T. F., & Ivory, J. D. (2015). It's not easy trying to be one of the guys: The effect of avatar attractiveness, avatar sex, and user sex on the success of help-seeking requests in an online game. *Journal of Broadcasting & Electronic Media*, *59*(1), 112–129.

West, C., & Zimmerman, D. H. (1987). Doing gender. *Gender & Society*, *1*(2), 125–151. 10.1177/0891243287001002002

West, C., & Zimmerman, D. H. (2009). Accounting for doing gender. *Gender & Society*, *23*(1), 112–122.

Willemsen, L. M., Neijens, P. C., & Bronner, F. (2012). The ironic effect of source identification on the perceived credibility of online product reviewers. *Journal of Computer-Mediated Communication*, *18*(1), 16–31.

Winter, L., Uleman, J. S., & Cunniff, C. (1985). How automatic are social judgments? *Journal of Personality and Social Psychology*, *49*(4), 904.

Wyer Jr, R. S., & Srull, T. K. (1988). *Advances in social cognition, Volume I: A dual process model of impression formation*. Psychology Press.

Yan, X., Wang, M., & Zhang, Q. (2012). Effects of gender stereotypes on spontaneous trait inferences and the moderating role of gender schematicity: Evidence from Chinese undergraduates. *Social Cognition*, *30*(2), 220–231.

Yee, N. (2014). *The Proteus paradox: How online games and virtual worlds change us-and how they don't*. Yale University Press.

Yee, N., Ducheneaut, N., Yao, M., & Nelson, L. (2011). Do men heal more when in drag? Conflicting identity cues between user and avatar. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 773–776. 10.1145/1978942.1979054

Young, D. M., Sanchez, D. T., & Wilton, L. S. (2013). At the crossroads of race: Racial ambiguity and biracial identification influence psychological essentialist thinking. *Cultural Diversity and Ethnic Minority Psychology*, *19*(4), 461–467. 10.1037/a0032565

Zheng, W., Yuan, C.-H., Chang, W.-H., & Wu, Y.-C. J. (2016). Profile pictures on social media: Gender and regional differences. *Computers in Human Behavior*, *63*, 891–898. 10.1016/j.chb.2016.06.041

Zosuls, K. M., Miller, C. F., Ruble, D. N., Martin, C. L., & Fabes, R. A. (2011). Gender development research in sex roles: Historical trends and future directions. *Sex Roles*, *64*(11), 826–842.

5 Differences between Spontaneous and Intentional Trait Inferences

James S. Uleman
New York University

Most of our thinking (i.e., cognizing, processing information) takes place without our awareness. This chapter is about how this unconscious thought affects our impressions of others and interacts with our conscious thought. In particular, it focuses on differences between conscious and unconscious thought and how they may interact, as shown in research on spontaneous trait inferences (STIs). Broadly, unconscious thought is more bottom-up and data driven, whereas conscious thought is more top-down, socially driven, and exemplified by the language that lets us share our thoughts with others. These are not new ideas, but the research that illustrates and supports them in this chapter is. The larger question which hovers over this work and which lies beyond the scope of this chapter is how we turn sensory data into meaning, represented for adults in most cases by language (one of our better understood conscious representational systems). Mere associations or evaluations are only a small part of the story. Our impressions of others are richer than that.

This is not a chapter about activating stored concepts or evaluations and associating them with other people. Such priming and evaluative conditioning processes do affect our impressions of others (e.g., Higgins, 1996), but they do not involve inferences about people; they do not take pieces of information, like words that are not traits or related concepts, and combine them to compute an emergent meaning like a behavioral description that implies a trait, which can then be later combined with a representation of an actor (Todorov & Uleman, 2002). This is also not a chapter about how words in a sentence, or silhouettes depicting actions (Fiedler & Schenck, 2001), are used to infer trait-implying behaviors' trait categories, even though linguistic and behavior categorization processes are central to producing STIs. Rather, in this chapter I take STI for granted and focus on evidence that their effects diverge from those of intentional trait inferences. Finally, this chapter is not about implicit versus explicit evaluations.

As we shall see, the literature on STIs over the past 40 years shows how to experimentally create unconscious thoughts about others, through incidental exposure to trait implying information. The same information can also be used to produce conscious thoughts by asking for impressions that are

DOI: 10.4324/9781003045687-6

measured through self-reports about the same targets. Discrepancies can arise between unconscious and conscious thoughts. All such instances known to me are reviewed below. Then the challenge is to formulate and test hypotheses about relationships between unconscious and conscious impressions of the same targets based on the same information. This chapter offers some preliminary suggestions about how they may be related and presents research that tests of one of these suggestions. Future research by others should offer more.

"Incidental" exposure to trait implying information is critical in producing STIs because it provides enough attention to the material to enable inferences about it, yet not such focused attention as to create clear awareness and memories of the inferences. Thus, the claim is made that participants in these studies are not "aware" of their inferences, because they do not have explicit memories of them when queried. (More research is needed to clarify relations between instructed or self-initiated attention, processing goals, and awareness at various time delays after making such inferences.)

It has been said that "thinking is for doing" (e.g., Fiske, 1992). But spontaneous inferences seem not to be for doing anything; they simply occur unintentionally and without the perceiver's awareness. They can be affected by the perceiver's goals (e.g., Uleman & Moskowitz, 1994), even unconsciously primed goals (Dijksterhuis & Aarts, 2010; Rim et al., 2013). However, they do not seem to be purposive, goal directed, or functional in any immediate sense that self-reports can reveal. (None of this precludes theorists attributing functions or purposes to STIs, but that's another matter.) Instead, we should say that "thinking does." It does produce inferences that may have multiple downstream consequences, depending on the situation. My interest is in *how* this occurs, mechanistically, not in *why* it occurs teleologically. (Although top-down goals affect social inferences, theorists attributing motives or goals to cognition is both dangerous and slippery. It is dangerous because it can give the appearance of explaining cognitions without really doing so. Attributing motives or goals can also be circular unless the motive is measured or manipulated at that time, shows clear relationships with the outcomes, and is shown to function as a motive as, for example, demonstrated in Rim et al. (2013) where satiation leads to cessation. It is slippery because it can insert a homunculus into the explanatory machinery without acknowledgment.)

Divergent Processes in Spontaneous and Intentional Trait Inferences

What is a spontaneous trait inference (STI)? Here is an example. When participants read that "He carried the old woman's groceries across the street" ostensibly within the context of a "memory test" or to "familiarize" themselves with the research materials, they infer that *He* is *helpful* even though they are unaware of making this inference. Attention to the sentence is

intentional but the inference is not, and the process of forming the inference usually remains unconscious. STI is a robust phenomenon, having produced a moderate to strong effect size of $d_z = 0.59$ in a meta-analysis over 97 studies with a total of 14,387 participants (Bott et al., 2021). The process of forming spontaneous trait inferences is moderated by multiple factors and produces many consequences (Uleman et al., 2012; Uleman et al., 2008). Bott et al. (2021) found that effect sizes for STIs varied by paradigm and were largest for false recognition (Todorov & Uleman, 2002) and savings-in-relearning (Carlston & Skowronski, 1994) $-d_z = 0.75$, $k = 69$, and $d_z = 0.62$, $k = 32$, respectively.

The memory probes or cues that are used in these two STI paradigms are based on pretesting for intentional inferences. Study participants rated the grocery carrier as helpful. Only sentences with high consensual trait implications are used in these studies. Nevertheless, brain activity differs for spontaneous and intentional trait inferences. Ma et al. (2011) used fMRI and found that STIs activated central mentalizing areas, the temporal-parietal junction and medial prefrontal cortex. Intentional trait inferences activated these and additional areas, suggesting "that intentional instructions invite observers to think more about the material they read…" (p. 123).

Other effects of these consensual inferences can differ depending on whether they are spontaneous or intended. Moskowitz and Roman (1992) had participants form impressions from trait-implying sentences, thereby producing intentional explicit trait inferences, or they read the sentences for a memory test, thereby producing spontaneous trait inferences. Then they reported their impressions of a different actor enacting ambiguous behaviors, for example, ones that could be interpreted in one of two ways such as *adventurous* or *reckless*. This allowed the prior conscious or unconscious (STIs) trait inferences to function as primes, thereby disambiguating the subsequent impressions intentionally formed from these ambiguous behaviors. Intentional inferences produced contrast effects; spontaneous inferences produced assimilation effects. Thus, conscious and unconscious inferences of the same traits had opposite effects as primes, depending on whether the inferences were conscious or not.

Conscious and unconscious trait inferences have opposite effects on how memories are organized as well. Ferreira et al. (2012) showed participants 24 trait-implying sentences about a single actor, John, unlike in most STI studies with multiple actors and behaviors. Consistent with past research, memory instructions (spontaneous inferences) produced more false recognition of implied traits than impression formation instructions (intended inferences), while the latter produced more clustering in free recall. That is, intentional impression formation organized memories around implied traits but spontaneous impressions did not. Additional studies used a forced choice sentence recognition paradigm which can prompt the retrieval of the trait inference monitoring processes, which researchers believe happens during encoding, even when controlling for memory of other aspects of the

sentence. Spontaneous inferences produced more source monitoring errors (false recognition), while intentional inferences reduced source confusion. Using Jacoby's (1991) process dissociation procedure (PDP), researchers estimated the contributions of controlled (C) and automatic (A) processes to recognition memory performance. As hypothesized, C was higher under intentional than spontaneous instructions, but A did not vary. Cognitive load reduced C, but left A unchanged. The explicit goal of forming impressions seems to activate additional inference and source monitoring processes (Johnson et al. 1993).

Divergent Content in Spontaneous and Intentional Trait Inferences

Even though STI behavioral sentences are selected on the basis of consensual explicit trait inferences, determined through pretesting with independent samples, there is evidence that the content of spontaneous and intentional trait inferences may differ. Zelli and his and colleagues chose participants in extreme quartiles on self-reports of being aggressive within the past year, identifying individuals who threatened, or actually cut another person with a knife, or shot a gun (Zelli et al., 1995). In the spontaneous inference condition, half of them read sentences (for a subsequent memory test) that had both hostile and non-hostile interpretations, e.g., "The electrician looks at his younger brother and starts laughing." In the intentional inference condition, the other half were also asked to think about why each actor did what s/he did. Zelli et al. (1995) found that in the spontaneous (memory only) condition, hostile cues (e.g., *ridicule*) more effectively retrieved hostile participants' memories for the sentences than did semantic associates of important sentence words (e.g., *wires*). But in the intentional condition involving thought about why each action occurred, this difference disappeared. Zelli et al. (1996) showed that this effect was restricted to hostile traits. They concluded that spontaneous processes are more sensitive than deliberative processes to individual differences in constructs' chronic accessibility (Zelli et al., 1995) because deliberative processes activate a wider range of concepts, thereby obscuring the first inferences that come to mind.

Note that "spontaneous trait inferences" occur when participants (intentionally) read trait implying sentences without the goal of inferring traits or anything else that entails trait inferences. Thinking about "why the actor did what s/he did" entails trait inferences, as does thinking about "the actor's personality," so trait inferences under these conditions are intentional and not spontaneous.

Another instance of individual differences in chronic preoccupations affecting STIs was provided by Narvaez et al. (2006). They chose participants from the extreme quartiles of a student population on moral chronicity—how chronically accessible moral traits are. Differences in moral chronicity predicted cued recall performance on a measure of spontaneous moral trait

inferences, but were not related to cued recall following intentional inferences. They also used lexical decision response times (RTs) to assess spontaneous activation of traits during readings. Those with high moral chronicity showed more activation of negative traits (e.g., *disloyal, selfish*) than low participants.

Spontaneous inferences from short stories about social injustice differ from intentional judgments. Ham and van den Bos's (2008) participants read short stories about just and unjust events, varying whether the story actors were high or low in personal relevance. They used the probe recognition paradigm to assess spontaneous activation of justice concepts, in which participants quickly judge whether probe words were explicit in the stories. They also got intentional ratings of the same events (e.g., *fair-unfair*). For example, an unjust high-relevance paragraph said that "You and your colleague do the same work. You make 1,400 euros a month, and your colleague 4,100 euros." Spontaneous inferences of justice concepts occurred most with high-relevance unjust stories, rather than low-relevance stories that reference others only rather than the self ("He and his colleague do the same…"). Explicit justice judgments showed no effects of relevance. Spontaneous inferences were also uncorrelated with explicit judgments in these two within-subjects studies, providing more evidence of a dissociation between spontaneous and intentional inferences. Ham and van den Bos (2011) showed that these effects were distinct from valence effects, assessed specifically through the probe words *positive* and *negative*, *friendly* and *hateful*.

Nevertheless, some research shows that spontaneous inferences are more sensitive to valence than intentional ones. Zhang and Wang (2018) used 12 behavioral sentences as trait implying stimuli. Half had implications that differed on the warm-cold dimension and half had implications on competence-incompetence. Explicit trait ratings showed no effects of either dimension or of moderation by valence on either dimension. ("Valence" here is not comparable to that in Ham and van den Bos, 2011.) But a probe recognition measure of trait activation revealed an interaction. The longest RTs (indicating more activation) occurred for cold traits (*impolite, selfish, indifferent*). In a second experiment, the same effect occurred with the false recognition paradigm, which measures traits' activation and binding to actors. Unlike intentional trait ratings, both spontaneous trait measures showed this "primacy-of-warmth effect." Negative instances on the social warm-cold dimension showed the largest effect. As in most studies cited previously, these effects for spontaneous inferences did not occur for intentional inferences.

What Are People Thinking, Consciously and Not?

These studies raise the question of what people are thinking when they infer traits intentionally, and "think more about the material they read…" (Ma et al., 2011). Are they checking relationships with related knowledge, or remembering similar events in their lives, or wondering what their inferences

reveal about themselves to the psychologist running this study or to another audience, or questioning their initial inference? None of these studies provide answers, but most of the discrepancies suggest some. Zelli et al. (1995, p. 407) suggested that impression formation instructions may change what participants attend to and therefore alter their inferences: "If people who characteristically make hostile inferences are explicitly asked to deliberate upon an encounter, they may consider situational cues that they otherwise would have ignored." Some participants had been asked to consider "why" the behaviors occurred and as such formed intentional inferences. Such "causality instructions may prompt individuals to consider a variety of alternate dispositional characteristics" (Zelli et al., 1996, p. 186). Narvaez et al. (2006) found that intentional inferences reduced the impact of individual differences, in this case in moral chronicity. They speculated that relative to a semantic cuing recall measure, "there were no differences between chronics and non-chronics in the deliberate [intentional] processing condition … because the impression formation instructions also directed non-chronics to attend to dispositional features of characters" rather than to semantic associates (p. 975). Ham and van den Bos (2008, 2011) found that justice concepts were spontaneously activated among all participants only when unjust events were self-relevant, but not when they were about strangers. Yet intentional judgments showed no such difference. They note that "people's spontaneous reactions are influenced by egocentric biases…[whereas] more controlled and explicit reactions to events are less egocentrically biased and more objective" (Ham & van den Bos, 2008, p. 699). Perhaps participants adopted a less egocentric viewpoint when reporting intentional judgments to appear more even-handed and less egocentric. Zhang and Wang's (2018) participants only showed valence effects on spontaneous inferences, not on intentional ones. Perhaps some sort of affect flattening was at work here (as well as in the other studies) when intentional judgments were made, in the service of impression management or appearing "rational." But the possibilities seem endless and are clearly post hoc. How can we make progress on this issue?

First, as an easy beginning, we might generate more post hoc possibilities by getting intentional inferences from sets of sentences that are already known to be sensitive to individual differences in STI. If intentional inferences have different determinants from spontaneous inferences, as in the research reviewed above (Zelli et al., 1995, 1996), this might suggest more post hoc possibilities. Uleman et al. (1986) published such sentences sensitive to differences in authoritarianism, and Crouch et al. (2010) did the same for parents at low and high risk for child abuse. But participants in these studies were not asked for their explicit impressions of the actors in these sentences, so these could not be contrasted with their spontaneous impressions. Such contrasts might suggest possible processing differences.

Second, if one thought that intentional trait inferences are affected by impression management, one could manipulate or measure beliefs about the

presumptive audience for those inferences, and contrast them with STIs under the same conditions. Cognitive tuning (Zajonc, 1960) and audience effects (e.g., Higgins, 1981) are well known but have not been examined with STIs.

Third, one could study sentences with variations known to produce differences in STIs, and gather intentional inferences from them as well. Fortunately, research on effects of stereotypes on STIs has already produced such sentences. For example, Wigboldus et al. (2003) compared STIs from sentences such as "X wins the science quiz," which implies *smart*, with STIs when X is either "the professor" or "the garbage man." Compared with the neutral actor X, the garbage man inhibited inferences of *smart* but the professor had no effect. The extensive literature on effects of stereotypes on STIs (reviewed below) generally confirms this pattern: STIs are inhibited when the stereotype of the actor conflicts with the behavior's implication, but are not enhanced when the actor stereotype is consistent with it. But none of these studies compared STIs with intentional impressions. Biernat's (2003, 2012) shifting standards model for explicitly judging stereotyped actors suggests that these impressions would be different from STIs, and it describes the mechanism that produces these differences.

So the rest of this chapter focuses on this third option. First, there is a review of all the research to date on effects of stereotypes on STIs. This, along with the papers reviewed previously, completes the review of all published research showing discrepancies between intentional and spontaneous impressions. Second, Biernat's shifting standards model is presented in detail. Third, five unpublished studies that show differences between intentional and spontaneous impressions are described. (These studies were done about 20 years ago, and are unpublished because some of the documentation now required for journal publication has been lost. These standards have shifted too.) Finally, some thoughts and speculations about future directions are presented. It seems very unlikely that one or two mechanism, such as shifting standards for stereotyped actors or audience effects or impression management concerns, can account for all the differences between spontaneous and intentional trait inferences noted here. Thus the research below serves as a proof of concept for one such mechanism, not as a way to account for all the discrepancies reviewed here or likely to be uncovered in the future, once future research focuses on this. "More research is needed," as we say. Fortunately there has been some, since I first speculated on how intentional and spontaneous impressions might differ (Uleman, 1989).

Stereotypes and STIs

In the original STI studies, actors' identities were deliberately unrelated to the behaviors and their implied traits (Winter & Uleman, 1984). Thus, behaviors with consensual explicit trait implications were paired with neutral actor names, occupations, or photos. But contexts such as actors' identities

can affect intentional trait attributions in many ways. Might stereotypes of actors' identities affect STIs if they were trait relevant?

This question was first addressed when Wigboldus et al. (2003) reported effects of stereotypes that are associated with an actor's identity on spontaneous trait inferences. Using a probe recognition RT paradigm, they found inhibition of STIs when the actor identity stereotypes conflicted with or mismatched implied traits. For example, "the professor wins the science quiz" spontaneously implied *smart*—as measured by participants taking longer to correctly decide that the probe "smart" was not in the sentence—as did "X wins the science quiz," a neutral sentence without a specified actor. But after "the garbage man wins the science quiz," RTs were significantly shorter to *smart*, indicating STIs were less activated, or to put it another way, they were inhibited. Inconsistencies inhibited STIs, but consistencies did not promote them, relative to neutral controls. These effects were more likely when cognitive capacity was low (Wigboldus et al., 2004).

Evidence consistent with this was reported by Stewart et al. (2003), using different stereotypes. They found STI inhibition when actor stereotypes were inconsistent with implied traits. Using the probe RT paradigm, they found that photos of White actors performing Black stereotypic behaviors inhibited activation of stereotypic Black traits, relative to a no-actor baseline for the same behaviors and traits, when participants were under high cognitive load (Exp. 4).

Stereotypes' effects on two kinds of spontaneous inferences were reported by Ramos et al. (2012). They studied how stereotyped behaviors' consistency with actor stereotypes affects STIs and spontaneous situational inferences (SSIs), in sentences that afford situational causes. Using probe RT in two studies, they found consistency facilitated STIs and inconsistency facilitated SSIs, relative to neutral no stereotype sentences. However these were more complex sentences, composed in order to afford alternative inferences. Wang et al. (2019) used this same paradigm with Chinese (rather than Portuguese) undergraduates. They replicated the findings of Ramos et al. (2012) for SSIs but not for STIs, suggesting that for Chinese participants, "SSIs may be more easily induced than STIs" (p. 7). More importantly, both papers show that trait-inconsistent actor stereotypes can spontaneously affect alternative inferences such as SSIs when they are afforded. STI results were inconsistent.

Effects of participants' mood and power, and gender, and elderly stereotypes were also studied in China. Wang et al. (2015) used probe RTs to study the effects of mood on STIs from gender-consistent or—inconsistent stimuli, among Chinese undergraduates. Consistent with past studies, they found that gender-inconsistency inhibited STIs, but this occurred only among participants in a positive mood. Participants in a negative mood showed no STIs at all. Wang and Yang (2017) used a similar procedure to examine effects of perceivers' power on STI. They found that effects of high power paralleled those of positive mood, using stimuli involving elderly (study 1) and gender stereotypes.

So actor stereotypes can both inhibit STIs—or more accurately, inhibit spontaneous trait activation—if they are inconsistent with the implied traits (Stewart et al., 2003; Wang et al., 2015; Wang & Yang, 2017; Wigboldus et al., 2003, 2004) and facilitate activation if they are consistent and situational inferences are afforded (Ramos et al., 2012). Note that in all these studies, STIs are compared under various conditions but there are no direct comparisons with intentional trait inferences (except in so far as pretesting involved intentional trait judgments).

However, Yan and Wang (2014) published the first studies that compared intentional with spontaneous trait inferences from gender-stereotyped behaviors. They used the probe RT paradigm. Participants read gendered and neutral behaviors paired with a man or woman photo. Half read for a memory test later, and half read to form impressions. Both RTs and error rates showed differences among sentence types. In the spontaneous condition, relative to neutral and gender-consistent sentences (which did not differ), gender-inconsistent sentences produced shorter RTs (i.e., inhibited STIs), as in studies cited previously. But in the intentional impression formation condition, gender-inconsistency produced longer RTs relative to neutral and consistent sentences, indicating that *inconsistency facilitated* trait activation. Error rates showed a similar pattern. Thus, relative to gender-neutral sentences, gender-inconsistent stimuli had opposite effects on both RTs and error rates for spontaneous and intentional trait inferences. No explanation was offered. But this study showed that spontaneous and intentional trait inferences from stereotyped actors' behaviors differ. Unfortunately, this probe RT paradigm does not tap the integration of trait inferences with actor representations, the way that both the false recognition and the savings-in-relearning paradigms do. It only measures spontaneous trait activation.

There are two obvious loose ends to this story. First, true STIs that include actor cues in the memory task have not been studied. Only trait activation has been measured. Second, there is no plausible theory of how intentional inferences could nullify or even reverse the inhibition of STI shown by Wigboldus et al. (2003, 2004), Stewart et al. (2003), Wang et al. (2015, 2017), and Yan and Wang (2014). Some sort of spreading activation, from behavior implication and actor stereotype implication, might account for the inhibition. But what mechanism can account for its nullification or reversal?

Biernat's Shifting Standards Model

Perhaps the answer lies in the standards of comparison that particular stereotypes bring to mind. Imagine you're interested in buying a horse. (I've never bought a horse, so this is completely imaginary for me.) You want to start small, and you spot ads for a "small Percheron" and a "small Clydesdale." But you also fancy race horses, and see an ad for a "large Arabian" horse. Which should you pursue with your limited feed and stable budget in mind? The "large Arabian" is the clear choice, because Arabians weigh 800 to 1,000 lbs, whereas Percherons and

Clydesdales weigh 1,400 to 2,000 lbs. The meanings of "large" and "small" differ, depending on the breed of horses they describe. People "shift their standards" to accommodate the range of values that apply to those described. A *large* turtle is *smaller* than a small elephant. Biernat's (2003, 2012) shifting standards model describes this phenomenon and how it affects the operation of stereotypes in explicit, intentional person judgments.

Biernat et al. (1991) ask participants to judge a series of male and female targets on subjective (Likert) scales or objective scales. Judgments on objective scales (inches of height, pounds of weight, dollars per year of income) reflected gender stereotypes, which are generally accurate (Swim, 1994). But subjective scale judgments (tall-short, heavy-light, wealthy-poor) did not reflect stereotypes. "[D]ifferent standards of height, weight, and financial success are used, even when respondents are explicitly instructed to make their judgments relative to the 'average person'" (p. 495). Such shifting standards were large enough to negate or in some cases reverse the effects of stereotypes on objective scales.

Trait judgments are on subjective scales, and the scales may differ for different categories (stereotypes) of people. One might try to make them objective by asking, for example, how often someone assaulted others instead of how aggressive the person is. But that's not the same thing (e.g., Block, 1989). And trait terms are inherently ambiguous in other ways (Uleman, 2005). However, their subjective nature makes them clear candidates for shifting standards, as already shown (Biernat, 2012). *Aggressive* for a woman does not mean the same thing as *aggressive* for a man.

Gender Stereotypes in STIs and Intentional Impressions

As noted previously, Yan and Wang (2014) obtained consistent results for spontaneous trait activation across two measures from the probe RT paradigm (RTs and error rates). When actor stereotypes conflict with behavioral trait implications, STIs are inhibited. But this does not ensure that the trait inferences were *about* the actors, rather than merely trait concepts activated by behaviors but unconnected with the actors. Most prior research on effects of stereotypes on STI found the same thing, with the same activation measures: stereotype inconsistent behaviors inhibit STI trait activation. What is missing is a) evidence from inconsistently stereotyped actors giving rise to spontaneous *representations about* the actors, not merely to trait activation; and then b) a comparison of such representations with explicit intentional impressions of the same actors. The studies below provide this missing evidence.

Uleman and Todorov (2021a, with Celia Gonzalez[1]) used the false recognition paradigm, which detects spontaneous trait inferences from behaviors that are incorporated into actor representations (Todorov & Uleman, 2002). As noted previously, this paradigm asks participants at encoding to read behavioral sentences that explicitly contain or merely imply traits, paired with photos of the actors who performed these behaviors. Then later

at test, they judge whether particular traits were explicitly present in the sentences previously shown with the test photos. On most test trials, they were not present. For example, at study (encoding) a photo of Mary might be paired with the sentence, "She organized a rent strike in her building." At test (retrieval), the photo of Mary would be paired with the trait *assertive* (which incidentally is gender incongruent). False recognition of *assertive* indicates that the trait had been inferred about Mary at encoding. However, false recognitions can occur for many reasons. So control trials are necessary, such as a photo of Joan paired with *assertive*. (Joan was present in different encoding trials.) The measure of STIs is the extent to which the number of false recognitions of *assertive* when paired with Mary exceeds false recognitions of *assertive* when paired with Joan. Control trials present previously seen actors (photos at study) paired with previous implied traits (at study), but for traits not previously implied about those particular actors.

In exploring effects of gender incongruencies between photos, behaviors, and traits, Uleman and Todorov (2021a) found other effects of inconsistencies among stimuli. As has long been known in the person perception literature (Hastie & Kumar, 1979), inconsistencies prompt deeper, more extensive processing under impression formation instructions. Uleman and Todorov found that this also occurs under spontaneous (memory) conditions. Several kinds of incongruency are possible on control trials, and to be adequate controls, they must match those on experimental trials. So in the first of three studies, they explored effects of various incongruencies on control trial false recognitions (as well as differences between experimental and control trials). Here are the results.

STI Control Trials

With 166 actor-behavior sentence pairs – including 24 fillers, 42 with explicit traits, 20 experimental trials, and 80 control trials – effects of three kinds of incongruencies on *control trials* were examined to see which incongruencies affected false recognition rates. These incongruencies arise from two pairs of stimuli, one pair at encoding study (photo and behavior) and one pair at the retrieval test (photo and trait). First, the sex of the photo at retrieval can be (in)congruent with the sex of the photo at study (about whom the trait was implied), as if John were presented instead of Joan in the example above. Second, the gender stereotypicality of the trait at retrieval can be (in)congruent with the gender stereotypicality of the behavior-implied trait at study. For example, the test trait might be *sloppy*, which like *assertive* is gender incongruent with Mary. So these traits are congruent with each other, in that both are gender incongruent with the actor. Third, the sex of the actor's photo and the gender stereotypicality of the implied trait on a control trial can be (in)congruent at *encoding*. (Note that a study trial at encoding becomes a control trial at test if the test photo or trait is not literally identical to that at study.) Call these factors photo congruency, trait

congruency, and photo-trait congruency, respectively. All three factors had significant effects on control trial false recognitions, in a 2 × 2 × 2 ANOVA.

In two main effects, control trial false recognitions were higher with photo congurency ($M = .267$, $SD = .18$, $F(1, 57) = 4.77$, $p = .033$, $\eta_p^2 = .077$), and with trait congruency ($M = .258$, $SD = .17$, $F(1, 57) = 12.51$, $p = .001$, $\eta_p^2 = .180$). These two factors interacted so that false recognitions were lowest with incongruency on both factors ($F(1, 57) = 10.16$, $p = .002$, $\eta_p^2 = .151$). There was also an interaction between the photo congruency and photo-trait congruency ($F(1, 57) = 13.84$, $p < .001$, $\eta_p^2 = .195$) such that when photos and traits were incongruent at encoding, false recognitions were much higher when photo congruency (between the study phase and test phase) was congruent rather than incongruent; this difference was slightly reversed when photo and trait were congruent at encoding. Although unpacking particularly this last interaction awaits future research, the basic implication of these results is that all three types of congruency affect false recognitions on control trials.

These results should be no surprise, even though they were not expected. It has long been known that the "probability of recall of an item is a direct function of the similarity between the recall situation and the original learning environment (e.g., Hollingworth, 1928; Melton, 1963)" (Tulving & Thomson, 1973, p. 359). Therefore appropriate control trials must incorporate (in)congruencies identical to the experimental trials in all respects except the one under investigation—in this case, gender (in)congruencies between the actor and the implied behavior at encoding. That is, control trials must differ from experimental trials only in the particular photos or traits presented at test, but not in their more general "learning environments." They do not control for particular variables, but for a variety of other (in)consistencies.

STI Experimental Trials and Difference Scores

Analysis of experimental trials *alone*, rather than as part of the traditional difference scores between experimental and control trials, found no effects of gender congruency between actors and implied traits at encoding. However, this analysis is misleading because it does not control for the multiple congruency effects noted above, nor for any familiarity effects that traditional difference scores take into account. However, analysis using traditional and appropriate difference scores found only a main effect of gender congruence, $F(1, 56) = 5.73$, $p = .02$, $\eta_p^2 = .093$, with a larger difference for congruent trials ($M = .11$, $SD = .19$) than incongruent trials ($M = .04$, $SD = .19$). Thus the false recognition paradigm, with appropriate controls, replicates the well-known inhibition of STIs for stereotype inconsistent behaviors reviewed previously. However, this study did not include gender-neutral behaviors to provide an informative baseline.

A second study, including gender neutral behaviors and only appropriate control trials, found only a main effect for congruence at encoding, $F(2, 120) = 6.34$, $p = .002$, $\eta_p^2 = .096$. False recognition was lower for gender-incongruent

pairs than for neutral and gender-congruent pairs, which did not differ. Replicating study 1, STIs did occur for gender-incongruent pairs, one-sample two-tailed $t(61) = 2.38$ for difference from zero, $p = .021$, $d = .302$, power = .647.

A third study used the same materials and procedure as study 2 except that exposure at study time was self-paced instead of being fixed to 5 s per photo-behavior pair. Participants took the same time to read incongruent and congruent pairs (both 6.02 s), and congruence did not affect reading times, $F(2, 118) = 1.33$, $p = .27$, $\eta_p^2 = .022$. As in study 2, there was a significant effect of congruence on the difference between experimental and control RTs, $F(2, 116) = 4.35$, $p = .015$, $\eta_p^2 = .070$. RTs were relatively shorter on incongruent ($M = .09$, $SD = .15$) than congruent trials ($M = .15$, $SD = .18$), $t(59) = 2.31$, $p = .025$, $d_z = .384$, and tended that way relative to neutral trials ($M = .13$, $SD = .16$), $t(59) = 1.94$, $p = .057$, $d_z = .30$, whereas congruent and neutral pairs did not differ. As in all three studies, STIs were significant in the incongruent condition, and participant gender had no effect.

So, these three studies (Uleman and Todorov, 2021a) showed that when stereotypes of the actors are incongruent with the gender stereotypes of the behaviors, STIs are inhibited but nevertheless present. No STI facilitation occurred, meaning that STIs were not more likely when actor stereotypes were congruent with behaviors' gendered trait implications. Neither participant gender nor stimulus reading times affected this. These STIs are trait inferences integrated into actor representations, unlike prior research which only showed inhibition of trait activation.

STIs and Intentional Inferences

Now the question arises of whether or not intentional impressions from these same materials might show Biernat's shifting standard effect when actor and behavior gender stereotypes are incongruent, thus reducing or reversing this inhibition of spontaneous inferences. Uleman and Todorov (2021b, also done with Gonzalez[2]) reported two studies that address this question. The first study was designed to mirror the procedure of Uleman and Todorov's (2021a) second study as closely as possible, except that participants were informed at the outset that they should form impressions of the people presented. Then they saw the same series of 126 photo-behavior pairs, each shown for 5 s with an intertrial delay of 2 s. Finally, they viewed the same series of photo-behavior pairs and rated each one on the target trait on a six-point scale. A 2 (participant gender) × 3 (stereotype: congruent, neutral, and incongruent) ANOVA showed that trait rating judgments were highest for stereotype incongruent behaviors ($M = 4.97$, $SD = .42$), moderate for stereotype congruent behaviors ($M = 4.92$, $SD = .46$), and lowest for stereotype neutral behaviors ($M = 4.89$, $SD = .43$). This overall effect was not significant, $F(2,90) = 1.34$, $p = .27$, $\eta_p^2 = .029$, but contrast analysis showed that judgments of stereotype-incongurent behaviors were reliably *higher* than judgments of stereotype neutral behaviors, $t(46) = 2.15$, $p = .037$, and did not differ from judgments of stereotype-congruent

behaviors, $t(46) = 1.35$, $p = .18$. Thus, after forming quick impressions of over eight photo-behavior pairs per minute for almost 15 minutes, and then rating the pairs on implied traits at their leisure, participants did not rate the incongruent pairs lower in the way suggested by the inhibition of STIs in Uleman and Todorov (2021a). Rather they rated them *higher* than the neutral pairs and as high as the congruent pairs. Apparently participants shifted standards when forming and/or reporting these intentional trait judgments. As Biernat's work suggests, this shift of standards negated or reversed effects of stereotypes on trait judgments on subjective scales.

A similar ANOVA of rating times revealed only a main effect for congruency, $F(2, 90) = 8.19$, $p = .001$, $\eta_p^2 = .154$. Paralleling the judgment data, participants made trait judgments from stereotype-incongruent behaviors faster (M = 4.50 s, *SD* = 1.53) than from stereotype-neutral behaviors (M = 4.78, *SD* = 1.42), $t(46) = 3.73$, $p = .001$. Judgments from stereotype-congruent behaviors (M = 4.56, *SD* = 1.55) were also faster than for neutral behaviors, $t(46) = 3.39$, $p = .001$. If shifting standards takes time, this paradigm is too insensitive to detect it. But then again, all judgments may require selecting frameworks or standards, so different times might not be expected. These results show that judgments from neutral behaviors took longer, as though lacking a stereotype framework of any kind slowed judgments.

Their second study was designed within-subjects, so that they could directly compare spontaneous and intentional impressions. Spontaneous impressions had to be obtained first. STIs were assessed as in Uleman and Todorov (2021a) with the same procedure and photo-behavior pairs. Then participants viewed all 96 behaviors, without the photos to avoid cuing STIs, but with the gendered actor names. They rated the actors on implied traits on a six-point scale. Results for STIs replicated previous findings in Uleman and Todorov (2021a), but only among the 60 women. The effect of congruency was significant, $F(2, 118) = 3.36$, $p = .038$, $\eta_p^2 = .054$. The critical difference between implied and control traits was reliably smaller in the incongruent condition (M = .07, *SD* = .13) than in the congruent condition (M = .12, *SD* = .16), $t(59) = 2.29$, $p = .026$, and the neutral condition (M = .12, *SD* = .19), $t(59) = 2.08$, $p = .042$, which did not differ. STIs occurred to some extent in all conditions.

Among the 31 men, there was no significant effect of congruency, $F(2, 60) = 1.02$, $\eta_p^2 = .033$. Difference scores indicated that STI occurred under only two conditions, relative to a difference of zero. For congruent pairs, M = .091, $t(30) = 3.21$, $p = .003$; for incongruent pairs, M = .085, $t(30) = 2.58$, $p = .005$. But for neutral pairs, M = .038, $t(30) = 5.11$, $p = .26$. This pattern is new and anomalous: equivalent STIs for congruent and incongruent pairs and no STIs for stereotype neutral pairs. It also arose in an unusually small sample, so the results are statistically under powered and it is difficult to be confident of the reliability of these results. Therefore, all the data from study 2 were combined with data from Uleman and Todorov's (2021a) studies 1 and 2 (*N*s = 58 and 62, respectively) that used the same stimuli and procedure. The expected significant consistency effect emerged and there was no interaction with

participant gender. That is, results from this small sample of men were anomalous, but similar enough to prior results that, when combined with two previous samples, they did not disrupt the usual findings of STIs being interfered with by incongruent pairs, equivalent for congruent and neutral pairs, and no effects of participant gender.

Intentional trait ratings were also analyzed in a 3 × 2 ANOVA. It showed a main effect of congruency, $F(2, 178) = 11.54$, $p < .001$. Stereotype-incongruent behaviors ($M = 4.94$, $SD = .57$) produced *higher* trait ratings than either stereotype-congruent ($M = 4.84$, $SD = .58$), $t(90) = 3.44$, $p < .001$, or gender-neutral behaviors ($M = 4.82$, $SD = .57$), $t(90) = 4.96$, $p < .001$. Stereotype-congruent and neutral behaviors did not differ, $t < 1$. This is consistent with the shifting standards model. More importantly, it also shows that for the same participants and stimuli, effects of actor and behavior gender stereotypes differ for spontaneous versus intentional trait inferences.

Consistent with the idea that spontaneous and intentional trait inferences reflect different processes, they were uncorrelated, $r(91) = .16$, *ns*. This was true regardless of the type of photo-behavior pair involved: stereotype-incongruent items, $r(91) = .10$, *ns*, stereotype-congruent items, $r(91) = .10$, *ns*, and stereotype-neutral items, $r(91) = .14$, *ns*.

This pattern of divergent effects of gender-inconsistent actors and behaviors replicates the effects found by Yan and Wang (2014) except that a) Yan and Wang's dependent variable was trait activation rather than full STIs incorporated into actor representations, and b) they could offer no explanation. It appears that a shifting judgment standard, when trait inferences are intentional, can account for this. Thus, STIs from gender-incongruent actors and behaviors are inhibited, but intentional inferences from this same information do not show this effect and may even be enhanced. Biernat's shifting standard model can account for this and suggests some of what people may "think more about" (Ma et al., 2011) when they make intentional inferences from stereotype incongruent information. More generally, Uleman and Todorov (2021a, 2021b) provide an example of empirically diverging spontaneous and intentional trait inferences accompanied by a theory to account for this. This theory does not account for most of the other divergences noted previously.

This pattern raises several interesting questions about reducing stereotypes. First, what kinds of groups carry their own standards. Must they be universal (like gender) or only socially important (like "race" or caste) or even inconsequential (like redheads)? And do counter-stereotypic behaviors reduce stereotypes more when they only produce spontaneous impressions, and do not give rise to intentional impressions? Conversely, are counter-stereotypic behaviors less likely to reduce stereotypes when they produce intentional impressions, because they are adjusted by shifting standards? More research, with dependent variables based on other behaviors, is needed. McCarthy and Skowronski (2011) showed that STIs can be the basis for predicting actors' future behaviors. They compared predictions following both intentional and spontaneous inference instructions (Exp. 3) and found no difference. However,

participants did not report their intentional inferences; they merely formed them in whatever amorphous framework or representational system came to mind. They did not have to confront a particular subjective Likert scale on which to rate them. Perhaps reporting such judgments on specific scales is what shifts judgment standards. Future research is needed here.

Further Speculations on Relations between Spontaneous and Intentional Thought

One might be tempted to think that spontaneous inferences are simply low energy cognitions that occur below some threshold for consciousness, and that intentions to form an impression push the process of forming spontaneous trait inferences over the threshold into consciousness. But the research reviewed previously shows that they can have different content; one is not just a louder version of the other. What variables control the processes that "edit" the content of unconscious thought as conscious thought is constructed? Perhaps Dennett's (1991) multiple drafts model of consciousness is relevant here.

One might also be tempted to think that spontaneous inferences always precede intentional inferences, in the kind of flow chart model that was popular decades ago (e.g., Uleman, 1989, p. 435; Gilbert, 1998, p. 113). But if one starts out with an intention to form an impression, spontaneous processes may be skipped or precluded, and unconscious adjustments such as shifting standards will be included from the outset. Intentions to do something – such as to form an impression in general and/or perhaps for a specific purpose (hiring, dating, befriending)—are goals. And goals activate relevant concepts and suppress irrelevant and distracting ones (Dijksterhuis & Aarts, 2010). To complicate matters further, goals can be primed unconsciously and affect STIs (e.g., Rim et al., 2013). So there appear to be multiple "sequences" depending on what a person is presented with, or what we chose as the starting point.

It is also unclear what role parallel processes or feedback loops may play, or whether this is even the right architecture (e.g., Freeman & Ambady, 2011). Furthermore, qualitatively distinct models may, when made more specific, turn out to be equivalent (e.g., Orghian et al., 2015).

So relations between spontaneous and intentional thought remain unclear. They are more complex than we thought. Knowing this is progress and grist for the mill of future research methods, results, and integrative chapters.

Notes

1 Data can be found at osf.io/qvfzx.
2 Data can be found at osf.io/qvfzx.

References

Biernat, M. (2003). Toward a broader view of social stereotyping. *American Psychologist*, 58(12), 1019–1027. doi: 10.1037/0003-066X.58.12.1019

Biernat, M. (2012). Stereotypes and shifting standards: Forming, communicating, and translating person impressions. *Advances in Experimental Social Psychology*, *45*, 1–59. Elsevier. doi: 10.1016/B978-0-12-394286-9.00001-9

Biernat, M., Manis, M., & Nelson, T. E. (1991). Stereotypes and standards of judgment. *Journal of Personality and Social Psychology*, 60(4), 485–499.

Block, J. (1989). Critique of the act-frequency approach to personality. *Journal of Personality and Social Psychology*, 56(2), 234–245. 10.1037/0022-3514.56.2.234

Bott, A., Brockmann, L., Denneberg, I., Henken, E., Kuper, N., Kruse, F., & Degner, J. (2021). *Spontaneous trait inferences from behavior: A systematic meta analysis*. Paper submitted for publication.

Carlston, D. E., & Skowronski, J. J. (1994). Savings in the relearning of trait information as evidence for spontaneous inference generation. *Journal of Personality and Social Psychology*, 66(5), 840–856.

Crouch, J. L., Milner, J. S., Skowronski, J. J., Farc, M. M., Irwin, L. M., & Neese, A. (2010). Automatic encoding of ambiguous child behavior in high and low risk for child physical abuse parents. *Journal of Family Violence*, *25*, 73–80. doi: 10.1007/s10896-009-9271-2

Dijksterhuis, A., & Aarts, H. (2010). Goals, attention, and (un)consciousness. *Annual Review of Psychology*, *61*, 467–490. doi: 10.1146/annurev.psych.093008.100445

Dennett, D. (1991). *Consciousness explained*. Boston: Little, Brown and Co.

Ferreira, M. B., Garcia-Marques, L., Hamilton, D., Ramos, T., Uleman, J. S., & Jerónimo, R. (2012). On the relation between spontaneous trait inferences and intentional inferences: An inference monitoring hypothesis. *Journal of Experimental Social Psychology*, *48*, 1–12.

Fiedler, K., & Schenck, W. (2001). Spontaneous inferences from pictorially presented behaviors. *Personality and Social Psychology Bulletin*, *27*(11), 1533–1546.

Fiske, S. T. (1992). Thinking is for doing: Portraits of social cognition from daguerreotype to laserphoto. *Journal of Personality and Social Psychology*, *63*(6), 877–889.

Freeman, J. B., & Ambady, N. (2011). A dynamic interactive model of person construal. *Psychological Review*, *118*(2), 247–279. doi: 10.1037/a0022327.

Gilbert, D. T. (1998). Ordinary personology. Ch 20. In D. T. Gilbert, S. T. Fiske & G. Lindzey (Eds.), *Handbook of Social Psychology* (4th edition, Vol. 2, pp. 89–150). Boston, MA: McGraw-Hill.

Ham, J. & van den Bos, K. (2008). Not fair for me! The influence of personal relevance on social justice inferences. *Journal of Experimental Social Psychology*, *44*, 699–705.

Ham, J., & van den Bos, K. (2011). On justice knowledge activation: Evidence for spontaneous activation of social justice inferences. *Social Justice Research*, *24*, 43–65. doi: 10.1007/s11211-011-0123-x

Hastie, R., & Kumar, P. A. (1979). Person memory: Personality traits as organizing principles in memory for behaviors. *Journal of Personality and Social Psychology*, 37(1), 25–38. 10.1037/0022-3514.37.1.25

Higgins, E. T. (1981). The "communication game": Implications for social cognition and persuasion. In E. T. Higgins, C. P. Herman & M. P. Zanna (Eds.), *Social cognition: The Ontario Symposium* (Vol. 1, pp. 343–392). Hillsdale, NJ: Erlbaum.

Higgins, E. T. (1996). Knowledge activation: Accessibility, applicability, and salience. Ch. 5. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 133–168). New York: Guilford.

Hollingworth, H. L. (1928). *Psychology: Its facts and principles*. New York: Appleton.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513–541.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Review*, *114*, 3–28.

Ma, N., Vandekerckhove, M., Van Overwalle, F., Seurinck, R., & Fias, W. (2011). Spontaneous and intentional trait inferences recruit a common mentalizing network to a different degree: Spontaneous inferences activate only its core areas. *Social Neuroscience*, 6(2), 123–138. doi: 10.1080/17470919.2010.485884

McCarthy, R. J., & Skowronski, J. J. (2011). What will Phil do next? Spontaneously inferred traits influence predictions of behavior. *Journal of Experimental Social Psychology*, *47*, 321–332. doi:10.1016/j.jesp.2010.10.015

Melton, A. W. (1963). Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior*, *2*(1), 1–21.

Moskowitz, G. B., & Roman, R. J. (1992). Spontaneous trait inferences as self-generated primes: Implications for conscious social judgment. *Journal of Personality and Social Psychology*, *62*(5), 728–738.

Narvaez, D., Lapsley, D. K., Hagele, S., & Lasky, B. (2006). Moral chronicity and social information processing: Tests of a social cognitive approach to the moral personality. *Journal of Research in Personality*, *40*, 966–985.

Orghian, D., Garcia-Marques, L., Uleman, J. S., & Heinke, D. (2015). A connectionist model of spontaneous trait inference and spontaneous trait transference: Do they have the same underlying processes? *Social Cognition*, *33*, 20–66.

Rim, S., Min, K. E., Uleman, J. S., Chartrand, T. L., & Carlston, D. E. (2013). Seeing others through rose-colored glasses: An affiliation goal and positivity bias in implicit trait impressions. *Journal of Experimental Social Psychology*, *49*(6), 1204–1209. doi: 10.1016/j.jesp.2013.05.007

Ramos, T., Garcia-Marques, L., Hamilton, D. L., Ferreira, M., & Van Acker, K. (2012). What I infer depends on who you are: The influence of stereotypes on trait and situational spontaneous inferences. *Journal of Experimental Social Psychology*, *48*, 1247–1256. doi: 10.1016/j.jesp.2012.05.009

Swim, J. (1994). Perceived versus meta-analytic effect sizes: An assessment of the accuracy of gender stereotypes. *Journal of Personality and Social Psychology*, 66(1), 21–36.

Stewart, T. L., Weeks, M., & Lupfer, M. B. (2003). Spontaneous stereotyping: A matter of prejudice? *Social Cognition*, *21*(4), 263–298.

Todorov, A., & Uleman, J. S. (2002). Spontaneous trait inferences are bound to actors: Evidence from false recognition. *Journal of Personality and Social Psychology*, *83*, 1051–1065.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*, 352–373.

Uleman, J. S. (1989). A framework for thinking about unintended thought. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 425–449). New York: Guilford.

Uleman, J. S. (2005). On the inherent ambiguity of traits and other mental concepts. In B. F. Malle & S. D. Hodges (Eds.), *Other minds: How humans bridge the divide between self and others* (pp. 253–267). New York: Guilford Publications.

Uleman, J. S., & Moskowitz, G. B. (1994). Unintended effects of goals on unintended inferences. *Journal of Personality and Social Psychology*, *66*, 490–501.

Uleman, J. S., Rim, S., Saribay, S. A., & Kressel, L. M. (2012). Controversies, questions, and prospects for spontaneous social inferences. *Social and Personality Psychology Compass*, *6*, 657–673.

Uleman, J. S., Saribay, S. A., & Gonzalez, C. (2008). Spontaneous inferences, implicit impressions, and implicit theories. *Annual Review of Psychology*, *59*, 329–360.

Uleman, J. S., & Todorov, A. (2021a). When behavior implications and actor stereotypes conflict, spontaneous trait inferences are weaker. Unpublished manuscript, New York University. Available at osf.io/qvfzx

Uleman, J. S., & Todorov, A. (2021b). Divergent effects of gender stereotypes on intentional and spontaneous inferences. Unpublished manuscript, New York University. Available at osf.io/qvfzx

Uleman, J. S., Winborne, W. C., Winter, L., & Shechter, D. (1986). Personality differences in spontaneous personality inferences at encoding. *Journal of Personality and Social Psychology*, *51*, 396–403.

Wang, M., Xia, J., & Yang, F. (2015). Flexibility of spontaneous trait inferences: The interactive effects of mood and gender stereotypes. *Social Cognition*, *33*(4), 345–358.

Wang, M., & Yang, F. (2017). The malleability of stereotype effects on spontaneous trait inferences: The moderating role of perceivers' power. *Social Psychology*, *48*(1), 3–18. doi: 10.1027/1864-9335/a000288

Wang, P., Tao, A., Gao, F., & Xie, Y. (2019). The effect of stereotype activation on spontaneous inferences. *Social Behavior and Personality: An International Journal*, *47*(8), e7470.

Wigboldus, D. H. J., Dijksterhuis, A., & van Knippenberg, A. (2003). When stereotypes get in the way: Stereotypes obstruct stereotype-inconsistent trait inferences. *Journal of Personality and Social Psychology*, *84*(3), 470–484. doi: 10.1037/0022-3514.84.3.470

Wigboldus, D. H. J., Sherman, J. W., Franzese, H. L., & van Knippenberg, A. (2004). Capacity and comprehension: Spontaneous stereotyping under cognitive load. *Social Cognition*, *22*(3), 292–309.

Winter, L., & Uleman, J. S. (1984). When are social judgments made? Evidence for the spontaneousness of trait inferences. *Journal of Personality and Social Psychology*, *47*, 237–252.

Yan, X., & Wang, M. (2014). Effects of gender stereotype on spontaneous and intentional trait inferences. *Chinese Journal of Clinical Psychology*, *22*(6), 976–980.

Zajonc, R. B. (1960). The process of cognitive tuning and communication. *Journal of Abnormal and Social Psychology*, *61*, 159–167.

Zelli, A., Cervone, D., & Huesmann, R. L. (1996). Behavioral experience and social inference: Individual differences in aggressive experience and spontaneous versus deliberate trait inference. *Social Cognition*, *14*, 165–190.

Zelli, A., Huesmann, L. R., & Cervone, D. (1995). Social inference and individual differences in aggression. Evidence for spontaneous judgments of hostility. *Aggressive Behavior*, *21*, 405–417.

Zhang, Q., & Wang, M. (2018). The primacy-of-warmth effect on spontaneous trait inferences and the moderating role of trait valence: Evidence from Chinese undergraduates. *Frontiers in Psychology*, *9*(2148). doi: 10.3389/fpsyg.2018.02148

6 Bridging the Gap between Spontaneous Behavior- and Stereotype-Based Impressions

Jacqueline M. Chen[1], *Kimberly A. Quinn*[2], *and Keith B. Maddox*[3]

[1]*University of Connecticut and University of Utah*
[2]*DePaul University*
[3]*Tufts University*

Since the earliest days of social psychology, researchers have been interested in understanding how processes involved in perceiving *people* differed from those involved in perceiving *objects* (Asch, 1946; Heider, 1958). The principle that came to guide the field of social cognition was that humans are inherently social beings, and consequently their minds may process information about other humans in distinctive or specialized ways (Hamilton & Stroessner, 2021). Indeed, there grew to be a large body of work documenting the cognitive mechanisms that demonstrated the social perceiver's propensity for quickly making sense of other individuals or groups of individuals, likely in anticipation of future interpersonal interaction. In this chapter, we highlight a gap between spontaneous behavior-based inferences about social targets and another critical form of social inference that we have each investigated in our own research: stereotype-based inferences. Our goal is to bridge this gap in the literature that explores spontaneous behavior-based inferences on the one hand and stereotype-based inferences, particularly racial stereotyping, on the other.

Spontaneous Trait Inference and Stereotypic Inference as Overlapping Processes

Social perception and judgment can involve a wide range of processes and outcomes. Here, we are interested in the subset of that literature focusing on the social cognitive processes by which we get to know and make judgments of others. Generally speaking, this literature is known by many names: person perception, person memory, and impression formation, among others. Within this literature, research varies in the inputs, processes, and outputs of theoretical and empirical focus. One foundational advance to this body of social cognition research was the discovery of the process known as *spontaneous trait inference* (spontaneous trait inference), the act of efficiently and effortlessly interpreting a person's behavior as indicative of their personalities (Winter &

DOI: 10.4324/9781003045687-7

Uleman, 1984). Work by Jim Uleman and his colleagues definitively showed that a perceiver witnessing Sally help an old lady cross the street would readily infer that Sally was a helpful person. Inferences of this nature share some features with another foundational social perception phenomenon: stereotyping.

Spontaneous Trait Inference vs. Stereotyping: Similarities and Differences

By generalizing from acts to dispositions, perceivers are actively interpreting targets' behaviors and "going beyond the information given" to generate trait-based target representations. From our perspective, these generalization processes sound a lot like another socially significant process: stereotyping. While definitions abound, stereotyping involves going beyond the information given by 1) inferring a person's social group membership from available cues (e.g., their appearance), and 2) subsequently using any traits and characteristics associated with that social group as a lens for judgments of the individual. As mentioned earlier, when people see Sally help an old lady cross the street, they infer that Sally is a helpful person (Todorov & Uleman, 2002, 2003; Uleman, 1987; Winter et al., 1985). In addition, when people meet Sally and learn that she is a nurse, they also spontaneously infer that she is a helpful person, without reference to information about her behavior (Chen et al., 2014). Both the *spontaneous trait inference* and the *spontaneous role inference* (extrapolating a person's traits based on their social role) constitute the perceiver actively interpreting knowledge about Sally (behavior- or role-based) as indicative of her character. More generally, we believe that the intersection of the processes of spontaneous trait inferences based on targets' *behavior* and stereotyping based on targets' *social group memberships* is under-examined and deserves direct theoretical and empirical comparison.

We highlight spontaneous trait inferences specifically as a form of behavior-based inference to honor Uleman's groundbreaking work, but also because spontaneous trait inferences enable the most direct comparisons between behavior- and stereotype-based inferences. Both occur spontaneously at early stages of information processing and are formed on the basis of similarly minimal cues (faces or category labels for stereotypic impressions; individual behavioral episodes for spontaneous trait inferences). Like stereotyping, spontaneous trait inference constitutes a precursor to other processes guiding the formation, representation, and retrieval of impressions in models of person memory (e.g., Hastie et al., 1980). Similarly, spontaneous trait inferences are likely formed prior to (and potentially influencing) attributional processes reflecting deeper consideration of causal forces, or information over time across episodes and individuals (e.g., covariation model; Kelley, 1967). In addition, spontaneous trait inferences reflect a focus on a single cue—behavior—at the expense of other potentially impactful elements guiding impression formation such as expectancies or situations, while

stereotyping arguably reflects a focus on a single cue— expectancies—at the expense of behavior or situations.

We explore this comparability below, documenting theoretical and methodological factors that have created a gap between the two literatures and considering several implications. We conclude by suggesting directions for research in an effort to bridge the gap and to make research exploring spontaneous trait inferences more inclusive and generalizable.

Diverging Inputs of Focus in Stage Models of Behavior- and Category-Based Impressions

Indeed, models of person perception and impression formation involve very similar processes regardless of whether the initial input is behavior (e.g., Gilbert et al., 1988) or membership in a group such as a socially significant category or role (e.g., Brewer, 1988; Fiske & Neuberg, 1990). With respect to person perception, Gilbert et al. (1988) conceptualized the process as involving three broad stages: categorization, characterization, and correction. First, the perceiver categorizes the target's behavior via dispositional inference, judging what kind of mental state the behavior implies. The perceiver then characterizes the target via trait inference, applying the dispositional inference. Finally, in the face of either contradictory information or knowledge of an acute situational pressure to act, and assuming motivation and capacity, the perceiver corrects their impression of the target based on behavior that suggests different trait inferences (see also Trope, 1986, for a similar stage model).

Similarly, Brewer (1988) and Fiske & Neuberg (1990) proposed models of stereotype-driven impression formation that describe similar processes of categorization, characterization, and correction. However, their focus on "inference" does not describe moving specifically from observed target behavior to a personality trait, but moving from a variety of observed target cues to a social category membership (i.e., what group in my mental representation does this target best match?). According to these impression formation models, the social perceiver first categorizes the *target themselves*, rather than their behavior, via social categorization and/or stereotype activation. The perceiver then characterizes the target via trait inference, but in this case the trait derives from the activated stereotype. Finally, processes of correction are applied in the face of contradictory information—in this case, stereotype-contradictory information, which may present in the form of stereotype-incongruent behavior from the target.

Tests of these models typically begin at the same starting point: the presentation of an unknown target. In the case of spontaneous trait inference, the target is often labeled with nothing more than a name; even when the stimuli include categorical cues—generally role or occupation labels—the disposition-suggestive behaviors are chosen to be low in stereotypicality (i.e., neither so stereotypical as to be redundant with category membership nor so

counterstereotypical as to be implausible; e.g., Winter & Uleman, 1984). In the case of stereotypic inference, the target is either visibly identifiable as a member of a particular social category or labeled as such. But despite this initial difference in input, the subsequent processing appears to be very similar, in terms of processing characteristics and moderators for behavior-based versus stereotype-based impressions. In both domains, for example, the processes of categorization and characterization are at least conditionally automatic, and are moderated by perceivers' capacity and motivations, including both social versus non-social processing goals (person perception: e.g., Uleman & Moskowitz, 1994; Uleman et al., 1992; Winter et al., 1985; stereotyping: e.g., Macrae et al., 2005; Moskowitz et al., 1999; Quinn & Macrae, 2005; Spencer et al., 1998). Ultimately, models of person perception and impression formation converge to claim that perceivers eventually place the target on a continuum ranging between a (trait or social) category-based versus individuated impression based on all of the information available (including behavior) with a bias toward confirming the initial categorization.

Non-Overlapping Empirical Focus in the Spontaneous Trait Inference and Stereotyping Literatures

Despite these similarities and hypothesized links between person perception and impression formation, we struggled to think of studies representing points of overlap where the stereotyping and impression formation were jointly investigated.[1] And, despite the fact that social cognition researchers are frequently interested in understanding issues related to stereotyping and prejudice, the area of behavior-driven person perception has not fully integrated knowledge about how social category stereotypes, and related processes such as egalitarian motives, impact how perceivers form impressions of diverse individuals. In particular, racially heterogeneous contexts and cross-race impression formation processes remain understudied. Historically speaking, the trait inference and stereotyping literatures have seemingly progressed independently, such that trait inferences were studied in largely racially homogeneous contexts and research on stereotyping was not explicitly linked to informing models of impression formation (but see McConnell et al., 1994, 1997; Sanbonmatsu et al., 1987 for direct comparisons between judgments of individuals vs. groups). For example, while the early impression formation literature acknowledges a differential impact of trait- versus category-based impression, only one study comes to mind that explored the effects of a racial category-based impression while using more traditional impression formation methodologies (Stewart et al., 1998). We see an opportunity to integrate these two literatures in ways that will broaden the scope and legacy of Uleman's signature contribution to the field. Furthermore, we comment on recent encouraging developments, spearheaded by Uleman and others, that begin to diversify the person perception literature.

What We Know about the Connections between Spontaneous Trait Inferences and Stereotyping

Stereotype Inconsistency Disrupts Spontaneous Trait Inferences

Despite the commonalities between spontaneous trait inferences and stereotyping, only a few studies have investigated their interplay. One example was reported by Wigboldus et al. (2003). They used a reaction-time paradigm in which sentences (including sentences that described behaviors with clear trait implications) were presented briefly on a computer screen, and then quickly replaced with a target word; the participant's task was to indicate whether the target word appeared in the preceding sentence. Importantly, the trait-implying sentences either described the target in categorical terms (e.g., "the girl") or were primed subliminally with a category label. Across five experiments and using a variety of targets (e.g., *girl*, *skinhead*, *professor*, *nurse*), Wigboldus et al. demonstrated that dispositional inferences were less likely to be applied to targets when those inferences were inconsistent with stereotypes about the target's social category—that is, spontaneous trait inference was less likely to occur. Gonzalez et al. (unpublished; as cited by Uleman et al., 2012), using Uleman's standard false-recognition paradigm, similarly demonstrated that trait implications were less likely to be bound to targets when those traits were inconsistent versus consistent with the target's gender stereotype. These findings converge with work by Yan et al. (2012), who found that gender stereotypic spontaneous trait inferences were more strongly formed than gender counter-stereotypic spontaneous trait inferences.

Spontaneous Trait Inferences Can Be Formed for Group Targets

Another extension of the spontaneous trait inference literature focused on how these processes could be applied to the perception of groups' behaviors, thereby "sowing the seeds" of stereotyping (Hamilton et al., 2015). Specifically, Hamilton et al. (2015) noted that spontaneous trait inference research uniformly presented individual targets, despite the fact that social perception regularly involves perceiving groups of individuals performing trait-implying behaviors (e.g., the teachers went on strike, the church group visited an impoverished village to provide aid, the CEOs lobbied Congress against tax increases). Adapting Uleman's false recognition probe paradigm, Hamilton et al. presented participants with groups (collections of four individuals) described as engaging in trait-implying behaviors similar to the aforementioned examples. After participants viewed a series of groups engaging in trait-implying behaviors (vs. filler behaviors that did not imply a trait), they were shown the same groups and a probe word, which they were asked to identify as having been in the sentence about that particular group or not. The probe word was either a matched trait (implied by that specific group) or a mismatched trait (implied by another group). Participants were

more likely to falsely recognize the matched traits than the mismatched traits, the signature evidence of spontaneous trait inferences. These results suggested that perceivers generate spontaneous trait inferences about groups (sometimes referred to as STIGs). In other words, learning that a fraternity group partied all weekend without any breaks might imply that the group is irresponsible, and perceivers who witness or learn about this behavior would readily form the spontaneous trait inference that this group is irresponsible. Importantly, Hamilton et al. demonstrated that spontaneous trait inferences generalized from the original four members of the group to a new group member who was not involved in the trait-implying behavior. Therefore, the spontaneous trait inference transfers between group members and the spontaneous trait inference itself can be considered a trait impression at the group level (not about any particular group member). The importance of generalization from the trait inference about the group to individual members, regardless of their role in the group's behavior, is an important mechanism by which stereotypes about a group can be formed.

Future Directions: Decolonizing the Spontaneous Trait Inference Literature by Expanding the Intersection of Spontaneous Trait Inference and Racial Stereotyping Research

Although the studies at the intersection of spontaneous trait inference and stereotyping research summarized above provide an initial understanding of how stereotypes constrain spontaneous trait inferences and how spontaneous trait inferences about groups may contribute to stereotype formation, there are many unanswered questions about the interplay between stereotypes and spontaneous trait inferences for future investigation. Additionally, we would be remiss not to point out that the divide we have described above between spontaneous trait inference and stereotyping research also reflects a notable divergence in the consideration of race, gender, ethnicity, and other social identities with respect to the research questions, stimuli, participants, and researchers engaged in work exploring social perception processes other than stereotyping (e.g., attribution, person memory) with only a few exceptions (e.g., Stewart et al., 1998; Xie et al., 2019). These concerns are not unique to the literature exploring social perception and judgment. Indeed, the field of psychological science finds itself in a time of uncertainty and change as we respond to critiques that challenge us to grow into a more open, reproducible, and inclusive endeavor (e.g., Arnett, 2008; Cheon et al., 2020; Cole, 2009; Murphy et al., 2020; Roberts et al., 2020; Syed & Kathawalla, 2020) and offer suggestions for how we can make progress on this front (Ledgerwood et al., 2022, 2021).

Keeping these goals in mind, this section outlines how the convergence of these two research literatures can contribute new knowledge with an admittedly biased focus on race and ethnicity. Thus, the directions for research

that we've outlined reflect questions of theoretical interest *and* contribute to a process of making spontaneous trait inference research more broadly inclusive and generalizable. Where applicable, we identify any existing efforts toward this goal reflected in the work of Jim and others and propose a few opportunities for continued progress.

Intersection of Behavior- and Stereotype-Based Inference Processes Unfolding over Time

A potential avenue for fruitful integration of the spontaneous trait inference and stereotyping literatures would be to examine the time course of processing and establish the conditions under which behavioral trait inferences versus stereotypic trait inferences might come to dominate impressions. The scant literature available on this question (Gonzalez et al., unpublished; Wigboldus et al., 2003; Yan et al., 2012) suggests that stereotypes constrain trait inferences, and there are reasons to expect that this would be the typical state of affairs. Although current models of person perception tend to posit parallel rather than serial processing of information (e.g., Freeman & Ambady, 2011) and thus allow for the theoretical possibility for category membership and behavior to exert equal or mutually constraining influence on impressions, evidence suggests that the accumulation of evidence for categorical versus individuated impressions differs in the time needed (e.g., Cloutier et al., 2005; Quinn et al., 2010). Based on this evidence, it seems reasonable to expect that the categorization of the target's group membership should typically precede the categorization of the behavior's dispositional implications simply because social categorization can proceed with very little input (e.g., on the basis of easily extracted superficial cues such as skin tone, which can be detected in a matter of milliseconds), whereas behavioral categorization would take longer simply because behaviors unfold over time.

This conjecture, however, ignores a number of interesting questions about the development and updating of impressions over time. One key difference between the spontaneous trait inference and stereotypic impression literatures is that stereotype-driven impressions are anchored by long-standing stereotypes, whereas the literature on spontaneous trait inferences documents a process whereby an impression is generated on the basis of behavior committed by a target for whom the perceiver has no pre-existing associations. These pre-existing associations have important implications. A great deal of research has suggested that stereotypes act as attentional filters that direct our processing vis-a-vis stereotype-consistent versus -inconsistent information. According to the encoding flexibility model (Sherman et al., 1998), for example, stereotypes serve to direct attention toward stereotype-inconsistent information, because the conceptual fluency of stereotype-consistent information means it can be comprehended and assimilated into the existing impression with little effort.

Thus, stereotype-based inferences are anchored from the beginning of the person perception process (assuming category membership is immediately accessible), and these stereotypes continue to direct attention with each new piece of information about the target. In the case of behavior-based inferences, however, absent long-standing stereotypes, there are no attentional filters to guide initial impression formation. An interesting question is whether the initial behavior-based inferences are strong enough to act as attentional filters and anchor subsequent inferences about new information, and the processes by which such updating of impressions might occur (see Moskowitz et al., Chapter 18, this volume). Some research suggests that they are not, in that implicit revision of impressions has been documented repeatedly (e.g., Gawronski & Bodenhausen, 2006; Wyer, 2010, 2016; for a review, see Cone et al., 2017) and can emerge even after a significant delay between initial behavioral encoding and subsequent reinterpretation (e.g., Mann & Ferguson, 2017). An open question, therefore, is whether there is a certain type or amount of behavior-based inference that is necessary for spontaneous trait inferences to have attentional filtering capacity.

Another interesting question about the time course of these processes is whether slower-to-accumulate behavioral information can override the stereotypes activated by the rapid detection of social category membership. Of course, stereotype-inconsistent behaviors are assumed to initiate correction processes (e.g., Brewer, 1988; Fiske & Neuberg, 1990). What we are referring to, however, is the capacity for dispositional attributions to compete successfully with stereotypic attributions in initial impressions. Stated differently, are there conditions that might facilitate the extraction of dispositional inferences from behavior, giving them a chance to compete with stereotypic inferences from categorization? One possibility rests on the question of what stereotypic attributions reflect. It seems intuitively reasonable that stereotypic attributions are a form of trait attribution, and that the trait being applied to the target is part of the stereotype content—a view that we endorsed at the outset of this chapter. In this case, spontaneous trait inferences and stereotype-based attributions are functionally equivalent, differing only in the ease with which the driving cues can be extracted. This intuition, however, is untested. Perhaps stereotypic attributions are more transient in nature, reflecting attributions of the target's current dispositions or goals. In this case, the two forms of attribution might not be functionally equivalent, and some evidence suggests that goal inferences are inferred faster than trait inferences (Malle & Holbrook, 2012; Smith & Miller, 1983; Van Overwalle et al., 2012).

This goes back to the question of whether spontaneous trait inferences can anchor construal of subsequently encountered information and hinges on the capacity of visual cues to activate trait-based inferences. One of the hallmarks of spontaneous trait inferences is that they are tagged to the actor (Todorov & Uleman, 2002, 2003, 2004). Specifically, in studies using the standard false-recognition paradigm, participants erroneously recall the presentation of trait

words implied by the behavioral descriptions more frequently when the recall cue is the actor versus another individual who was also presented with the behavioral description, and regardless of whether the actor image presented at recall is the one presented at encoding or another image of the same actor. If spontaneous trait inferences are linked to the actor and if visual representations of actors have implications for subsequent processing of new information (which the small literature on evaluative generalization to perceptually similar faces suggests; e.g., Gawronski & Quinn, 2013; Verosky & Todorov, 2010), then perhaps the impact of rapid categorical processing can be constrained. Indeed, evidence suggests social categorization can be offset by target familiarity, such that for familiar targets, even early construals are based on identity more than social category (Quinn et al., 2009).

Spontaneous Trait Inferences in Racially Diverse Contexts

Importantly, the majority of past work has examined social roles about which there are stereotypes (e.g., nurse, skinhead, professor) or focuses on creating stereotypes about novel groups experimentally. Although some work has examined the role of gender on spontaneous trait inferences (Gonzalez et al., unpublished; Yan et al., 2012), there is a large gap in understanding how racial group memberships impact spontaneous trait inferences. To our knowledge, the only published study that has explored the effects of a racial category-based impression while using more traditional impression formation methodologies was embedded in Stewart et al. (1998). They investigated whether inferences drawn from one person's behavior would influence a future judgment of the same person. In their Experiment 3, they asked White participants to make timed trait inferences from behaviors (Black-stereotypic or neutral) first associated with photographs of Black or White actors, and then again with the same or a different Black or White actor. Analyses of the facilitation scores revealed a predicted actor–context effect: Trait inferences were facilitated when a behavior was associated with the same vs. a different actor, and even more so when the behavior was stereotypic vs. neutral. However, this experiment studied deliberate rather than spontaneous trait inferences. While it is possible that prior spontaneous trait inference work using faces has incorporated greater variability in the racial and ethnic identities of the faces used, this information is rarely specified and never the empirical focus (e.g., Todorov & Uleman, 2002, 2003). Thus, research on the influence of racial group membership on spontaneous trait inference formation is needed.

Another interesting research avenue that is sparked by considerations of cross-race impression formation is the influence of social motivations in the processing of spontaneous trait inferences. Whereas there is some work on motivated impression formation (see Fein, 1996, on how suspicion of ulterior motives disrupts spontaneous trait inferences), most of this work focuses on intra-group impression formation. One exception is Otten and Moskowitz

(2000), who used a minimal group paradigm to investigate whether mere group membership facilitated or inhibited spontaneous trait inferences. They found evidence of ingroup favoritism, such that positive spontaneous trait inferences were facilitated for ingroup members' behaviors, though they did not find evidence for outgroup derogation (no facilitation of negative spontaneous trait inferences for outgroup members' behaviors; for a review, see Brewer, 1979). Their findings indicate that the motivations based in social group memberships could facilitate or inhibit spontaneous trait inferences. Yet, to date, we are aware of no research that considers how impression formation processes in cross-race situations. Perhaps outgroup derogation in spontaneous trait inference formation would occur within an interracial context, in particular for stereotype-consistent spontaneous trait inferences. Furthermore, consideration of interracial impression formation raises interesting questions with respect to other types of social motivations, such as motivation to be non-prejudiced (Moskowitz & Li, 2011; Plant & Devine, 1998) or motivation to maintain or reinforce group stereotypes (Oldmeadow & Fiske, 2007). However, we propose that intergroup considerations pose many interesting and currently unanswered questions for the processing of behavioral information.

For instance, imagine a scenario where a White perceiver, Elizabeth, is forming an impression of her new co-worker, Tyrone, who is a Black man. It is personally important to Elizabeth that she is non-prejudiced (Plant & Devine, 1998). In his first days at work, Tyrone submits an assignment to the supervisor where he clearly misunderstood the instructions. Does Elizabeth make a spontaneous trait inference about Tyrone, concluding that he is incompetent? In this case, Elizabeth's egalitarian motives may disrupt classic spontaneous trait inference processes and result in more deliberative processing of the behavioral information, consistent with Gilbert et al.'s (1988) model described earlier. Whereas past research has shown that dispositional (chronic) goals can impact the likelihood of spontaneous trait inference formation (Uleman et al., 1985), another possibility has not yet been addressed. Specifically, Elizabeth's motives could change the nature of the spontaneous inference made, perhaps facilitating a *spontaneous situational inference* (e.g., others did not clearly explain expectations). Current goals could either inhibit a correspondent spontaneous trait inference or facilitate other, more goal-congruent spontaneous inferences. Finally, Elizabeth may still form stereotype-consistent spontaneous trait inferences that may or may not be overridden by deliberative processing. Whether correction occurs may depend on other factors such as the extent to which Tyrone otherwise fits the prototype of a Black man in Elizabeth's mind (e.g., Maddox, 2004). This example illustrates the sudden relevance of group-based motivations to impression formation processes once the interaction moves from same-race to cross-race contexts. It also highlights the understudied and complicated interaction of perceiver motives, behavior type, and target factors influencing person perception.

Continuing the example in the other direction, Tyrone's stereotypes about White people and whether they are to be trusted (Major et al., 2016) may influence his impressions of Elizabeth. If she behaves in a friendly manner toward him, Tyrone could make the correspondent spontaneous trait inference—that she's a friendly person—or that she is just trying to appear non-prejudiced to him and other co-workers. Seminal research by Crocker et al. (1991) suggests that Tyrone will not infer that Elizabeth is a friendly person and instead will experience a drop in self-esteem because he attributes her behavior to the external motivation to appear non-prejudiced. However, those researchers were examining the affective consequences of explicit attributions, whereas we are interested in the role of stereotypic beliefs on spontaneous trait inferences. *We know of no research that has investigated whether and under what conditions perceivers of color make spontaneous trait inferences about White targets*, let alone examine the role of stereotypic beliefs or motivations that could shape people of color's impressions of White individuals. Because research suggests that suspicion of ulterior motives disrupts spontaneous trait inferences (Fein, 1996), it may be that perceivers of color who associate White people with disingenuity may resist forming spontaneous trait inferences about White targets broadly.

In sum, the racial stereotypes and race-based motivations of perceivers could play important roles in shaping cross-race impression formation, and especially the likelihood of spontaneous trait inference formation. Here the predictions will depend not only on who the target is, but also on the valence of the trait-implying behavior, and the social motivation of the perceiver.

Expanding Participant Populations

One exception to the dearth of research on perceivers of color in the person perception literature is research examining cross-cultural variability on the inference and attribution processes (e.g., Choi & Nisbett, 1998; Masuda & Kitayama, 2004; Miyamoto & Kitayama, 2002). Early research in this area showed that East Asian perceivers are less likely to form trait inferences from trait-implying behavior, in part due to increased attention to the context as a causal force on targets' behaviors. Notably, Jim Uleman has been involved in conducting cross-cultural research in spontaneous trait inference formation for decades (Rhee et al., 1995, 1996; see also Na & Kitayama, 2011) and has continued these efforts throughout his career. Uleman's recent work with colleagues Yuki Shimizu and Hajin Lee demonstrates a recognition of this need to examine spontaneous trait inferences through a lens other than that of undergraduate students in the United States forming impressions of White male targets. Shimizu et al. (2017), for example, compared the responses of U.S. American and Japanese undergraduates, and found that Japanese participants made fewer spontaneous trait inferences, and that their responses were driven less by automatic processing, compared to their U.S. American

counterparts (see also Lee et al., 2015, 2017). More recently, Shimizu and Uleman (2021) replicated the finding that White American participants generated more spontaneous trait inferences than both Japanese participants and Asian American participants and used eye-tracking analysis to provide evidence for cultural differences in attention allocation as a mediator of cultural variation in spontaneous trait inference.

While Uleman's efforts to diversify the spontaneous trait inference literature are admirable, much more research is needed. Although some recent studies have endeavored to determine the global generalizability of *face-based* trait inferences (Jones et al., 2021), we believe that focused experimental work with more racially diverse samples is also needed. As developed in the previous section, examining cross-race spontaneous trait inference formation will provide a more comprehensive understanding of the role of perceiver factors, target factors, and their interactive effects on person perception.

Methodological Traditions and Opportunities

A hallmark of spontaneous trait inference research, and Jim's legacy, is methodological rigor (e.g., Winter & Uleman, 1984). Researchers create new stimulus sentences of behaviors that could imply traits (e.g., "Sean slipped money into his wife's purse") that are carefully pretested to ensure that they reliably imply a trait (e.g., "generous") among the sample population. Furthermore, researchers have made sure that the implied trait words are not contained, in any form, within the sentence itself. Probe words used in experiments are frequently examined so that they are not too unique or pedantic (e.g., "benevolent" is a less ideal probe word than "generous"). Within an experiment, sentences must imply unique traits (not implying the same trait multiple times, as this would obscure spontaneous trait inference formation about a particular target) and there is always a balance between positive and negative traits implied. Meticulous attention to methodological detail is a hallmark of the spontaneous trait inference tradition.

Explore Behavior Ambiguity

Now that the spontaneous trait inference process has been firmly established, it is time to re-evaluate the methods that have become standard practice in this research area. We observe that, in the previously described approach, any ambiguity is essentially pre-tested out of the stimuli. Behaviors can be ambiguous with respect to their implications; sometimes a behavior could be potentially friendly or unfriendly depending on numerous factors (e.g., the classic "Donald" paragraph featuring ambiguously hostile behaviors; Srull & Wyer, 1979; see also McCarthy et al., 2018, 2021). Because Uleman et al.'s original research was aimed at establishing that spontaneous trait inferences occur, it became paradigmatic to focus on clearly trait-implying behaviors.

Yet ambiguity can be a researcher's tool in trying to understand the underlying process. For example, using ambiguous behaviors could clarify the person- and situation-based factors that lead to spontaneous trait inferences based on ambiguous behaviors. Using behaviors that "strongly" implied traits, Moskowitz (1993) found that people who are high in the personal need for structure were more likely to use spontaneous trait inferences. Relatedly, Olcaysoy Okten and Moskowitz (2020) found that political conservatives, who are more motivated to seek consistency, were more likely to form spontaneous trait inferences than political liberals (see also Olcaysoy et al., 2018). Yet, since Moskowitz's and Olcaysoy Okten's contributions, there has been limited follow-up research on the dispositional factors that can influence the strength of spontaneous trait inference formation. It is possible that researchers who have sought to follow up on individual variability in spontaneous trait inference formation have found null person-by-situation interactions when using strong trait-behavior stimuli. We speculate that the lack of ambiguity in spontaneous trait inference sentences may decrease the chances of finding effects of individual difference characteristics.

Explore Target Appearance Variability

Moving away from clearly trait-implying behaviors enacted by White targets (the predominant spontaneous trait inference paradigm) has the capacity to move the field forward in other ways. The usefulness of stimulus variability has recently been harnessed in the growing literature on face perception and judgment. For example, a person can also be racially ambiguous, in that their particular racial group membership is difficult to discern (Chen & Hamilton, 2012) or not consensually agreed upon by others (Chen et al., 2018). The use of racially ambiguous faces has clarified a number of interesting top-down and bottom-up processes involved in the racial categorization process (e.g., Chen et al., 2014; Freeman et al., 2011; Johnson et al., 2012). In addition, an overlooked aspect of racial bias concerns within-race variation on race-related cues among unambiguous targets. When considering physical appearance, racial phenotypicality bias reflects a tendency to use within-race variation in facial appearance to guide judgments and behavior (Maddox, 2004). For example, Black people are stereotypically thought to have dark skin tone, broad noses, full lips, and tightly curled hair. Research exploring within-race variation on these features contends that, across a variety of domains, Black people with more stereotypical appearance are more closely associated with category stereotypes (Hinzman & Maddox, 2017; Maddox & Gray, 2002) and have poorer societal outcomes compared to their less stereotypical counterparts (for reviews, see Adams et al., 2016; Maddox, 2004, Maddox & Dukes, 2008).

Consequently, we believe that these approaches exploring racial phenotypicality and ambiguity could be fruitful for understanding perceptions of behaviors and spontaneous trait inference formation, particularly when

coupled with manipulations of behavior ambiguity and stereotypicality. Racial phenotypicality and racial ambiguity reflect the strength of association between a target's appearance and a potential racial categorization. As such, these cues have the potential to constrain or facilitate the formation of spontaneous trait inferences depending on the ambiguity and stereotypicality of the behavior. For example, an ambiguously aggressive behavior could elicit increasingly stronger trait inferences when performed by a racially ambiguous, low-, or high-phenotypically Black person. Likewise, behavior ambiguity and stereotypicality may constrain or facilitate target categorization depending on facial ambiguity and phenotypicality. A Black-White racially ambiguous person who excels in athletics might be categorized as Black instead of as White, whereas another ambiguous person who excels in academics might be seen as more White than Black. This kind of spontaneous trait inference work could contribute to our understanding of stereotype change by influencing the likelihood that counterstereotypic behavioral information is associated with an individual and generalize to inferences about the group (e.g., Hinzman & Maddox, 2017; Maurer et al., 1995).

These examples illustrate the potential interplay between social category and behavioral information that we believe is ripe for future research. However, in order for the proposed work to occur, researchers will need to draw on large sets of racially diverse stimuli. Fortunately, in recent years, the number of faces of various different races, including ambiguous faces, has grown substantially, in large part due to the Chicago Face Database and the American Multiracial Faces Database (Chen et al., 2021; Ma et al., 2015, 2021). With these stimulus sets, in combination with published spontaneous trait inference behavioral stimuli, it will be possible for researchers to examine how racial stereotypes and motivations shape impression formation.

Diversifying the Research

Scholars of color are drawn to topics that enable them to study intergroup disparities, and the spontaneous trait inference literature could do more to integrate the kinds of questions that interest scholars of color. As we have already noted, the methodological strength of the existing spontaneous trait inference literature—its use of carefully controlled stimuli—might also be an impediment to future progress in developing a more nuanced understanding of behavior-based impression formation. The spontaneous trait inference literature might better be described in less general terms: as a literature documenting behavior-based impressions of White targets, or targets assumed to be White, engaging in unambiguous behavior.

Early research on spontaneous trait inferences used behavioral descriptions without photos of the actor. Although this work did not explicitly narrow its focus to White targets, plenty of research has documented that perceivers assume that a person is White (and male) when their race and gender are unspecified (see Bailey et al., 2020; Zarate & Smith, 1990; Stroessner, 1996).

Once the paradigms incorporated faces, many studies focused exclusively on White male targets for the sake of experimental control (e.g., Hamilton et al., 2015), and others did not specify the demographic composition of their stimuli (a serious methodological oversight, in our opinion).

Of course, not all the spontaneous trait inference literature focuses on White male targets, but when preponderance of evidence seems to imply or describe a particular type of target, scholars who do not see themselves as represented in the research—or in the community of scholars conducting that research—might reasonably assume that the topic is irrelevant to them. And yet, we know that developing comprehensive accounts of psychological phenomena requires a diversity of perspectives, questions, and methods—and what better way to achieve that than by setting up the conditions to encourage scholars of diverse backgrounds to join the community?

To that end, we advocate for researchers in the spontaneous trait inference tradition to not only diversify their stimuli but also to actively engage with scholars of color. We are calling for an expansion of the indigenous psychology approach (Allwood & Berry, 2006; Berry & Kim, 1993) to include not only broad cultures and language, but also to recognize the critical importance of differences in lived experiences even within cultures. Indigenous psychology elevates the knowledge and beliefs that people have about themselves and their experiences as starting points for inquiry. To the extent that the diverging lived experiences of scholars of color and White scholars promote different perspectives on how to interpret behavior, bringing more scholars of color into the spontaneous trait inference research community will facilitate new discoveries about the mechanisms by which behavioral inference shapes impressions (see also, Ledgerwood et al., 2022, 2021).

On a final, personal note, we believe that diversifying the researchers is not solely driven by the content of the research. Author KM, an African American man, met Jim Uleman as an advanced graduate student in 1996 while attending the Society for Experimental Social Psychology (SESP) Conference. Even though Black people and members of other racial and ethnic minorities were severely underrepresented at the conference, I will always remember that SESP as the time, early in my career, that I started to feel a part of the field. This was in large part due to Jim. We were among a large group of primarily senior faculty members who went out for dinner at a restaurant. KM sat next to Jim the entire evening chatting one-on-one at times, and in the larger group. Jim showed genuine interest in me and in my work that was at times intimidating, but also comforting and validating. As a graduate student, I had some concerns about keeping the bill manageable but ultimately decided not to worry too much about it. But for those of us who have been in that situation with little disposable income, it's hard not to be a little preoccupied throughout the evening before the arrival of the check. And as the only person of color in the group, this preoccupation was even more intense as I considered the negative lens through which others might see me were I to speak out, balanced with the financial implications if

I didn't. When the moment arrived and the check finally came, the dreaded recommendation to split the bill evenly among our party was announced. Instead, Jim suggested that the faculty split the bill and buy dinner for the graduate students. I will never forget that he did that, and the tremendous relief from the burden that I felt. Reflecting on that time, my concerns may have been exaggerated in my mind, but at the time I lacked that perspective. Since then, I've tried to do the same, whenever possible, for others who might have similar concerns in these situations like that one. But *before* we order, and with explicit reference to the kind of rumination I experienced back then. Doing so makes explicit my understanding of their potential concerns and takes a bit of the edge off a potentially threatening experience.

Conclusion

There are several structural and interpersonal barriers to the inclusion and representation of underrepresented minority scholars in social cognition that we absolutely must address to fulfill the promise of a broad and generalization theory of human thought and behavior (Ledgerwood et al., 2021). Awareness of these challenges and good intentions to enact change are not enough. We need scholars who will channel awareness into urgency, urgency into intent, and intent into action. However, we believe that diversifying the field would be a little bit easier with more inquisitive, clever, thoughtful, and compassionate scholars like Jim in it.

Acknowledgments

The authors are grateful to the editors and to Dave Hamilton for their feedback on earlier drafts of this chapter. We would also like to thank Abigail Pomeranz and Jailekha Zutshi for editorial assistance.

Note

1 The focus of our analysis is to examine the potential interplay between behavior-based and social group-based inferences about a target individual. For another literature that compares and contrasts differences in the perception of *individual* targets with *group* targets, see Hamilton and Sherman (1996).

References

Adams, E. A., Kurtz-Costes, B. E., & Hoffman, A. J. (2016). Skin tone bias among African Americans: Antecedents and consequences across the life span. *Developmental Review*, *40*, 93–116. 10.1016/j.dr.2016.03.002

Allwood, C. M., & Berry, J. W. (2006). Origins and development of indigenous psychologies: An international analysis. *International Journal of Psychology*, *41*(4), 243–268. 10.1080/00207590544000013

Arnett, J. J. (2008). The neglected 95%: Why American psychology needs to become less American. *American Psychologist*, *63*, 602. 1037/0003-066X.63.7.602

Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology*, *41*, 258–290. 10.1037/h0055756

Bailey, A. H., LaFrance, M., & Dovidio, J. F. (2020). Implicit androcentrism: Men are human, women are gendered. *Journal of Experimental Social Psychology*, *89*, 103980. doi: 10.1016/j.jesp.2020.103980

Berry, J. W., & Kim, U. (1993). The way ahead: From indigenous psychologies to a universal psychology. In U. Kim & J. W. Berry (Eds.), *Cross-cultural research and methodology series, Vol. 17. Indigenous psychologies: Research and experience in cultural context* (pp. 277–280). SAGE.

Brewer, M. B. (1979). In-group bias in the minimal intergroup situation: A cognitive-motivational analysis. *Psychological Bulletin*, *86*(2), 307–324. 10.1037/0033-2909.86.2.307

Brewer, M. B. (1988). A dual-process model of impression formation. In R. S. Wyer, Jr. & T. K. Srull (Eds.), *Advances in social cognition* (Vol. 1, pp. 1–36). Erlbaum.

Chen, J. M., & Hamilton, D. L. (2012). Natural ambiguities: Racial categorization of multiracial individuals. *Journal of Experimental Social Psychology*, *48*(1), 152–164. 10.1016/j.jesp.2011.10.005

Chen, J. M., Moons, W. G., Gaither, S. E., Hamilton, D. L., & Sherman, J. W. (2014). Motivation to control prejudice predicts categorization of multiracials. *Personality and Social Psychology Bulletin*, *40*(5), 590–603. 10.1177/0146167213520457

Chen, J. M., Norman, J. B., & Nam, Y. (2021). Broadening the stimulus set: Introducing the American multiracial faces database. *Behavior Research Methods*, *53*(1), 371–389. 10.3758/s13428-020-01447-8

Chen, J. M., Pauker, K., Gaither, S. E., Hamilton, D. L., & Sherman, J. W. (2018). Black+ White= Not White: A minority bias in categorizations of Black–White multiracials. *Journal of Experimental Social Psychology*, *78*, 43–54. 10.1016/j.jesp.2018.05.002

Cheon, B. K., Melani, I., & Hong, Y. (2020). How USA-centric is psychology? An archival study of implicit assumptions of generalizability of findings to human nature based on origins of study samples. *Social Psychological and Personality Science*, *11*(7), 928–937. 10.1177/1948550620927269

Choi, I., & Nisbett, R. E. (1998). Situational salience and cultural differences in the correspondence bias and actor-observer bias. *Personality and Social Psychology Bulletin*, *24*(9), 949–960. 10.1177%2F0146167298249003

Cloutier, J., Mason, M. F., & Macrae, C. N. (2005). The perceptual determinants of person construal: Reopening the social-cognitive toolbox. *Journal of Personality and Social Psychology*, *88*, 885–894. 10.1037/0022-3514.88.6.885

Cole, E. R. (2009). Intersectionality and research in psychology. *American Psychologist*, *64*, 170–180. 10.1037/a0014564

Cone, J., Mann, T. C., & Ferguson, M. J. (2017). Changing our implicit minds: How, when, and why implicit evaluations can be rapidly revised. In J. M. Olson (Ed.), *Advances in experimental social psychology* (Vol. 56, pp. 131–199). Academic Press. 10.1016/bs.aesp.2017.03.001

Crocker, J., Voelkl, K., Testa, M., & Major, B. (1991). Social stigma: The affective consequences of attributional ambiguity. *Journal of Personality and Social Psychology*, *60*, 218–228. 10.1037/0022-3514.60.2.218

Fein, S. (1996). Effects of suspicion on attributional thinking and the correspondence bias. *Journal of Personality and Social Psychology*, *70*, 1164–1184. 10.1037/0022-3514.70.6.1164

Fiske, S. T., & Neuberg, S. L. (1990). A continuum model of impression formation from category-based to individuated processes: Influences of information and motivation on attention and interpretation. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 1–74). Academic Press. 10.1016/S0065-2601(08)60317-2

Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review*, *118*, 247–279. 10.1037/a0022327

Freeman, J. B., Penner, A. M., Saperstein, A., Scheutz, M., & Ambady, N. (2011). Looking the part: Social status cues shape race perception. *PloS one*, *6*(9), e25107. 10.1371/journal.pone.0025107

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*, 692–731. 10.1037/0033-2909.132.5.692

Gawronski, B., & Quinn, K. A. (2013). Guilty by mere similarity: Assimilative effects of facial resemblance on automatic evaluation. *Journal of Experimental Social Psychology*, *49*, 120–125. 10.1016/j.jesp.2012.07.016

Gilbert, D. T., Pelham, B. W., & Krull, D. S. (1988). On cognitive busyness: When person perceivers meet persons perceived. *Journal of Personality and Social Psychology*, *54*(5), 733–740. 10.1037/0022-3514.54.5.733

Gonzalez, C. M., Todorov, A., Uleman, J. S., & Thaden, E. P. (forthcoming). *A dissociation between spontaneous and intentional stereotyped trait inferences*. Unpublished manuscript, New York: New York University.

Hamilton, D. L., Chen, J. M., Ko, D. M., Winczewski, L., Banerji, I., & Thurston, J. A. (2015). Sowing the seeds of stereotypes: Spontaneous inferences about groups. *Journal of Personality and Social Psychology*, *109*(4), 569–588. 10.1037/pspa0000034

Hamilton, D. L., & Sherman, S. J. (1996). Perceiving persons and groups. *Psychological Review*, *103*(2), 336–355. 10.1037/0033-295X.103.2.336

Hamilton, D. L., & Stroessner, S. J. (2021). *Social cognition: Understanding people and events*. SAGE.

Hastie, R., Ostrom, T. M., Ebbesen, E. B., Wyer, R. S., Jr., Hamilton, D. L., & Carlston, D. E. (1980). *Person memory: The cognitive basis of social perception*. Erlbaum.

Heider, F. (1958). *The psychology of interpersonal relations*. Wiley.

Hinzman, L., & Maddox, K. B. (2017). Conceptual and visual representations of racial categories: Distinguishing subtypes from subgroups. *Journal of Experimental Social Psychology*, *70*, 95–109. doi: 10.1016/j.jesp.2016.12.012

Johnson, K. L., Freeman, J. B., & Pauker, K. (2012). Race is gendered: How covarying phenotypes and stereotypes bias sex categorization. *Journal of Personality and Social Psychology*, *102*, 116–131. 10.1037/a0025335

Jones, B. C., DeBruine, L. M., Flake, J. K., Liuzza, M. T., Antfolk, J., Arinze, N. C., … & Sirota, M. (2021). To which world regions does the valence–dominance model of social perception apply? *Nature Human Behaviour*, *5*, 159–169. 10.6084/m9.figshare.7611443.v1

Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska symposium on motivation* (Vol. 15, pp. 192–238). University of Nebraska Press.

Ledgerwood, A., da Silva Frost, A, Kadirvel, S., Maitner, A. T., Wang, Y. A., & Maddox, K. B. (2021). Methods for advancing an open, replicable, and inclusive science of social cognition. Chapter to appear. In D. Carlston, K. Johnson, & K. Hugenberg (Eds.), *Oxford handbook of social cognition*. Oxford University Press.

Ledgerwood, A., Hudson, S. T. J., Lewis, N.A., Jr., Maddox, K. B., Pickett, C. L., Remedios, J. D., Cheryan, S., Diekman, A. B., Dutra, N. B., Goh, J. X., Goodwin, S. A., Munakata, Y., Navarro, D. J., Onyeador, I. N., Srivastava, S., & Wilkins, C. L. (2022). The pandemic as a portal: Reimagining psychological science as truly open and inclusive. Perspectives on Psychological Science, 174569162110366. doi: 10.31234/osf.io/gdzue

Lee, H., Shimizu, Y., Masuda, T., & Uleman, J. S. (2017). Cultural differences in spontaneous trait and situation inferences. *Journal of Cross-Cultural Psychology*, *48*(5), 627–643. 10.1177/0022022117699279

Lee, H., Shimizu, Y., & Uleman, J. S. (2015). Cultural differences in the automaticity of elemental impression formation. *Social Cognition*, *33*(1), 1–19. 10.1521/soco.2015.33.1.1

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135. 10.3758/s13428-014-0532-5

Ma, D. S., Kantner, J., & Wittenbrink, B. (2021). Chicago face database: Multiracial expansion. *Behavior Research Methods*, *53*(3), 1289–1300. doi: 10.3758/s13428-020-01482-5

Macrae, C. N., Quinn, K. A., Mason, M. F., & Quadflieg, S. (2005). Understanding others: The face and person construal. *Journal of Personality and Social Psychology*, *89*, 686–695. 10.1037/0022-3514.89.5.686

Maddox, K. B. (2004). Perspectives on racial phenotypicality bias. *Personality and Social Psychology Review*, *8*(4), 383–401. 10.1207/s15327957pspr0804_4

Maddox, K. B., & Dukes, K. N. (2008). Social categorization and beyond: How facial features impact social judgment. In N. Ambady & J. J. Skowronski (Eds.), *First impressions* (pp. 205–233). Guilford.

Maddox, K. B., & Gray, S. A. (2002). Cognitive representations of Black Americans: reexploring the role of skin tone. *Personality and Social Psychology Bulletin*, *28*(2), 250–259. doi: 10.1177/0146167202282010

Major, B., Kunstman, J. W., Malta, B. D., Sawyer, P. J., Townsend, S. S., & Mendes, W. B. (2016). Suspicion of motives predicts minorities' responses to positive feedback in interracial interactions. *Journal of Experimental Social Psychology*, *62*, 75–88. 10.1016/j.jesp.2015.10.007

Malle, B. F., & Holbrook, J. (2012). Is there a hierarchy of social inferences? The likelihood and speed of inferring intentionality, mind, and personality. *Journal of Personality and Social Psychology*, 102(4), 661–684. 10.1037/a0026790

Mann, T. C., & Ferguson, M. J. (2017). Reversing implicit first impressions through reinterpretation after a two-day delay. *Journal of Experimental Social Psychology*, *68*, 122–127. 10.1016/j.jesp.2016.06.004

Masuda, T., & Kitayama, S. (2004). Perceiver-induced constraint and attitude attribution in Japan and the US: A case for the cultural dependence of the correspondence bias. *Journal of Experimental Social Psychology*, *40*(3), 409–416. 10.1016/j.jesp.2003.08.004

Maurer, K. L., Park, B., & Rothbart, M. (1995). Subtyping versus subgrouping processes in stereotype representation. *Journal of Personality and Social Psychology*, *69*(5), 812–824. 10.1037/0022-3514.69.5.812

McCarthy, R., Gervais, W., Aczel, B., Al-Kire, R. L., Aveyard, M., Baraldo, S. M., … & Zogmaister, C. (2021). A multi-site collaborative study of the hostile priming effect. *Collabra: Psychology*, *7*(1), 18738. 10.1525/collabra.18738

McCarthy, R. J., Skowronski, J. J., Verschuere, B., Meijer, E. H., Jim, A., Hoogesteyn, K., … & Yıldız, E. (2018). Registered replication report on Srull and Wyer (1979). *Advances In Methods and Practices in Psychological Science*, *1*(3), 321–336. 10.1177/2515245918777487

McConnell, A. R., Sherman, S. J., & Hamilton, D. L. (1994). On-line and memory-based aspects of individual and group target judgments. *Journal of Personality and Social Psychology*, *67*, 173–185. 10.1037/0022-3514.67.2.173

McConnell, A. R., Sherman, S. J., & Hamilton, D. L. (1997). Target entitativity: Implications for information processing about individual and group targets. *Journal of Personality and Social Psychology*, *72*, 750–762. 10.1037/0022-3514.72.4.750

Miyamoto, Y., & Kitayama, S. (2002). Cultural variation in correspondence bias: The critical role of attitude diagnosticity of socially constrained behavior. *Journal of Personality and Social Psychology*, *83*(5), 1239–1248. 10.1037/0022-3514.83.5.1239

Moskowitz, G. B. (1993). Individual differences in social categorization: The influence of personal need for structure on spontaneous trait inferences. *Journal of Personality and Social Psychology*, *65*(1), 132–412. 10.1037/0022-3514.65.1.132

Moskowitz, G. B., Gollwitzer, P. M., Wasel, W., & Schaal, B. (1999). Preconscious control of stereotype activation through chronic egalitarian goals. *Journal of Personality and Social Psychology*, *77*, 167–184. 10.1037/0022-3514.77.1.167

Moskowitz, G. B., & Li, P. (2011). Egalitarian goals trigger stereotype inhibition: A proactive form of stereotype control. *Journal of Experimental Social Psychology*, *47*(1), 103–116. 10.1016/j.jesp.2010.08.014

Murphy, M. C., Mejia, A. F., Mejia, J., Yan, X., Cheryan, S., Dasgupta, N., … & Harackiewicz, J. M. (2020). Open science, communal culture, and women's participation in the movement to improve science. *Proceedings of the National Academy of Sciences*, *117*(39), 24154–24164. 10.1073/pnas.1921320117

Na, J., & Kitayama, S. (2011). Spontaneous trait inference is culture-specific: Behavioral and neural evidence. *Psychological science*, *22*(8), 1025–1032. 10.1177%2F0956797611414727

Olcaysoy Okten, I., & Moskowitz, G. B. (2018). Goal versus trait explanations: Causal attributions beyond the trait-situation dichotomy. *Journal of Personality and Social Psychology*, *114*(2), 211–229. 10.1037/pspa0000104

Olcaysoy Okten, I., & Moskowitz, G. B. (2020). Spontaneous goal versus spontaneous trait inferences: How ideology shapes attributions and explanations. *European Journal of Social Psychology*, *50*(1), 177–188. 10.1002/ejsp.2611

Oldmeadow, J., & Fiske, S. T. (2007). System-justifying ideologies moderate status = competence stereotypes: Roles for belief in a just world and social dominance orientation. *European Journal of Social Psychology*, *37*(6), 1135–1148. 10.1002/ejsp.428

Otten, S., & Moskowitz, G. B. (2000). Evidence for Implicit Evaluative In-Group Bias: Affect-Biased Spontaneous Trait Inference in a Minimal Group Paradigm. *Journal of Experimental Social Psychology*, *36*(1), 77–89. 10.1006/jesp.1999.1399

Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, *75*(3), 811–832. 10.1037/0022-3514.75.3.811

Quinn, K. A., & Macrae, C. N. (2005). Categorizing others: The dynamics of person construal. *Journal of Personality and Social Psychology*, *88*, 467–479. 10.1037/0022-3514.88.3.467

Quinn, K. A., Mason, M. F., & Macrae, C. N. (2009). Familiarity and person construal: Individual knowledge moderates the automaticity of category activation. *European Journal of Social Psychology*, *39*, 852–861. 10.1002/ejsp.596

Quinn, K. A., Mason, M. F., & Macrae, C. N. (2010). When Arnold is "The Terminator," we no longer see him as a man: The temporal determinants of person perception. *Experimental Psychology*, *57*, 27–35. 10.1027/1618-3169/a000004

Rhee, E., Uleman, J. S., & Lee, H. K. (1996). Variations in collectivism and individualism by ingroup and culture: Confirmatory factor analysis. *Journal of Personality and Social Psychology*, *71*, 1037–1054. 10.1037/0022-3514.71.5.1037

Rhee, E., Uleman, J. S., Lee, H. K., & Roman, R. J. (1995). Spontaneous self-descriptions and ethnic identities in individualistic and collectivistic cultures. *Journal of Personality and Social Psychology*, *69*, 141–152. 10.1037/0022-3514.69.1.142

Roberts, S. O., Bareket-Shavit, C., Dollins, F. A., Goldie, P. D., & Mortenson, E. (2020). Racial Inequality in Psychological Research: Trends of the Past and Recommendations for the Future. *Perspectives on Psychological Science*, *15*(6), 1295–1309. 10.1177/1745691620927709

Sherman, J. W., Lee, A. Y., Bessenoff, G. R., & Frost, L. A. (1998). Stereotype efficiency reconsidered: Encoding flexibility under cognitive load. *Journal of Personality and Social Psychology*, *75*, 589–606. 10.1037/0022-3514.75.3.589

Sanbonmatsu, D. M., Sherman, S. J., & Hamilton, D. L. (1987). Illusory correlation in the perception of individuals and groups. *Social Cognition*, *5*(1), 1–25. 10.1521/soco.1987.5.1.1

Shimizu, Y., Lee, H., & Uleman, J. S. (2017). Culture as automatic processes for making meaning: Spontaneous trait inferences. *Journal of Experimental Social Psychology*, *69*, 79–85. 10.1016/j.jesp.2016.08.003

Shimizu, Y., & Uleman, J. S. (2021). Attention allocation is a possible mediator of cultural variations in spontaneous trait and situation inferences: Eye-tracking evidence. *Journal of Experimental Social Psychology*, *94*, 104–115. 10.1016/j.jesp.2021.104115

Smith, E. R., & Miller, F. D. (1983). Mediation among attributional inferences and comprehension processes: Initial findings and a general method. *Journal of Personality and Social Psychology*, *44*(3), 492–505. 10.1037/0022-3514.44.3.492

Spencer, S. J., Fein, S., Wolfe, C. T., Fong, C., & Duinn, M. A. (1998). Automatic activation of stereotypes: The role of self-image threat. *Personality and Social Psychology Bulletin*, *24*, 1139–1152. 10.1177/01461672982411001

Srull, T. K., & Wyer, R. S., Jr. (1979). The role of category accessibility in the interpretation of information about persons: Some determinants and implications. *Journal of Personality and Social Psychology*, *37*(10), 1660–1672. 10.1037/0022-3514.37.10.1660

Stewart, T. L., Doan, K. A., Gingrich, B. E., & Smith, E. R. (1998). The actor as context for social judgments: Effects of prior impressions and stereotypes. *Journal of Personality and Social Psychology*, *75*(5), 1132–1154. 10.1037/0022-3514.75.5.1132

Stroessner, S. J. (1996). Social categorization by race or sex: Effects of perceived nonnormalcy on response times. *Social Cognition*, *14*(3), 247–276. 10.1521/soco.1996.14.3.247

Syed, M., & Kathawalla, U. (2020, February 25). Cultural psychology, diversity, and representation in open science. doi: 10.31234/osf.io/t7hp2

Todorov, A., & Uleman, J. S. (2002). Spontaneous trait inferences are bound to actors' faces: Evidence from a false recognition paradigm. *Journal of Personality and Social Psychology*, *83*, 1051–1065. 10.1037/0022-3514.83.5.1051

Todorov, A., & Uleman, J. S. (2003). The efficiency of binding spontaneous trait inferences to actors' faces. *Journal of Experimental Social Psychology*, *39*, 549–562. 10.1016/S0022-1031(03)00059-3

Todorov, A., & Uleman, J. S. (2004). The person reference process in spontaneous trait inferences. *Journal of Personality and Social Psychology*, *87*, 482–493. 10.1037/0022-3514.87.4.482

Trope, Y. (1986). Identification and inferential processes in dispositional attribution. *Psychological Review*, *93*(3), 239–257. doi: 10.1037/0033-295X.93.3.239

Uleman, J. S. (1987). Consciousness and control: The case of spontaneous trait inferences. *Personality and Social Psychology Bulletin*, *13*(3), 337–354. 10.1177/0146167287133004

Uleman, J. S., & Moskowitz, G. B. (1994). Unintended effects of goals on unintended inferences. *Journal of Personality and Social Psychology*, *66*, 490–501. 10.1037/0022-3514.66.3.490

Uleman, J. S., Newman, L. S., & Winter, L. (1992). Can personality traits be inferred automatically? Spontaneous inferences require cognitive capacity at encoding. *Consciousness and Cognition*, *1*, 77–90. 10.1016/1053-8100(92)90049-G

Uleman, J. S., Rim, S., Adil Saribay, S., & Kressel, L. M. (2012). Controversies, questions, and prospects for spontaneous social inferences. *Social and Personality Psychology Compass*, *6*(9), 657–673. 10.1111/j.1751-9004.2012.00452.x

Van Overwalle, F., Van Duynslaeger, M., Coomans, D., & Timmermans, B. (2012). Spontaneous goal inferences are often inferred faster than spontaneous trait inferences. *Journal of Experimental Social Psychology*, *48*(1), 13–18. 10.1016/j.jesp.2011.06.016

Verosky, S. C., & Todorov, A. (2010). Generalization of affective learning about faces to perceptually similar faces. *Psychological Science*, *21*, 779–785. 10.1177/0956797610371965

Wigboldus, D. H., Dijksterhuis, A., & van Knippenberg, A. (2003). When stereotypes get in the way: Stereotypes obstruct stereotype-inconsistent trait inferences. *Journal of Personality and Social Psychology*, *84*(3), 470–484. 10.1037/0022-3514.84.3.470

Winter, L., & Uleman, J. S. (1984). When are social judgments made? Evidence for the spontaneousness of trait inferences. *Journal of Personality and Social Psychology*, *47*, 237–252. 10.1037/0022-3514.47.2.237

Winter, L., Uleman, J. S., & Cunniff, C. (1985). How automatic are social judgments? *Journal of Personality and Social Psychology*, *49*, 904–917. 10.1037/0022-3514.49.4.904

Wyer, N. A. (2010). You never get a second chance to make a first (implicit) impression: The role of elaboration in the formation and revision of implicit impressions. *Social Cognition*, *28*, 1–19. 10.1521/soco.2010.28.1.1

Wyer, N. A. (2016). Easier done than undone…by some of the people, some of the time: The role of elaboration in explicit and implicit group preferences. *Journal of Experimental Social Psychology*, *63*, 77–85. 10.1016/j.jesp.2015.12.006

Xie, S. Y., Flake, J. K., & Hehman, E. (2019). Perceiver and target characteristics contribute to impression formation differently across race and gender. *Journal of Personality and Social Psychology*, *117*(2), 364–385. 10.1037/pspi0000160

Yan, X., Wang, M., & Zhang, Q. (2012). Effects of gender stereotypes on spontaneous trait inferences and the moderating role of gender schematicity: Evidence from Chinese undergraduates. *Social Cognition*, *30*(2), 220–231. 10.1521/soco.2012.30.2.220

Zarate, M. A., & Smith, E. R. (1990). Person categorization and stereotyping. *Social cognition*, *8*(2), 161–185. 10.1521/soco.1990.8.2.161

7 The Secret Life of Spontaneous Trait Inferences: Emergence, Puzzles, and Accomplishments

Leonel Garcia-Marques[1], *Mário B. Ferreira*[1], *Sara Hagá*[1], *Daniel Marcelo*[2], *Tânia Ramos*[3], *and Diana Orghian*[4]

[1]*Center for Research in Psychological Science (CICPsi), Faculdade de Psicologia, Universidade de Lisboa, Portugal*
[2]*Backend Developer, Oswald Digital, Lisboa, Portugal*
[3]*Lead UX Researcher, Springer Nature Group - SN Digital, Lisbon, Portugal*
[4]*Lead User Experience Researcher, Feedzai, Lisbon Office, Portugal*

It may strike some readers as ironic that the concept of *trait*, or more precisely the concept of *perceived trait*, has pervaded research in impression formation, person memory, and stereotypes in social psychology (e.g., Moskowitz & Olcaysoy Okten, 2016). The irony derives from the position, often advanced by important authors, that the principal message of our discipline is that individual differences provide weak accounts of behavior, at best, and that situations are much more powerful in that regard (Ross & Nisbett, 1991). Thus, the need for studying the way we perceive traits in others was not readily apparent in the development of social psychology. Nevertheless, we argue that the study of spontaneous trait inferences (STIs) was a missing piece of the puzzle of impression formation and attribution—a kind of secret ingredient that everybody felt was absent, but that nobody could put their finger on it. The literature focused on how impressions were organized around existing traits and behaviors, and how traits cohered together to form narratives and implicit personality theories. Yet the question of where the inferences comprising the impressions came from was missing.

This chapter is organized as follows. After painting, in broad strokes, the context in which the study of STIs emerged, we present a brief critique (i.e., main advantages and limitations) of the two kinds of research paradigms that have helped STIs researchers to further probe STI in creative ways: online probe recognition and memory-based procedures. We then describe in more detail the contributions made by research from our own labs, including previous findings (e.g., Ferreira et al., 2012; Ramos et al., 2012), and the recent proposal of two new ways to approach the STI debate. One is the connectionist model of associative trait inference and trait transference

DOI: 10.4324/9781003045687-8

(MATIT; Orghian et al., 2015). The other is a new experimental paradigm to study STI (Orghian et al., 2017). We finish with a discussion on new research paths on STIs that we are currently pursuing, which emphasize the inspiring role of text comprehension literature. More specifically, we argue that the minimalist hypothesis of McKoon and Ratcliff (1992) and formal language analysis may provide critical inputs into the understanding of STI.

Forming Impressions of Personality: Inferring Traits from Inferred Traits

Asch, a pioneer of impression formation, was fascinated by the spontaneity and ease of trait inference. In 1946, Asch famously wrote:

> We look at a person and immediately a certain impression of his character forms itself in us. A glance, a few spoken words are sufficient to tell us a story about a highly complex matter. We know that such impressions form with remarkable rapidity and with great ease. (p. 258)

However, although Asch's words have often been quoted in the social psychology literature, social psychologists went on to explore slightly different problems.

Indeed, Asch asked remarkable research questions and applied experimental methods to the overly complex matter of how people combine traits and integrate them in gestalt-like impressions. For instance, in Asch's studies (1946), participants easily formed a complete and cohesive personality impression of someone described by a list of traits. Moreover, switching just one trait (e.g., warm) for its opposite (e.g., cold) in a list of seven, dramatically changed the formed impressions. Asch's work paved the way for researchers, who explored the main empirical effects that emerged from those studies and tinkered with detailed different accounts of them. But impression formation would feel incomplete without the grasp of the process by which trait inferences occur in the first place, and STIs would be able to keep the secrecy of their life for some more time.

In a natural extension of Asch's work, researchers brought to light the implicit personality theories that people have (e.g., Rosenberg et al., 1968). These theories refer to the beliefs that common people hold about how personality traits relate to each other (Bruner & Tagiuri, 1954). By looking into a variety of such beliefs, this research revealed that one uses something like a map to make sense of other people (Rosenberg et al., 1968). However, instead of using north-south and east-west dimensions for orientation, people use good-bad intellectual and good-bad social dimensions for placing others. This means that knowing that someone is imaginative—a desirable intellectual trait—allows one to place this person towards the good pole of the intellectual dimension and infer other traits that share this space, like intelligent. This common structure of the implicit personality theories allows

people to infer traits from other traits. But how are the initial traits inferred and how is the implicit trait theory network formed to begin with?

Forming Impressions versus Memory: Explaining the Unknown with Another Unknown

In another extension of Asch's work, the person memory literature showed that when participants are asked to form impressions of someone from a list of behaviors (with no mention of a future memory test) or to memorize the same list, performance at a recall test is better in the case of impression formation relative to the memory condition. Participants not only recalled more items (i.e., behaviors), they also showed more evidence of trait clustering—grouping behaviors by their underlying trait. The classic literature in person memory went on to suggest that these differences in recall were due to differences in the likelihood of trait inference: A process that is much more likely when participants are asked to form impressions (Hamilton et al., 1980; Klein & Loftus, 1990).

Thus, although trait inference was not directly studied in this literature, it was supposedly the critical signature of impression formation. Trait inference was, therefore, given a crucial role even before its underlying process was pinpointed or circumscribed. Aristotelian Physics was famously criticized by Boyle (1666) for *ignotum per æque ignotum* (explaining the unknown for the equally unknown). The role of trait inference in explaining differences in recall performance between impression and memory conditions apparently followed this illustrious precedent.

Attribution: When What Precedes Why

In a parallel area of study, namely causal attribution, the focus was on how people's behaviors are accounted for. So, understandably, the first theories of how one infers invariant internal dispositions of people from observing the continuously varying stream of their behaviors emerged in that area of study (Heider, 1958). Although Heider (1958) was mainly concerned with motives, intentions, and sentiments (even if he sometimes referred to traits and skills), his followers almost always used dispositions as a proxy for traits (e.g., Moskowitz & Olcaysoy Okten, 2016). The theory of correspondent inference (Jones & Davis, 1965) described how perceivers move from the observation of a behavior to the inference of stable individual dispositions (or traits). This model posited that, in order to conclude that a behavior reflected the actor's traits, alternative situational accounts must be discarded, and a number of other pre-conditions met. Thus, this approach was hard to reconcile with Asch's intuition that people infer traits with great ease and spontaneity.

But more problematically for this theory, people seem too ready to infer traits and too often with little consideration for the situational circumstances in

which the behavior occurred. This propensity was dubbed the correspondence bias (Gilbert & Malone, 1995). The correspondence bias was demonstrated in numerous studies, even in cases in which participants were the direct instigators and regulators of the actor's behavior (Gilbert & Jones, 1986).

The surprising ubiquity of correspondent (trait) inference was even more puzzling, when considered conjointly with main findings of the spontaneous attribution literature. According to this literature, the specific conditions that trigger causal attribution search are only rarely met in daily life and therefore attributions should only seldomly occur (Enzle & Schopflosher, 1978; Hastie, 1984; Lau & Russell, 1980; Pyszczynski & Greenberg, 1981; Wong & Weiner, 1981). Thus, a trait inference model that could explain how correspondent inference can recurrently occur, requiring apparently little information or effort, and even when no one was seemingly trying to account for a behavior or an event, seems to be resoundingly missing from these literatures. Attribution research authors would soon have this puzzle solved by the addition of one missing piece—the enigmatic STIs.

Text Comprehension Comes to Rescue

Smith and Miller (1979a, 1979b, 1983) examined the ANOVA causal attribution model of Kelley (Kelley, 1967, 1973, Orvis et al., 1975) and showed that participants were faster to answer to trait attribution queries than to causal locus queries (in contrast with classical attribution models). In this research, Smith & Miller (1983) imported theoretical and methodological approaches from text comprehension both to inform their research methods and to account for their findings. Research in text comprehension suggested people routinely make inferences while reading or listening to a text, without being required to do so, without necessarily realizing they are doing so, and even without needing to invest considerable cognitive resources while doing so. In text comprehension, these spontaneous causal inferences occur to aid people's comprehension of oral or written text, in their effort after meaning (e.g., Kintsch, 1974, 1975; Norman, Rumelhart, & the LNR Research Group, 1975; Schank, 1975). Accordingly, inferences from the text, including trait inferences, presumably help people understand others and their behaviors.

Based on these assumptions, Winter and Uleman (1984) developed a paradigm that combined text comprehension research and memory research to study trait inference. In this paradigm, participants read a list of trait implicative sentences with the ostensible goal of learning them for a memory test. These sentences were descriptions of behaviors that implied traits. At test, participants were provided different extra-list retrieval cues: a word associated with the actor of the behavior, with the action verb, or the implied trait. Compared with the no-cued sentences, sentences cued with traits were more often recalled. Notably, these sentences were even recalled at least as often as sentences provided with other cues. This result supposedly illustrates Tulving's encoding specificity principle (Tulving & Thomson, 1973), according to

which an effective retrieval cue for a piece of information is another piece of information that was encoded at the same time, in the same context. Thus, the effectiveness of trait cues was due to trait inferences that occurred during the original encoding of the sentences (but see D'Agostino & Beegle, 1996).

Shortly after STIs had come to light, a trait inference model was proposed—the categorization-characterization-correction model (Gilbert et al., 1988). This model came from the lineage of the theory of correspondent inferences (Jones & Davis, 1965), but simplified the parent theory, reversed the order between consideration of situational constraints and trait inferences (following Quattrone, 1982), and followed the guide of STI by turning the initial steps in the attributional path to automatic, effortless, and unconscious trait inferences. By suggesting that social perceivers first attribute behaviors to traits of the actors and only later, more deliberately, correct these assumptions with other kinds of information, this model succeeded in explaining the correspondence bias.

However, how does it make sense to combine a frequent process that begins by an initial trait inference to seldom occurring processes of attributional search? The STIs that occur naturally were apparently the missing ingredient. Attributional search would only be triggered in specific conditions but, more often than not, after a trait inference drawn for the sake of meaning construal was already available. The availability of a previous (implicitly drawn) trait inference allowed social perceivers to save time and resources from their cognitively busy life (Gilbert, 1998). Attributional search was no longer seen as a dedicated ad hoc process, but as a process that worked in tandem with more general meaning construal processes.

Thus, one might say that the suggesting and unveiling of the existence of STIs was the secret ingredient, in a mix of other ingredients, to make sense of Asch's (1946) original intuitions: "We look at a person and immediately a certain impression of his character forms itself in us" (p. 258) because we spontaneously, quickly, and easily infer traits from this person's behaviors (i.e., we draw STIs). We do this to better understand what this person is doing and why (i.e., similar to effort for meaning in text comprehension). Once traits are inferred, we use them to organize the information we commit to memory about this person, and we have a whole set of other traits ready to be inferred in order to complete an impression (i.e., we use implicit personality theories). Finally, we may, or we may not, shape these trait inferences by taking the circumstances in which the behavior occurred into account (i.e., we consider causes that may not lie within the actor).

Intentional and Spontaneous Trait Inferences and a Solution for Apparently Contradictory Findings

As previously described, the study of STIs succeeded in bringing together different spheres of the literature. However, some divergences remained. More recently, a residual divergence between the person memory (Hamilton

et al., 1980; Klein & Loftus, 1990) and the STIs literature (Winter & Uleman, 1984; Winter et al., 1985) was shown to be more apparent than real. In fact, the differences in recall performance and trait clustering that occur between participants who form impressions of a target person from a list of behaviors and participants who memorize the same information seem to derive not from the existence of trait inferences per se, but from differences in monitoring of trait inferences. These differences in inference monitoring, in turn, allow for a greater awareness of the trait inferences made under an impression formation instructional set (see Ferreira et al., 2012).

To explore these hypotheses, in one of the experiments, Ferreira et al. (2012) combined the Winter and Uleman (1984) and Hamilton et al. (1980) studies into a single paradigm. Specifically, participants read a series of behaviors representing four different trait categories under memory or impression formation instructions. Later, participants recalled these behaviors. For half of the participants, the four traits were provided as memory cues during the recall task. We hypothesized that under an impression goal, participants should be more aware of the trait inferences made during the process of impression formation. This heightened awareness made them better able to effectively use traits as cues during retrieval (whether or not these traits are provided as cues). The same should not occur for memory goal participants—they were hypothesized to be able to use traits as retrieval cues only if they were provided with them at recall. Results showed that with no cues provided at recall, the usual superiority of impression goal relative to memory participants occurred both in recall and trait clustering, but these differences vanished when traits were provided as cues at recall. Ferreira et al. (2012) interpreted these results as indicating differences in awareness of trait inference between conditions of impression formation and memorization. Such differences were assumed to derive from disparities in inference monitoring that are critical to grasping what distinguishes intentional from spontaneous trait inference.

Thus, according to this inference monitoring hypothesis, spontaneous and intentional inferences share the same largely automatic inferential process that allows for the efficient extraction of traits from trait implying behaviors, but differ in the monitoring (i.e., attentional focus) on the outcomes of this initial process (i.e., the traits). Ferreira et al. (2012) tested the inference monitoring hypothesis in a set of studies. They used a forced recognition paradigm, in which participants picked the behavioral descriptions they had previously seen from pairs of descriptions that differed only in the explicit inclusion of the implied trait. Ferreira et al. (2012, Studies 2 and 3) then predicted that inference monitoring would facilitate forced recognition performance and indeed found that impression formation instructions (i.e., explicit inferences) as opposed to memory instructions reduced the proportion of false recognitions (i.e., recognizing traits that were implied but not present during encoding). In addition, using the process dissociation procedure (PDP; Jacoby, 1991), they showed that impression formation (versus

memory) instructions and cognitive overload affected trait inference monitoring (measured by the controlled component of the PDP) but left the automatic inferential processes unchanged. These results are consistent with the inference monitoring hypothesis and support the notion that STIs are implicit inferences often occurring outside awareness.

In sum, we argue that STIs were and are crucial for the convergence of the study of impression formation, attribution, person memory and trait inference processes. But as the secret life of STI went on being revealed, it presented researchers with further unexpected identity conundrums.[1]

Homer or a Greek with the Same Name

As a well-known anecdote goes, there was this historian who spent their life trying to demonstrate that the author of *Odyssey* was not Homer but another Greek author with the same name. Well, in the study of STIs, questions about Homer versus another Greek with the same name occurred often.

As we described earlier, Winter and Uleman (1984) presented the first STI paradigm. The critical feature of this paradigm was an imaginative application of the encoding specificity principle (Tulving & Thomson, 1973), according to which, the effectiveness of using traits as retrieval cues for the recall of trait implicative sentences was the occurrence of spontaneous trait inferences during the original encoding of the sentences. However, an equally plausible account of the results was simply to say that participants used the traits to probe their memory in search of good examples of sentences that implied the traits. Since the sentences used by Winter and Uleman (1984) were examples of the implied traits, the search was bound to be successful (Wyer & Srull, 1986). Furthermore, although the results obtained with this experimental paradigm may indeed suggest that trait terms are associated with previously learned behaviors, this does not necessarily imply that they are associated with the actors of these behaviors (Bassili, 1989; Carlston & Skowronski, 1994; Higgins & Bargh, 1987). The first of these issues (but not the second) was greatly overcome with the probe recognition paradigm (Uleman et al., 1996; first proposed by McKoon & Ratcliff, 1986). In this paradigm, participants are probed about the presence of a specific word immediately after the presentation of each trait-implicative sentence. When this word was the implied trait, performance deteriorated with longer response times and/or more errors because participants had to discriminate trait inferences from the presence of the traits in the sentences. This paradigm allowed the study and measurement of STIs in the moment they were being made, overcoming the possibility of a backward trait-sentence association as a strategy for sentence recall.

The second problem was solved by the saving in relearning paradigm (Carlston & Skowronski, 1994; Carlston et al., 1995). In this paradigm, after being exposed to pairs of trait-implicative descriptions and photos, participants are requested to learn several photo-trait pairs, involving the same

photos and traits previously implied or not by the descriptions. When the traits correspond to the traits implied in the initial descriptions, learning was facilitated. As the task implied a direct association between the trait and actor, the possibility of mere association between the trait and the sentence's action verb was discarded. The false memories paradigm later introduced by Todorov and Uleman (2002, 2004) avoided the same problem by presenting participants with pairs of photos and trait-implicative sentences (sometimes including the implied trait in the sentence). At test, participants saw the photo and the implied trait and responded whether the implied trait was included in the sentence previously paired with the photo. When the sentence did not include the implied trait, participants showed nevertheless a tendency to commit false memories by responding positively. However, as these two problems were solved, other problems emerged. First, the interpretation of results obtained through online paradigms such as the probe-recognition paradigm is susceptible to confounds resulting from word-based priming activation of the trait (e.g., Ham & Vonk, 2003; Van Overwalle et al., 1999). In other words, the apparent inference may be the result of processing specific words in the sentence that are individually associated with the trait (Keenan et al., 1990; Orghian et al., 2019). A possible way to avoid this confound is to include control sentences that contain roughly the same words as the trait-implying ones, but rearranged in such a way that the sentences as a whole no longer imply the trait (as Uleman et al., 1996 did in the original probe recognition STI studies).

Secondly, the memory-based paradigms, which were conceived to directly explore the target-trait inference link, also produced evidence of a possible association between the implied trait and non-target persons (e.g., bystanders or informants; Carlston et al., 1995; Goren & Todorov, 2009; Skowronski et al., 1998), or even objects present in the original encoding context (e.g., bananas; Brown & Bassili, 2002)—the so-called spontaneous trait transference effect. As a consequence, the nature of the link between the target and the trait became itself a new question of research because simple associative processes might explain the occurrence of STIs (Bassili, 1989; Brown & Bassili 2002). Carlston and Skowronski (2005; see also, Crawford, Skowronski, & Stiff, 2007; Crawford et al., 2008; Goren & Todorov, 2009) have claimed that, whereas the link between traits and appropriate actors entails the intervention of causal attributional processes, the link between traits and other irrelevant elements are dependent on simple associative processes or spontaneous trait transferences (STT). Three main types of evidence have been advanced to argue that STIs and STTs are underlined by different cognitive processes (Crawford, Skowronski, Stiff, & Scherer, 2007). First, STIs are generally stronger than STTs. Second, generalization (halo effects) to other personality traits is more likely in STI than in STT. Third, the finding that STTs are eliminated when the actor of the behavior is included in the same context as the communicator has been interpreted as a sign that attributional processes activated by the presence of the actor

preclude the establishment of arbitrary associations (Crawford, Skowronski, Stiff, & Schere, 2007; Goren & Todorov, 2009; Todorov & Uleman, 2004).

However, Orghian et al. (2015) proposed a connectionist model of associative trait inference and trait transference (MATIT) that was able to successfully simulate the three aforementioned empirical differences between STI and STT. MATIT simulates variations on the attention paid to the stimuli when the target is the actor of the trait implying behavior (relevant target) and when the target is the informant or bystander (irrelevant target) via activation weights of the links actor-behavior, actor-trait, and trait-behavior. Accordingly, telling participants that the person in the photo is the actor of the described behavior makes them pay more attention to the person, resulting in larger activation weights between actor, behavior, and trait. Conversely, if the person presented together with the behavior is said to be less relevant to the behavior, attention to this person is reduced, and the activation links are weaker. This difference in activation for relevant and irrelevant targets produces differences in the strength of associations between the behaviors and the target, and between the person and the trait implied by the behavior. Later in the test phase, when only persons (photos) and traits are presented, the resulting activation weights usually benefit STI more than STT.

However, as the authors noted: "We do not intend to (a) present a model that describes STI and STT phenomena in their intrinsic complexity; (b) explain all the differences between STI and STT; or (c) defend a single process view" (Orghian et al., 2015; p. 25). Instead, they intended to show that the evidence used to suggest the existence of two processes is easily reproduced by a simple and purely associative model. And they added:

> Our point was not that we were able to come up with an associative model that could explain previous results. After all, given some theoretical latitude and/or ad hockery, any type of model can simulate (mimic) any pattern of data (Anderson, 1978; Garcia-Marques & Ferreira, 2011). In that sense, finding a simulation model that simulates a data pattern is like fitting a statistical model. It will be a meaningless achievement unless the model can be falsified by plausible data (e.g., Roberts & Pashler, 2000). As it was demonstrated with MATIT, the advantage of using a simple (baseline) associative model is that, when it fits the data, it can provide clear guidelines for obtaining data that will challenge the model, that is, more diagnostic data. (p. 25)

All in all, MATIT's role is to suggest that clarifying the nature of the target-trait link is not as clear-cut as it may seem, it will always be dependent on the specification of the assumptions used and of the sophistication of the available theoretical models.

A final important methodological issue is the contamination problem (Jacoby, 1991). Specifically, STI measures are said to be contaminated when

the participants adopt an intentional retrieval strategy or if they are aware that the tested material is related to the studied material. All major STI paradigms, including the cued-recall (e.g., Winter & Uleman, 1984), the probe recognition (e.g., Uleman et al., 1996; Van Overwalle et al., 1999), the savings in relearning (e.g., Carlston & Skowronski, 1994) and the false recognition paradigm (Todorov and Uleman (2002), are vulnerable to the contamination problem to some degree. The more straightforward way to avoid contamination is to refrain from using memory-based paradigms (i.e., paradigms that require participants to recall past events). Instead, implicit tasks that make no reference to the study phase, and in which the explicit retrieval of the previous material does not benefit or influence performance in the task should be used.

Furthermore, given that a) the process taking place at encoding in the case of STI is conceptually driven (i.e., the meaning of a whole sentence has to be comprehended in order for the trait to be inferred); and b) there are no perceptual features at encoding of the trait to be inferred (i.e., the trait is only implied and is not physically presented), it follows that a conceptual task that only minimally depends of the perceptual features of the target will be a more sensitive measure of STI. With these concerns in mind, we proposed and developed a new STI paradigm.

The New Paradigm in the Block: Trait Word Association

Inspired by Hourihan and MacLeod's (2007) modified word association paradigm, Orghian et al. (2017) proposed a new paradigm that simultaneously overcomes the contamination problem of memory measures and the dependency on perceptual-driven processing.

Hourihan and MacLeod (2007) created the modified word association paradigm. This paradigm is meant to be a conceptually driven measure of implicit memory. In the learning phase, participants either generated words from meaningful cues (e.g., "the piece of furniture used for sitting—c?") or merely read the words (e.g., "chair"). In a test phase, participants performed a word association task where they said aloud the first word to come to mind upon sight of a prompt word that was either generated or read in the learning phase. At encoding, generating a word activates the lexical network relative to that word more than just reading the same word. During a critical period of time, presenting the generated word as a test prompt facilitates access to that lexical network relative to word prompts that were only read at encoding (Nelson & Goodmon, 2002). Translated to trait inference research, the premise is that reading a trait-implying sentence primes the implied trait word and its semantic neighbors. Thus, when the inferred trait is encountered in the free association task, this delivers faster production of an associate compared to traits not primed (i.e., not inferred from the trait-implying sentence; Orghian et al., 2017).

In experiment 1, after reading trait-implying and rearranged sentences (i.e., sentences that contain roughly the same words as the trait-implying ones but do not imply the trait), participants were presented with targets and instructed to say the first word that came to their mind. The reaction times (RTs) to generate a word upon the sight of a target trait were shorter when the target followed a sentence that implied the trait than when the target followed a sentence that did not imply that trait. We interpreted this difference in RTs as being due to the fact that the trait is spontaneously inferred during the reading of the implying sentences and, as a consequence, the lexical network of that trait becomes activated, facilitating the generation of associates in the free association task.[2] In a second experiment, we extended these results to delayed tests (Orghian et al., 2017, Experiment 2). In a third experiment, we used pairs of photos and trait implicative sentences at the learning phase and pairs of the same photos and traits, implied and not implied by the sentences, as free association prompts at the test. In this experiment, we were able to show that the results of previous studies were replicated only for STIs (i.e., when the person depicted in a photo that accompanies the trait implicative sentence is said to be the actor in the sentence) and not for STTs (i.e., the person depicted in a photo that accompanies the trait implicative sentence is said to be different from the actor in the sentence). We interpreted this result as a demonstration of the sensitiveness of the paradigm to actor-trait associations that are supposed to develop when trait inferences are made.

We believe that this new paradigm offers a number of promising features: i) free association is such an easy task that strategic considerations are irrelevant and of little help for performance; ii) it is very flexible, in the sense that it can be used immediately after the encoding trait-implicative sentence stimuli or in a delayed fashion and combining a target photo with the free association trait prompt or not; iii) it is able to discriminate relevant and irrelevant targets and thus replicate the differences found in spontaneous trait inference and transference.

However, we also believe that further tests of this new paradigm will bring to life new challenges to STI research and to its own validity and/or usefulness. As Uleman et al. (1996) pointed out, no experimental paradigms applied to study STI are able to discard all alternative explanations of their results. That, however, has not prevented research in STI to develop and flourish. In this spirit, we will next provide promising new paths for STI research. As it happened with several research contributions of James Uleman and collaborators, the new proposed extensions in STI research were greatly inspired by parallel work developed in linguistics and text comprehension. We think that continuing to explore the relevant connections between STI research and linguistics and text comprehension is important because it can help us to both consider the features that STIs share with other inferences and the aspects that, by contrast, make them unique.

New Extensions in STI Research: Text Comprehension, Semantics, and beyond Semantics

In most STI paradigms, researchers use language as a means of presenting the trait-implying information. Language is not the sole way to convey a message and information, and actually some studies have used, for instance, silhouettes to effectively find occurrences of STI (Fiedler & Schenck, 2001; Fiedler et al., 2005; see Uleman et al., 2008). However, since language is the most used one, it has recently become a more consistent focus of attention in STI research.

The perspective that language analysis is important for controlling and studying STI, though recent, is not completely new or surprising. Some authors have hinted that linguistic-related factors were of interest and begged for further research to clarify them. Uleman was no exception. Throughout some of his works, we find several comments regarding linguistic elements (such as verbs and adjectives) and their encoding and comprehension that try to make sense of language as a system in STI research (Uleman et al., 1996; Uleman, 1999, 2015; Uleman et al., 2008).

Uleman (2015) mentions the concept of causal schemata being part of the verb and its informational structure. For example, some verbs like "to bore" or "to admire" have informational schemata with different focus points. Lee et al. (2017) also discuss linguistic features regarding language abstractness. The authors refer that adjectives are more abstract than verbs and that verbs serve to facilitate concrete thinking (see also, Semin & Fiedler, 1988). Another fact that many researchers agree on is that STIs are attributional and that they are attributed to the actor of the behavior (Todorov & Uleman, 2002, 2004). However, the notion of "actor," when using language, is ambiguous, since it could be used to refer to a logical position in the sentence (i.e., subject) or a semantic role played by the entity (i.e., agent). Finally, Orghian et al. (2019) discuss two ways of reaching an inference: through word activation or through text comprehension. However, the authors also acknowledge that "very little is known about the text-processing mechanisms by which the trait is activated" (p. 560).

All of this, and the fact that linguistic inferences are studied by social psychologists and linguists alike (McKoon & Ratcliff, 1992; Graesser et al., 1994), shows that language factors are relevant for STI research. The biggest challenge to using language is a difficulty in finding tools and empirical parameters that can be controlled and tested. This branch of research has developed a core idea that might bring STI research to a multidisciplinary level with linguistics. We provide two illustrations of our own research aiming at a greater integration between the two fields.

The Minimalist Hypothesis: Implications for STI Research

How stored knowledge interacts with new information to generate new inferences is perhaps the greatest challenge for researchers in the text

comprehension domain. Among the different theoretical approaches (e.g., Schank & Abelson, 1977 scripts framework; Graesser et al., 1994 constructionist perspective), the minimalist hypothesis (McKoon & Ratcliff, 1992) strikes us as particularly relevant for three main reasons. First, the minimalist hypothesis was developed to account for the type of inferences that are automatically generated during text comprehension and has thus direct implications for STI research. Second, the minimalist hypothesis configures a processing model proposal that defines the conditions of automatic inference encoding, allowing for specific predictions that can be empirically tested. Third, the minimalist hypothesis has received considerable empirical support (McKoon & Ratcliff, 1981, 1986; for a review see McKoon & Ratcliff, 1992).

The key assumption of the hypothesis is that readers are minimal inference encoders that automatically generate only two types of inferences in the absence of specific goals: a) inferences that are easily available; and b) inferences that are necessary to establish the local coherence of the text. Easily available information is defined as text information that is still in working memory or that is retrieved from long-term memory via passive activation mechanisms (McKoon et al., 1996). The need for local coherence concerns making sense of information mentioned in the text that is simultaneously held in working memory.

Another defining feature of the minimalist approach is the rejection of an all-or-none view of inference generation. Instead, inferences vary in a continuum of strength (Cook et al., 2001; Keefe & McDaniel, 1993; McKoon & Ratcliff, 1986, 1989). Thus, rather than asking whether an inference had occurred or not, it is important to explore the degree to which an inference is encoded. According to this view, some inferences may end up fully encoded, while others may be only weakly or partially encoded into memory.

It follows from the main assumptions of the minimalist approach that the automatic inference process is highly dynamic and context dependent. This means that it may often be impossible to distinguish, solely from its cognitive nature, between a self-generated thought content that will later become a full-fledged inference from self-generated thought content that will merely fade away without changing the representation of the target (with which it was briefly associated). The difference between automatic inferences and mere activation crucially depends on the context and the requirements of the cognitive tasks.

Based on the McKoon and Ratcliff minimalist framework (1992, 1995; McKoon et al., 1996; Ratcliff & McKoon, 2008), Ramos et al. (2012; Ramos, 2009) attempted to position STI research in a broader framework. More specifically, Ramos et al. (2012; Ramos, 2009) have proposed a gradual view of STI generation. According to the minimalist view, STI are considered to vary in a continuum of encoding strength making the difference between concept activation and inference subtler than usually assumed. At the lowest level of strength, STI are concept transient activations not linked in a long-lasting manner to the representation of the actor (in this sense, hardly

indistinguishable from STT or other transient goal or state activation). At the highest strength level, however, STIs are encoded in long-term memory as part of the representation of the actor. Furthermore, it follows from the the principle of local coherence that such full-fledged spontaneous inferences will be more or less likely to occur depending on their contribution to a coherent integration of the social information presented. This view agrees with the notion that STI (and other spontaneous inferences) regularly occur as part of our habitual quest to impose meaning to the social events that surround us (Uleman, Newman & Moskowitz, 1996; Uleman et al., 2008), but it also asserts that STI (and other spontaneous inferences) may be constrained when they fail to contribute to form a coherent view of others and of the local context where the social narrative unfolds.

Research in this minimalist view of STIs is still scarce. However, Ramos et al. (2012) nicely illustrated the latter point (local coherence) by showing that trait implying-behaviors that were inconsistent with a stereotype associated with the actor (e.g., the garbage man wins the science quiz) inhibited the STI (intelligent) while prompting spontaneously situational inferences (SSI) that facilitate making sense of the behavior. In essence, the stereotype information constrained the occurrence of stereotype-inconsistent STI and promoted SSI in order to develop a coherent view of the social environment (see also, Wigboldus et al., 2003).

In sum, we propose that STI research may be informed by a minimalist view of text comprehension and put into a more general perspective since the same fundamental problem underlies both domains: How do people build up and update meaningful knowledge representations from an ever-changing environment (being it the reading of a text or the real-life unfolding of human behavior)?

Language as a "Game Changer"

The second illustration involves more recent work by Marcelo et al. (2019). Marcelo et al. (2019) studied STI and how the traits might be activated in reading simple sentences. The introduction of trait-rich words, such as adverbs of manner, was sufficient to change the traits inferred from a sentence. At the same time, the adverb had to make some sense in the sentence. For instance, completely opposite adverbs of manner regarding the trait implied by the verb were not taken into account for inference, and their traits were therefore not as activated for inference as the trait implied by the verb. One clear example of that is the combination of "tripping cautiously," where the inferred trait is "clumsy" (the verb trait) and not "careful" (the adverb trait). This suggests that STI are text-comprehension effects that can be influenced by word activation, but not overpowered by them.

These results and questions open new perspectives regarding further research looking at linguistic materials as the conveyors of the information for STI. If they are comprehension-based effects, the principles of reading,

language comprehension, and language structure are of the utmost importance for this branch of investigation.

Recently, one critical question being asked in STI research, in order to understand how inferences work in text comprehension processes, is the linguistic role that syntactic and semantic components play in the process of trait inference and how this can be replicated in non-linguistic methodologies in the future. One way to start answering this question is to test inferences for the *agent who performs a trait-implicative action (i.e., the subject of the sentence describing the action)* and the *target person who receives* that action (i.e., the object of the sentence) in active and passive voices—two syntactic structures studied by linguistics as mirror versions of one other. For example, "John called Carl" and "Carl was called by John." Changes in the propensity to make STI might indicate the extent to which these inferences are sensitive to meaning and language structure. Active and passive voices have also been studied from a comprehension side in social cognition by looking at how participants perceive rapists and rape victims when being described by one of these sentence voices (Henley et al., 1995). For instance, in one of their studies, Henley et al. (1995) had participants read mock news reports on rape and other crimes (e.g., battery, robbery, and murder). Participants were then asked to rate the degree of harm caused to victims and the perpetrator responsibility, after each crime. With passive voice, males (but not females) attributed less victim harm and perpetrator responsibility for violence against women than with active voice.

To explore this question, we recently ran a study with a forced choice task similar to the paradigm used by Ferreira et al. (2012). In this task, participants saw a pair of pictures (one for the agent, another for the target) in the order they appear in the sentence, with their respective names, and a sentence with low linguistic complexity, with the same number of elements—a subject, a verb, and an object. There were three types of trait-implicative sentences: sentences that implied traits for both entities, sentences that only implied traits for agents, or that only implied traits for targets. The full sentence set also included trait-neutral fillers, and filler sentences without human agency.

The pattern of results matched linguistic principles for trait activation and attribution for both agent and target. After encoding all the stimuli (trait-implying sentences, trait-neutral fillers, and non-human fillers) and going through a distractor task for mental calculus, participants were given one of the pictures and had to choose which form of the sentence they saw in the first stage—the sentence with or without the trait. The pattern of results matched linguistic principles for trait activation and attribution for both agent and target, following syntactic prominence. In the active voice, agents are more prominent than the targets of the action, because agents serve the syntax function of subject. Conversely, in the passive voice, agents serve as verb objects and have thus a less privileged position than the targets of the action. As expected from this linguistic approach, in trait-implying

sentences, participants produced significantly more STIs about agents in the active voice sentences than in the passive voice sentences. The reverse was true for STIs about the targets of the action. Moreover, in the active voice, participants drew more STIs about agents than about targets of the action, whereas in the passive voice they drew more STIs about the targets of the action than about agents.

Inside Out and Outside In

Kahneman and Lovallo (1993) distinguished between two modes of forecasting the evolution of a given scenario: the inside and outside views. The inside view is generated by focusing on the specific scenario and by considering the obstacles to the desired end state, the prospects of future progress, and by extrapolating current trends. The outside view, quite on the contrary, ignores specific elements and involves no attempt at detailed forecasting of the future for the scenario at hand. Instead, it focuses on the general characteristics of a class of scenarios chosen to be similar in relevant respects to a target scenario. We believe that this distinction is useful not only for forecasting but for problem-solving and for development and progress of a field of inquiry. Trying to advance a given field of inquiry by concentrating on what makes that field unique or by concentrating on the general features that the field shares with other fields highlights two vastly different progress strategies.

In that sense, we argue that the progress and advancement of research in STI is characterized by the outside view strategy followed by its main mentor, James S. Uleman. We wonder whether the STI secret ingredient, which we previously argued to be missing from many well-intended recipes aimed at better grasp of several person perception phenomena, would ever be found without this outside view. Not only impression formation and attribution, but also text comprehension and implicit memory have been, of course, the sister disciplines in our endeavor. And we contend that they should remain so, together with linguistics and other dependable road companions. And of course, to every fruitful outside view moment, a long soul-searching inside view follows.

But we have built this outside view almost exclusively from the methodological bricks, importing smart techniques and procedures, one after the other, more often than not with good effect. However, we urge researchers interested in STIs to also consider other valuable resources generated outside our discipline: theories and models. Problems like the nature of the person-trait link (associative versus attributional), the differences between inference trait and transient trait word activation and the distinction between STIs and situational, verb or state inferences, are complex. These problems are not likely to be answered in a meaningful way without being adequately framed in terms of specific theoretical models with the provision of unambiguous

assumptions and hypotheses. STI has been one of the areas benefiting from an outside view in psychological research. Let us ensure that it remains so, with even better insights to enrich this fertile soil that Jim has been working on, so passionately, for all these years.

Conclusion

The research on spontaneous traits inferences has been characterized by a staggering breadth of scope, importing and adapting new methods and techniques, but also being able to provide safe haven for vivid and diverse polemics while remaining a truly integrative and cumulative scientific endeavor. As it often happens, the main qualities of a research field reflect the nature of the key contribution of some of its pioneers and continuing mentors. That is certainly the case for STIs research and James S. Uleman. We would like to take the chapter as a means to acknowledge our intellectual indebtedness and thank Jim's partnership over the years.

Notes

1 These topics are delved into more deeply in other chapters of this volume.
2 In a similar vein, Moskowitz and Roman (1992) used a judgment task to tap into the downstream consequences of STI. The rationale was that STI could function as self-generated primes that would facilitate access to related semantic content and promoting assimilation effects on judgments based on that information.

References

Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology*, *41*(3), 258–290. 10.1037/h0055756

Anderson, J. R. (1978). Arguments concerning representations for mental imagery. *Psychological Review*, *85*(4), 249–277. 10.1037/0033-295X.85.4.249

Bassili, J. N. (1989). Trait encoding in behavior identification and dispositional inference. *Personality and Social Psychology Bulletin*, *15*(3), 285–296. 10.1177/01461 67289153001

Brown, R. D., & Bassili, J. N. (2002). Spontaneous trait associations and the case of the superstitious banana. *Journal of Experimental Social Psychology*, *38*(1), 87–92. 10.1006/jesp.2001.1486

Boyle, R. (1666). The Origin of forms and Qualities (according to the corpuscular philosophy). From J. Bennett (text preparer) (2017). Some Texts From Early Modern Philosophy. https://www.earlymoderntexts.com/authors/boyle

Bruner, J. S., & Tagiuri, R. (1954). The perception of people. In G. Lindzey (Ed.), *Handbook of social psychology* (Vol. 2, pp. 634–654). Addison-Wesley, Reading.

Carlston, D. E., & Skowronski, J. J. (1994). Savings in the relearning of trait information as evidence for spontaneous inference generation. *Journal of Personality and Social Psychology*, *66*(5), 840–856. 10.1037/0022-3514.66.5.840

Carlston, D. E., & Skowronski, J. (2005). Linking versus thinking: Evidence for the different associative and attributional bases of spontaneous trait transference and

spontaneous trait inference. *Journal of Personality and Social Psychology*, *89*(6), 884–898. 10.1037/0022-3514.89.6.884

Carlston, D. E., Skowronski, J. J., & Sparks, C. (1995). Savings in relearning: II. On the formation of behavior-based trait associations and inferences. *Journal of Personality and Social Psychology*, *69*(3), 420–436. 10.1037/0022-3514.69.3.429

Crawford, M. T., Skowronski, J. J., & Stiff, C. (2007). Limiting the spread of spontaneous trait transference. *Journal of Experimental Social Psychology*, *43*, 466–472. 10.1016/j.jesp.2006.04.003

Crawford, M. T., Skowronski, J. J., Stiff, C., & Leonards, U. (2008). Seeing, but not thinking: Limiting the spread of spontaneous trait transference II. *Journal of Experimental Social Psychology*, *44*(3), 840–847. 10.1016/j.jesp.2007.08.001

Crawford, M. T., Skowronski, J. J., Stiff, C., & Scherer, C. R. (2007). Interfering with inferential, but not associative processes underlying spontaneous trait inference. *Personality and Social Psychology Bulletin*, *33*, 677–690. 10.1177/0146167206298567

Cook, A. E., Limber, J. E., & O'Brien, E. J. (2001). Situation-based context and the availability of predictive inferences. *Journal of Memory and Language*, *44*(2), 220–234. 10.1006/jmla.2000.2744

D'Agostino, P. R., & Beegle, W. (1996). A reevaluation of the evidence for spontaneous trait inferences. *Journal of Experimental Social Psychology*, *32*(2), 153–164. 10.1006/jesp.1996.0007

Enzle, M. E., & Schopflosher, D. (1978). Instigation of attribution processes by attributional questions. *Personality and Social Psychology Bulletin*, *4*(4), 595–599. 10.1177/014616727800400420

Ferreira, M. B., Garcia-Marques, L., Hamilton, D., Ramos, T., Uleman, J. S., & Jerónimo, R. (2012). On the relation between spontaneous trait inferences and intentional inferences: An inference monitoring hypothesis. *Journal of Experimental Social Psychology*, *48*, 1–12. 10.1016/j.jesp.2011.06.013

Fiedler, K., & Schenck, W. (2001). Spontaneous inferences from pictorially presented behaviors. *Personality and Social Psychology Bulletin*, *27*(11), 1533–1546. 10.1177/01461672012711013

Fiedler, K., Schenck, W., Watling, M., & Menges, J. I. (2005). Priming trait inferences through pictures and moving pictures: the impact of open and closed mindsets. *Journal of Personality and Social Psychology*, *88*(2), 229–244. 10.1037/0022-3514.88.2.229

Garcia-Marques, L., & Ferreira, M. B. (2011). Friends and foes of theory construction in psychological science: vague dichotomies, unified theories of cognition, and the new experimentalism. *Perspectives on Psychological Science*, *6*(2), 192–201. 10.1177/1745691611400239

Gilbert, D. T. (1998). Speeding with ned: A personal view of the correspondence bias. In J. M. Darley & J. Cooper (Eds.), *Attribution and social interaction: The legacy of Edward E. Jones* (pp. 5–66). American Psychological Association. 10.1037/10286-001

Gilbert, D. T., & Jones, E. E. (1986). Perceiver-induced constraint: Interpretations of self-generated reality. *Journal of Personality and Social Psychology*, *50*(2), 269–280. 10.1037/0022-3514.50.2.269

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, *117*(1), 21–38. 10.1037/0033-2909.117.1.21

Gilbert, D. T., Pelham, B. W., & Krull, D. S. (1988). On cognitive busyness: When person perceivers meet persons perceived. *Journal of Personality and Social Psychology*, *54*(5), 733–740. 10.1037/0022-3514.54.5.733

Goren, A., & Todorov, A. (2009). Two faces are better than one: Eliminating false trait associations with faces. *Social Cognition*, *27*(2), 222–248. 10.1521/soco.2009.27.2.222

Graesser, A., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*(3), 371–395. 10.1037/0033-295X.101.3.371

Ham, J., & Vonk, R. (2003). Smart and easy: Co-occurring activation of spontaneous trait inferences and spontaneous situational inferences. *Journal of Experimental Social Psychology*, *39*(5), 434–447. 10.1016/S0022-1031(03)00033-7

Hamilton, D. L., Katz, L. B., & Leirer, V. O. (1980). Cognitive representation of personality impressions: Organizational processes in first impression formation. *Journal of Personality and Social Psychology*, *39*(6), 1050–1063. 10.1037/h0077711

Hastie, R. (1984). Causes and effects of causal attribution. *Journal of Personality and Social Psychology*, *46*(1), 44–56. 10.1037/0022-3514.46.1.44

Heider, F. (1958). The naive analysis of action. In F. Heider, *The psychology of interpersonal relations* (pp. 79–124). John Wiley & Sons. 10.1037/10628-004

Henley, N., Miller, M., & Beazley, J. (1995). Syntax, semantics, and sexual violence: Agency and the passive voice. *Journal of Language and Social Psychology*, *14*(1–2), 60–84. 10.1177/0261927X95141004

Higgins, E. T., & Bargh, J. A. (1987). Social cognition and social perception. *Annual Review of Psychology*, *38*, 369–425. 10.1146/annurev.ps.38.020187.002101

Hourihan, K. L., & MacLeod, C. M. (2007). Capturing conceptual implicit memory: Produce an association. *Memory & Cognition*, *35*, 1187–1196. 10.3758/BF03193592

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*(5), 513–541. 10.1016/0749-596X(91)90025-F

Jones, E. E., & Davis, K. E. (1965). From acts to dispositions the attribution process in person perception. In *Advances in experimental social psychology* (Vol. 2, pp. 219–266). Academic Press.

Kahneman, D., & Lovallo, D. (1993). Timid choices and bold forecasts: A cognitive perspective on risk taking. *Management Science*, *39*(1), 17–31. 10.1287/mnsc.39.1.17

Keefe, D. E., & McDaniel, M. A. (1993). The time course and durability of predictive inferences. *Journal of Memory and Language*, *32*(4), 446–463. 10.1006/jmla.1993.1024

Keenan, J. M., Potts, G. R., Golding, J. M., & Jennings, T. M. (1990). Which elaborative inferences are drawn during reading? A question of methodologies. In D. A. Balota, G. B. Flores d'Arcais & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 377–402). Lawrence Erlbaum Associates.

Kelley, H. H. (1967). Attribution theory in social psychology. *Nebraska Symposium on Motivation*, *15*, 192–238.

Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist*, *28*(2), 107–128. 10.1037/h0034225

Kintsch, W. (1974). *The representation of meaning in memory*. Lawrence Erlbaum.

Kintsch, W. (1975). Memory representations of text. In R. L. Solso (Ed.), *Information processing and cognition*. Erlbaum.

Klein, S. B., & Loftus, J. (1990). Rethinking the role of organization in person memory: An independent trace storage model. *Journal of Personality and Social Psychology*, *59*(3), 400–410. 10.1037/0022-3514.59.3.400

Lau, R. R., & Russell, D. (1980). Attributions in the sports pages. *Journal of Personality and Social Psychology*, *39*(1), 29–38. 10.1037/0022-3514.39.1.29

Lee, H., Shimizu, Y., Masuda, T., & Uleman, J. (2017). Cultural differences in spontaneous trait and situation inferences. *Journal of Cross-Cultural Psychology*, *48*(5), 627–643. 10.1177/0022022117699279

Marcelo, D., Garcia-Marques, L., & Duarte, I. (2019). Language as a 'game changer' for spontaneous trait inference. *Cognitive Linguistic Studies*, 6(1), 185–209. 10.1075/cogls.00035.mar

McKoon, G., Gerrig, R. J., & Greene, S. B. (1996). Pronoun resolution without pronouns: Some consequences of memory-based text processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(4), 919–932. 10.1037/0278-7393.22.4.919

McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(1), 82–91. 10.1037/0278-7393.12.1.82

McKoon, G., & Ratcliff, R. (1989). Semantic associations and elaborative inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*,*15*(2), 326–338.

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, *99*(3), 440–466. 10.1037/0033-295X.99.3.440

McKoon, G., & Ratcliff, R. (1995). The minimalist hypothesis: Directions for research. In C. A. Weaver III, S. Mannes & C. R. Fletcher (Eds.), *Discourse comprehension: Essays in honor of Walter Kintsch* (pp. 97–116). Lawrence Erlbaum Associates.

Moskowitz, G. B., & Olcaysoy Okten, I. (2016). Spontaneous goal inference (SGI). *Social and Personality Psychology Compass*, *10*, 64–80. 10.1111/spc3.12232

Moskowitz, G., & Roman, R. (1992). Spontaneous trait inferences as self-generated primes: Implications for conscious social judgment. *Journal of Personality and Social Psychology*, *62*(5), 728–738. 10.1037/0278-7393.12.1.82

Nelson, D. I., & Goodmon, L. B. (2002). Experiencing a word can prime its accessibility and its associative connections to related words. *Memory and Cognition*, *30*(3), 380–398. 10.3758/BF03194939

Norman, D. A., Rumelhart, D. E., & the LNR Research Group. (1975). *Explorations in cognition*. Freeman.

Orghian, D., Garcia-Marques, L., Uleman, J. S., & Heinke, D. (2015). A connectionist model of spontaneous trait inferences and spontaneous trait transference: Do they have the same underlying processes?. *Social Cognition*, *33*(1), 20–66. 10.1521/soco.2015.33.1.20

Orghian, D., Ramos, T., Garcia-Marques, L., & Uleman, J. (2019). Activation is not always inference: Word-based priming in spontaneous trait inference. *Social Cognition*, *37*(2), 557–585. 10.1521/soco.2019.37.2.145

Orghian, D., Smith, A., Garcia-Marques, L., & Heinke, D. (2017). Capturing spontaneous trait inference with the modified free association paradigm. *Journal of Experimental Social Psychology*, *73*, 243–258. 10.1016/j.jesp.2017.07.004

Orvis, B. R., Cunningham, J. D., & Kelley, H. H. (1975). A closer examination of causal inference: The roles of consensus, distinctiveness, and consistency information. *Journal of Personality and Social Psychology*, *32*(4), 605–616. 10.1037/0022-3514.32.4.605

Pyszczynski, T. A., & Greenberg, J. (1981). Role of disconfirmed expectancies in the instigation of attributional processing. *Journal of Personality and Social Psychology*, *40*(1), 31–38. 10.1037/0022-3514.40.1.31

Quattrone, G. A. (1982). Overattribution and unit formation: When behavior engulfs the person. *Journal of Personality and Social Psychology*, *42*(4), 593–607. 10.1037/0022-3514.42.4.593

Ramos, T. (2009). *A flexible view of spontaneous trait inferences* [Doctoral Dissertation, ISCTE-IUL]. ISCTE-IUL Repository.

Ramos, T., Garcia-Marques, L., Hamilton, D. L., Ferreira, M. B., & Van Acker, K. (2012). What I infer depends on who you are: The influence of stereotypes on trait and situational spontaneous inferences. *Journal of Experimental Social Psychology*, *48*, 1247–1256. 10.1016/j.jesp.2012.05.009

Ratcliff, R., & McKoon, G. (1981). Does activation really spread? *Psychological Review*, *88*(5), 454–462. 10.1037/0033-295X.88.5.454

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*(4), 873–922. 10.1162/neco.2008.12-06-420

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, *107*(2), 358–367. 10.1037/0033-295X.107.2.358

Rosenberg, S., Nelson, C., & Vivekananthan, P. S. (1968). A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology*, *9*(4), 283–294. 10.1037/h0026086

Ross, L., & Nisbett, R. (1991). *The person and the situation: Perspectives of social psychology*. McGraw-Hill.

Schank, R. C. (1975). The structure of episodes in memory. In D. G. Bobrow & A. Collins (Eds.), *Representation and understanding: Studies in cognitive science* (pp. 237–272). Elsevier. 10.1016/B978-0-12-108550-6.50014-8

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals and understanding: An inquiry into human knowledge structures*. Lawrence Erlbaum Associates.

Semin, G., & Fiedler, K. (1988). The cognitive functions of linguistic categories in describing persons: Social cognition and language. *Journal of Personality and Social Psychology*, *54*(4), 558–568. 10.1037/0022-3514.54.4.558

Skowronski, J. J., Carlston, D. E., Mae, L., & Crawford, M. T. (1998). Spontaneous trait transference: Communicators take on the qualities they describe in others. *Journal of Personality and Social Psychology*, *74*(4), 837–848. 10.1037/0022-3514.74.4.837

Smith, E. R., Miller, F. D. (1979a). Attributional information processing: A response time model of causal subtraction. *Journal of Personality and Social Psychology*, *37*(10), 1723–1731. 10.1037/0022-3514.37.10.1723

Smith, E. R., & Miller, F. D. (1979b). Salience and the cognitive mediation of attribution. *Journal of Personality and Social Psychology*, *37*(12), 2240–2252. 10.1037/0022-3514.37.12.2240

Smith, E. R., & Miller, F. D. (1983). Mediation among attributional inferences and comprehension processes: Initial findings and a general method. *Journal of Personality and Social Psychology*, *44*(3), 492–505. 10.1037/0022-3514.44.3.492

Todorov, A., & Uleman, J. S. (2002). Spontaneous trait inferences are bound to actors' faces: Evidence from a false recognition paradigm. *Journal of Personality and Social Psychology*, *83*(5), 1051–1065. 10.1037/0022-3514.83.5.1051

Todorov, A., & Uleman, J. S. (2004). The person reference process in spontaneous trait inferences. *Journal of Personality and Social Psychology*, *87*(4), 482–493. 10.1037/0022-3514.87.4.482

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*(5), 352–373. 10.1037/h0020071

Uleman, J. S. (1999). Spontaneous versus Intentional Inferences in Impression Formation. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 141–160). Guilford.

Uleman, J. (2015). Causes and causal attributions: Questions raised by Dave Hamilton and spontaneous trait inferences. In S. J. Stroessner & J. W. Sherman (Eds.), *Social perception from individuals to groups* (pp. 52–70). Psychology Press.

Uleman, J. S., Hon, A., Roman, R. J., & Moskowitz, G. B. (1996). On-line evidence for spontaneous trait inferences at encoding. *Personality and Social Psychology Bulletin*, *22*(4), 377–394. 10.1177/0146167296224005

Uleman, J., Saribay, S., & Gonzalez, C. (2008). Spontaneous inferences, implicit impression, and implicit theories. *Annual Review of Psychology*, *59*, 329–360. 10.1146/annurev.psych.59.103006.093707

Uleman, J. S., Newman, L. S., & Moskowitz, G. B. (1996). People as flexible interpreters: Evidence and issues from spontaneous trait inference. In M. P. Zanna (Ed.), *Advances in experimental social psychology*, (Vol. 28, pp. 211–279). Academic Press.

Van Overwalle, F., Drenth, T., & Marsman, G. (1999). Spontaneous trait inferences: Are they linked to the actor or to the action?. *Personality and Social Psychology Bulletin*, *25*(4), 450–462. 10.1177/0146167299025004005

Wigboldus, D. H. J., Dijksterhuis, A., & Van Knippenberg, A. (2003). When stereotypes get in the way: Stereotypes obstruct stereotype-inconsistent trait inferences. *Journal of Personality & Social Psychology*, *84*(3), 470–484.

Winter, L., & Uleman, J. S. (1984). When are social judgments made? Evidence for the spontaneousness of trait inferences. *Journal of Personality and Social Psychology*, *47*(2), 237–252. 10.1037/0022-3514.47.2.237

Winter, L., Uleman, J. S., & Cunniff, C. (1985). How automatic are social judgments? *Journal of Personality and Social Psychology*, *49*(4), 904–917. 10.1037/0022-3514.49.4.904

Wong, P. T., & Weiner, B. (1981). When people ask "why" questions, and the heuristics of attributional search. *Journal of Personality and Social Psychology*, *40*(4), 650–663. 10.1037/0022-3514.40.4.650

Wyer, R. S., & Srull, T. K. (1986). Human cognition in its social context. *Psychological Review*, *93*(3), 322–359. 10.1037/0033-295X.93.3.322

8 Predictively Coding Objects and Persons

Ethan Ludwin-Peery and Yaacov Trope
New York University

When we meet someone for the first time, we quickly make inferences about them. If they flash a Duchenne smile, we infer that they are warm and friendly. If they give a firm handshake, we infer that they are steadfast and reliable, that we can count on them in the future if we need their help. From these cues, we go right to personality; not only how they are acting and feeling right now, but who they are, what they are like underneath, what we can expect of them. All this despite the fact that a warm smile or a firm handshake, in reality, provide only minimal information.

As a result, psychologists have long asked, what does it mean that we draw such broad inferences from these minor cues? While early person perception research has focused on the organization of personality impressions (Asch, 1946) and their accuracy (Cronbach, 1955), Heider's (1958) *The Psychology of Interpersonal Relations* and Jones and Davis' (1965) *From Acts to Dispositions: The Attribution Process in Person Perception* put forth personality trait inferences as the core issue for person perception research. Building on Heider and Jones' contributions, subsequent work further explored the perceptual identifications and inferential calculus underpinning trait attributions (Trope, 1974, 1986) and their consequences for the over-attribution of behavior to personality traits (Gilbert & Malone, 1995). But it was Jim Uleman's seminal work on spontaneous trait inference that spurred a paradigmatic shift highlighting trait inference as what Tversky and Kahneman (1974) called a "natural assessment," one that is ubiquitous and unintentional.

These approaches are well in line with a new paradigm which takes the perspective that all cognition is based on the use of abstraction in the service of prediction. To begin with, we review this paradigm and discuss its explanatory power in a domain in which it has so far been most successful—namely, perception. Following this, we present some new data from our lab on the role of abstract construals in predicting events across psychological distance, and link the entire perspective back to social cognition, in the discussion of face perception and trait inference.

DOI: 10.4324/9781003045687-9

Predictive Coding

Being a passive consumer of information is rarely a good strategy, for a couple of reasons. For one, going out and collecting information for yourself is usually a better way to stay informed than just sitting around and waiting for information to come to you. But second, the passive approach can be extremely costly, because many parts of the world are very predictable.

It is true that the world is very complicated. At any moment, many things are happening all around us, and most of them are happening quite quickly. As a result, perceiving the world directly is computationally intensive. But, fortunately for us, and for intelligent life in general, the world is not a total random mess. Many parts of this turmoil can be predicted, often with a high degree of accuracy. Instead of trying to drink from the firehose of raw sensory information, the perceptual system instead makes educated guesses about what is happening in the world and uses perceptual input to check and adjust its guesses (Bar et al., 2006; Clark, 2015; Friston, 2005; Helmholtz, 1860). This predictive system ends up being a very efficient and largely accurate approach, but it does have certain side effects and limitations, which we will discuss.

Predictive theories of perception go by many names. One common term is *predictive coding*, which is the term we will use in this chapter. According to this perspective, prediction is a fundamental process of cognition, involved not only in perception but also attention, action, but possibly even higher-level aspects of cognition, like decision making (Clark, 2015). This task would be hard enough on its own, but the raw visual field is a complete mess. The eyes capture a pair of 2-D images, horribly distorted towards the edges, with a massive blindspot in their center. Color vision quickly fades out as we leave the center of the visual field; red and green cones are almost entirely absent by 10° out from the fovea (Johnson, 1986). And this is only for vision—the other senses are just as bad, if not worse. Something is needed to coerce this assault into meaningful, manageable representations.

Raw sense data is noisy, and the key insight of predictive coding is that one of the ways to make sense of the cacophony of sensation is to have a good idea of what is coming next, by building abstractions that allow you to make predictions (Clark, 2013). Successful prediction is fast and efficient. If you have a good guess about what you're about to see, then the only things you need to register are things which were unexpected—the surprises. Everything else can be understood to be what was predicted, merely business as usual, and consequently ignored.

Since Bartlett (1932), psychologists have posited in their theories the important role of predictions for navigating everyday, but complex, environments through the use of schemas (Markus, 1977; Schank & Abelson, 1977), stereotypes (Fiske & Trope, 1999), heuristics (Tversky & Kahneman, 1973), and mindsets (Heckhausen, 1986). When your prediction is correct, you can continue as planned, saving a huge amount of time. Consider the experience of

entering your home after a long day at work. You already have a very high-quality model of what sorts of furniture and other objects are likely to be there when you open the door, and you are very confident about their locations. You don't need to stand there wide-eyed and take in every part of your living room every single evening; it would be a huge waste of time and processing resources. Instead, you can assume that things remain largely as they always are, unless you happen to observe evidence to the contrary, perhaps by stubbing your toe on a misplaced piece of furniture. These are the benefits of being able to predict the world around you, and filter out what you already know to be there.

Systems that model and try to anticipate their inputs can quickly become much more efficient than systems which are simply passive or reactive. When the mind notices a consistent pattern in the world, it develops abstractions to predict and account for that consistency. In the most general sense, an abstraction is the determination that two things that are subjectively distinguishable are interchangeable (Gilead et al., 2019). The percepts of the same coffee mug seen from two different angles are subjectively distinguishable; despite this, I abstract them as both being percepts of the same object. The winter of 2016 and the winter of 2017 were subjectively distinguishable; despite this, I abstract them as both being examples of "winter." Since the world is at least somewhat predictable, over time the use of these abstractions is very efficient, even when it's not perfectly accurate. Social schemas (Markus 1977; Schank & Abelson, 1977), stereotypes (Fiske & Trope, 1999), and even traits themselves (Trope & Higgins, 1993a) are all examples of abstractions that social psychologists posit exist as mental representations for the purpose of organizing and grouping the world into meaningful sets of things (see review by Moskowitz, 2005).

These abstractions do not just make predictions. They also adjust information coming from the senses, usually transforming that information so that it appears more in line with prior information than it really is. We don't experience the world directly; instead, our perception is mediated. What we actually experience is a combination of our expectations, observations that have been corrected to be more in line with our models, and the few surprises that get through (Bartlett, 1932; Bruner, 1957; Cantor & Mischel, 1979; Helmholtz, 1860; Markus, 1977; Meyer & Schvaneveldt, 1971).

The predictive coding system can be understood as an ongoing process of negotiation between bottom-up signals and top-down models or expectations. This negotiation involves evidence from the world and the internal understanding of what is likely to be out there, with these two systems constantly attempting to come to an agreement. To make this work, the mind is organized hierarchically, into what can be understood as layers. At each layer, a top-down set of expectations from the model above is compared to a new set of bottom-up evidence from the layer below, a stream of information originating in the senses. In most cases, this is very effective and helps to filter out a lot of the boring consistencies that are natural to

perception. Close to the "bottom" of this structure, the mind is in direct contact with perceptual inputs, and mechanisms at this layer attempt to predict low-level visual information, such as brightness and orientation. Higher layers combine this information into gradually more and more abstract concepts. Each layer is in constant communication with the layers above and below it, so there is a constant stream of data moving through the system in both directions (Bar et al., 2006; Friston, 2005).

In reality, of course, predictions and sense data never match perfectly. Even when seeing a room we have viewed a thousand times, there will be some small change in the light or new scuff on the rug. There will always be some disagreement, and so for discrepancies of sufficiently small magnitude, the top-down model overrides the evidence, and the bottom-up signal proceeds no further. As previously discussed, abstraction can be considered the process of selectively deleting information, in order to gain efficiency. This is the cause of phenomena such as change blindness, where observers often fail to notice changes to scenes or objects (Simons & Levin, 1997). If a change is insignificant enough, it can be filtered out entirely.

Deleting information in this way may seem to be maladaptive, but it is actually essential. Your senses are registering thousands of events at any given time, but most of these are not important, and it would be distracting to have them constantly in your awareness. You are probably not aware of the feeling of your clothing on your skin, the hum of the air conditioning, the smell of the room you're sitting in, the pumping of blood through your veins, your blinking, the image of your nose in your peripheral vision, and so on, even though at some level all of these stimuli are being registered by your senses. Imagine trying to write an email or cook dinner with all these sense-data pushing through to your conscious awareness. At the very least, it would be very distracting (Lippmann, 1922).

When a disagreement between top-down and bottom-up processes is too large to "cook the books," too significant to filter out, then the model has failed to predict the world. This means two things must occur. First, the model must be updated, in hopes that it can make better predictions next time (see Moskowitz, Olcaysoy, Okten, & Schneid, this volume, for a discussion of when and if updating occurs). Second, the discrepancy must be dealt with (e.g., Chaiken et al., 1989; see Sherman, this volume, for a review of how people deal with such discrepancies). Both of these are accomplished by the system sending a prediction error signal to the levels above the point of disagreement. Notably, this is the only case where information is passed up to higher levels; each level of analysis takes as input only that information which is not explained by the mechanisms below it (Clark, 2013, 2015).

Predictive coding implies that perception is "controlled hallucination" (Clark, 2015). What we actually experience is a combination of our expectations, observations that have been corrected to be more in line with our models, and the few surprises that get through. At each layer, a top-down set of expectations from the model above it is compared to a new bottom-up set

of evidence. The two streams attempt to come to an agreement about the state of the world, in a process sometimes called the "perceptual handshake" (Kleinschmidt et al., 2012).

When sense data contradicts a highly confident model, these abstractions continue to blindly adjust perception. As a result, sometimes the predictive coding solution leads you to see things that aren't quite what's there. We know these edge cases of perception as perceptual illusions, which can be understood to be a side effect of these principles and this approach (Clark, 2013; King et al., 2017). Such illusions reflect the influence of top-down processes, and the strength of these illusions is a measure of the balance between top-down and bottom-up processes.

This is consistent with many perceptual illusions, and helps explain why some illusions are not experienced by individuals outside of "WEIRD" cultures. Illusions having to do with 90° angles (e.g., Müller-Lyer, 1889), for example, don't appear to affect people who were raised without exposure to heavily carpentered environments (Henrich et al., 2010). Visual illusions of this type occur when clear evidence is overruled by an extremely confident predictive model. One consequence of this is that they provide a convenient measure of just how confident a model is willing to be in its predictions when they disagree with evidence coming from the senses (see Trope & Thompson, 1997, Cameron & Trope, 2004).

Predictions Spanning across Psychological Distance

Conceiving of cognition as predictive coding raises the question of how far from me in the here-and-now the things it serves to predict are. Stimuli that are right in front of us are full of detail. They are close enough to observe, to interact with, they might move or be moved, and so on. If it is foggy, or raining, or even just dark out, our view of them changes. Because these percepts are so changeable, they can be harder to model reliably—in essence, they are harder to predict. But because we can observe nearby stimuli directly, we don't need to rely so heavily on our models in the first place. Predictions still come into play because they are efficient; but since the stimuli are less predictable, the models are more often overruled.

Stimuli that are further away in time or in space, however, necessitate a greater reliance on abstraction. Often this is because they cannot be observed directly; when they can be observed, it is at a much lower level of detail. If thinking about my plans for this afternoon, it would make sense to look out my window in the morning, as the weather now is at least somewhat indicative of the weather in a couple hours. Anticipating plans in my city a year from today, I would do better to think about the seasonal weather in general. The same is true if thinking of travelling to a far-off location. The weather here and now will be less informative. As a result, predictions about these stimuli are more useful.

This insight is central to construal level theory (CLT), which asserts that abstraction is the process that allows people to transcend their current circumstances and think about remote places, times, minds, and possibilities (Liberman & Trope, 2014; Trope & Liberman, 2010). CLT considers both spatial and temporal distance to be forms of psychological distance, along with other psychological dimensions associated with increased uncertainty, such as hypotheticality (Wakslak & Trope, 2009) and politeness (Stephan et al., 2010). In this model, psychological distance refers not only to the experience of thinking about something physically distant from oneself, but also thinking about the past or future, or people who are very socially dissimilar. It's already the case that people use distance metaphors colloquially when referring to these dimensions of uncertainty. For example, distance metaphors are deeply ingrained with how we talk about time ("That part of my life is behind me now."; "I look forward to working together.") and social relationships ("He acted above his station."; "They're really very close.") (Casasanto & Boroditsky, 2008). Greater psychological distance favors greater abstraction, because abstract construals apply to a greater variety of targets.

Why is this so? Greater distance means greater uncertainty about specifics. When you look out your window at the house across the road, you can't see the shape of every shingle on their roof, or the shade of every brick set in their chimney. Yesterday and tomorrow are harder to observe than today is. When there is a great deal of uncertainty you are probably better off sticking with your model, so you give abstractions more weight.

Objects and events that are closer in time and space are more concrete, and possess specific details that distinguish them as special and idiosyncratic. My coffee mug has specific features (green trim, tragically empty), as does the experience of typing these words (the specific keystrokes involved). Objects and events at greater distance, by contrast, must be treated more abstractly, ignoring peripheral details and highlighting the essential or targeted features that are more likely to be true across time and space. As a result, one factor that influences the strength of top-down processing is psychological distance (Gilead et al., 2019).

Because of the increased uncertainty associated with greater psychological distance, CLT notes that people bring a more abstract perspective to bear on any situation which is extreme on one or more of these dimensions (Liberman & Trope, 2014). The type of distance is interchangeable; to a certain extent it does not matter if a person is considering something far off in time, something far away in space, or some absurd hypothetical. All of these involve a large amount of uncertainty, and so all are approached with a high level of abstraction.

A person thinking about a vacation a year in advance would do well to consider the big-picture aspects of their trip. What sort of destination sounds best? The mountains or the beach? How long should the vacation be for? Should they travel alone, or with family? With friends? And how to get there—a plane, a train, rent a car?

A person thinking about their vacation that starts a week from today will have other concerns. What will the weather be like? What should they pack? What time do they need to get up in order to catch their flight, and how do they get to the airport? What should they set as their out-of-office email message? (see also Vallacher & Wegner, 1987, for hierarchical levels of action identification)

Every vacation has some things in common. The vacationer goes somewhere, possibly with companions. They are gone for some length of time, and they use some method of transportation to get there and to get home again. Other things are different pretty much every time: the supplies you bring, the schedule, and the transportation you use to arrive at your destination.

Considering these aspects in the opposite order seems immediately wrong. Who, in planning next year's vacation, would worry about setting their alarm for that morning, or what clothes to pack? And who, in preparing for a vacation next week, would begin to wonder where they might go and how to get there? Such is the importance of scope!

Objects and events can be represented at varying levels of abstraction (Liberman & Trope, 2008, 2014; Trope & Liberman, 2010), and lower-level construals are relatively more concrete representations that spotlight those specific details that distinguish an object or event as special and idiosyncratic. Higher-level construals, by contrast, are relatively more abstract representations that ignore peripheral details and instead highlight the core and essential features that are true of all possible manifestations of an object or event. Thus, whereas construing a dog as a "Chihuahua" highlights those features that distinguish one dog from another, construing the same dog as a "pet" highlights instead those features that are common of all animal companions (including dogs but also cats and guinea pigs).

It is extremely adaptive to construe objects and events in a manner congruent with their scope. While there do exist a small number of perverse cases where abstract thought is helpful for local problems, or vice versa, the expansion of abstract thought with psychological distance is an effective heuristic. As with the vacationer in the example from above, it keeps us from spending too much time and effort worrying about details that may change before we encounter them, and from spending the same on big-picture questions when there are more pressing details.

A high-level construal treats alternative subordinate lower-level instantiations as being equivalent to each other and to some extent substitutable. Because the central and general aspects of an experience tend to be those that remain invariant across time, space, and perspective, high-level tools that incorporate centrality and generality should allow people to transcend the particularities of the here-and-now, make informed predictions, and therefore to regulate effectively in pursuit of distant ends. Higher level construals are especially useful for making predictions about psychologically distant objects because they are more likely than low-level construals to remain unchanged as one gets closer to an object or farther away

from it. For example, more people use communication devices than cell phones, and therefore the former construal is more useful for predicting the actions of socially distant individuals who may or may not have a cell phone, but who probably have a communication device. Even maintaining perceptual constancy across spatial distance requires abstract, high-level construals: Identifying an object in near and distant locations as being the same requires forming an abstract concept (e.g., a chair) that omits incidental features (e.g., perspective-specific appearances and contextual variations, such as the way a chair's shade falls upon the floor and its retinal size) while retaining essential, relatively invariant features (e.g., the object's overall shape and proportions; Liberman & Trope, 2014). Higher-level, abstract construals thus enable people to think and make predictions about objects across a wider and more expansive range of situations.

According to the predictive coding perspective, greater abstraction means a stronger relative force of top-down or model-based processes. People observing stimuli at high psychological distance should be more influenced by their models of the world, as well as by contextual factors that would interact with those models. Because perceptual illusions depend on the negotiations that occur as top-down and bottom-up processes converge and disagree, they are a measure of the relative strength of the processes. Psychological distance should therefore augment perceptual illusions.

Levin (2015) found that this was the case for a classic geometrical-optical illusion, the Müller-Lyer illusion (Müller-Lyer, 1889). This illusion consists of a straight horizontal line with "arrowheads" at either end, either pointing in towards the center of the line, or out. The orientation of these arrowheads distorts the perception of the line's length, causing it to be perceived as either longer or shorter than it really is. While the effect is quite consistent, over repeated trials participants eventually become more and more accurate when making judgments of the length of these figures. Levin had participants first complete a category-exemplar task, in which they were asked to give examples for each of a series of items, or to name a category to which each item belongs. For example, if the item in question was "Senator," a participant in the category (high-level construal) condition might say "Politician," while a participant in the exemplar (low-level construal) condition might say, "Bernie Sanders." Following the manipulation, participants judged the length of a series of Müller-Lyer figures. Consistent with the predictive coding explanation for this illusion, participants were faster to overcome the illusion at a low level of construal, when they had listed examples of categories rather than categories for examples. Participants from both conditions became more accurate with repeated trials, but those from the exemplar condition became more accurate more quickly.

Recent work from our lab has extended this same design to other illusions and to other methods of manipulating mental abstraction (Ludwin-Peery & Trope, 2022). While mindset manipulations like the category-exemplar task can be very useful, they are not the most reliable method of influencing

Figure 8.1 Oppel-Kundt figure.

mental abstraction. More importantly, it's difficult to use a mindset manipulation in a within-subjects design, and as a result it can be very tricky to use mindset manipulations to detect small or subtle effects.

An alternative to mindset manipulations is embedding a stimuli in a context that itself serves as the manipulation. This allows for within-subjects designs, and is particularly well-suited to psychological distance, as distance can be made immediately apparent in a stimuli by means of occlusion, perspective cues, and so on.

To investigate this, we needed an illusion that could easily appear to be at a greater or lesser distance in the stimuli. We found our answer in the Oppel-Kundt illusion, one of the earliest forms of geometric-optical illusion studied (Kundt, 1863; Oppel, 1855). In this illusion, a space or line divided into multiple sections appears wider than it actually is. For example, consider the Oppel-Kundt figure below (Figure 8.1), with 15 dividers in the filled AB space. In this case, BC is 15% wider than AB, though the two spans often appear equal to observers. In a series of experiments, we manipulated the apparent distance of Oppel-Kundt figures to investigate whether this influenced the strength of the illusion, as is predicted by the theories described above.

To manipulate apparent distance, Oppel-Kundt figures were placed within landscape photographs with a clear foreground and background. This ensured that the apparent distance of the figures from the observer could be readily manipulated, thus manipulating the psychological distance of the stimulus. The figures were placed so that they appeared to be either in a near or far spatial position relative to the viewer, and in such a way that they could potentially be perceived as objects within the scene (cans, fence posts, trees, smokestacks, etc.). This is similar to previous research, where psychological distance has been manipulated by situating stimuli in abstract images (Amit et al., 2012) or landscape scenes (Bar-Anan et al., 2007) with strong perspective cues (Figure 8.2).

Participants were trained to consider the Oppel-Kundt figures as though they were objects within the image, such as fenceposts or smokestacks, to help emphasize their apparent distance and enhance the manipulation. As is normal for this illusion, the dependent variable involved asking participants to estimate the size of the section between lines B and C (the empty space on the right) as it related to the section between lines A and B (the filled space on the left). Because filled spaces appear to be larger than their actual size in

Figure 8.2 Oppel-Kundt figures placed within landscape photographs with a clear foreground and background.

this illusion, when comparing filled and empty spaces to one another, observers tend to estimate the empty spaces to be smaller, in terms of their ratio, than they actually are.

This approach leads to a ratio estimate for each figure, which allows us to estimate the strength of the illusion for each trial. As a result, stimuli created using Oppel-Kundt figures are amenable to continuous reporting, which increases statistical power over a binary choice or "do you see this illusion or don't you" approach. Unsurprisingly, continuous reporting has been used in previous research on the illusion (Mikellidou & Thompson, 2014; Wackermann & Kastner, 2010).

We applied this same approach in a series of studies. In our first experiment, the Oppel-Kundt figures were adjusted in size between conditions, in order to serve as a distance cue, such that the figures in the "far" condition were physically smaller on the viewing screen than figures in the "near" condition. As this change in size was matched to distance condition, this presented a potential confound. As a result, in our second experiment, we experimentally controlled for this confound by fixing pairs of figures at identical on-screen sizes.

In both cases, we observed an effect of perceived distance on judgments of the ratio of our Oppel-Kundt figures, such that the illusion of divided space was stronger when the figures appeared to be situated further away from the viewer. Figure 8.3 below shows the 95% confidence intervals for the estimates of the effect of distance condition for two of the experiments. Neither

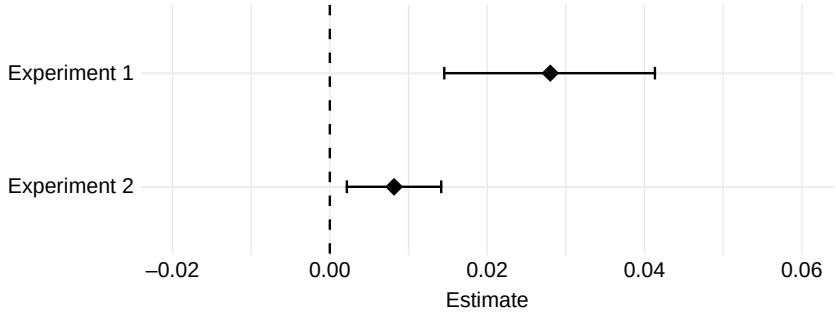

Figure 8.3 The 95% confidence intervals for the estimates of the effect of distance condition for two of the experiments.

confidence interval overlaps with zero, and both estimates are positive, indicating that the illusion was more pronounced in the "far" condition than in the "near" condition. These findings are very much in line with the predictions made by these theories of abstraction (Ludwin-Peery & Trope, 2022).

Cognitive illusions provide particularly clear evidence for this viewpoint, as they very closely match the theory's account of sense data being systematically adjusted by an extremely strong prior. But there's no reason to suspect that the explanatory power of this theory is limited to illusions, or even to perception. Notably, it fits findings from across cognitive science and social cognition. Social cognition is neither objective nor rational, and there's good evidence that our social thought is guided by our expectations. We should expect to see similar predictive effects in many other domains, and in fact we do.

Consistent with this perspective, Hansen (2019) found that ostensibly irrelevant contextual information had a greater impact on judgment when participants experienced higher psychological distance. In a series of experiments, participants were primed with high-level or low-level construal mindset using a category-exemplar task (as above). Following the manipulation, participants sampled beverages in cups of different colors and rated the beverage on several dimensions. It was found that participants at a high level of construal rated sparkling water as more refreshing when it was in a blue cup, an energy drink as more sour when it was in a yellow cup, and coffee as hotter when it was in a red cup, compared to control colors. The author interpreted these results from the perspective of multisensory integration, but the results are also consistent with a predictive coding account of perception. When subjects were in a more abstract mindset, the effect of context had more influence on the interpretation of new stimuli, just as in the illusions.

The theory also clearly explains why people, even experts, have their perceptions heavily shaped by context, and find themselves unable to shut

information out (Dror et al., 2006). It also explains our striking blindness when encountering situations that we do not expect to experience (Chabris et al., 2011; Simons & Chabris, 1999).

Person Perception

The top-down processes engaged by predictive coding manifest not only in how we perceive the physical object but also in how we perceived the social world. A large amount of research has demonstrated that our perception of other people depends on our expectancies about them, our stereotypes about the groups they belong to, and the social context in which they are observed (see review by Jost & Trope, 2013; Trope & Higgins, 1993a). Social cognitive researchers have been particularly interested in how we perceive human faces. The ability to read others' mental states and dispositions from their faces confers considerable adaptive advantages. It would enable you to tell, in advance, whether someone holding a knife intends to use it to attack you or to defend you. Indeed, faces are a source of visual information that we heavily rely on for drawing inferences about people. We use them to tell us about a host of others' personal characteristics—their emotions, intentions, attitudes, and personality dispositions (Todorov, 2017).

Top-Down Influences on Face Perception

From a predictive coding perspective, however, the perception of the face may itself be affected by top-down processes triggered by personal expectations about others, their group stereotypes, and the social context. Two lines of research on face perception fit and anticipate this perspective, and illustrate such top-down influences.

The first is research on physiognomy, specifically on what Hassin and Trope (2000) called reading into the face (RIF). The example they use is that we are unlikely to think that Einstein's forehead is short. If true, they ask, is this because there is something about his forehead, or is it because of something we know about Einstein? The RIF hypothesis suggests that it is what we know about Einstein that shapes the way we perceive his facial features. Without knowing about Einstein, "…he might be judged to be an amiable old man, who invents things that never work, ruining everything he lays his hands on and being a great nuisance to Aunt Jane, that practical and efficient old lady." This man's forehead, RIF suggests, will be perceived as smaller than Einstein's.

Consistent with RIF, Hassin and Trope (2000, Study 5) found that people have well-developed expectations about facial features of people with different personality traits. Kind people, compared to mean people, are expected to have rounder eyebrows, bigger eyes, lower cheek bones, and a wider face. In a study directly testing RIF, Hassin and Trope (2000, Study 6) described a person as either mean or kind. The mean description included statements like

"His friends note that he is extremely cynical, and that his critical sense of humor offends many of his acquaintances… he just enjoys seeing people squirm." The kind description indicated that "His friends say that his kindness is exceptional, and that he cannot say 'no' to any of his friends or family's requests… His pleasantness and kindness are both very special and very rare," Participants' then saw the person's face and were asked to judge his facial features. Consistent with RIF, participants' perception of the person's facial features depended on the description they read. For example, the same face was perceived as rounder, fuller, shorter, and wider when the person was described as kind, compared to the face of the person described as mean.

Further evidence for top-down processing of faces comes from more recent research on the effect of context on the perception of emotions in human faces (Aviezer et al., 2012a). We were specifically interested on the influence of a person's body on the identification of the emotion expressed by that person's face. The research used peak-intensity facial expressions elicited in a wide variety of emotional situations. For example, one study presented participants with peak expressive reactions to winning and losing points in professional high-stakes tennis matches that typically evoke strong affective reactions. Participants saw either the full image (face + body), the body alone, or the face alone.

When they saw the face alone, participants could not tell if the athlete had just won or lost a point and failed to rate the affective valence of winners as more positive than the affective valence of losers. However, they succeeded at distinguishing between winners and losers when they saw the the body and the face together. Most remarkably, they also succeeded when they saw the body alone, with the face totally obscured. This is especially impressive because it ran counter to people's explicit predictions. A full 80% of Aviezer et al.'s participants said that the face would be most diagnostic for affective valence discrimination, 20% chose the full image of the face in combination with the body, and none chose the body alone. In short, people told the authors that most or all the relevant information would be in the athlete's face.

Participants' belief in the face as a "window to the soul" was apparently so strong that they misattributed positive or negative affect to an ambiguous face, despite the fact that the source of their perception was actually the top-down influence of the affect in the body context (see also, Aviezer et al., 2012b). In a second study, participants were shown either the face of an athlete who had just won a point on the body of an athlete who had just lost a point, or the face of an athlete who had just lost a point on the body of an athlete who had just won a point. When participants saw a loser's face on a winner's body, they were likely to say that person's facial expression was positive, and they saw a winner's face on a loser's body, they were likely to say that that person's expression was negative. As before, the actual facial expression made little difference.

Even more striking are the results of Aviezer et al.'s more stringent test of whether participants' perception of the faces actually changes depending on

the body. Rather than rating the faces, participants were asked to simulate in their own face the exact facial movements of the tennis players. Aviezer et al. found that the simulation of identical faces shifted depending of the body's affective valence. Specifically, losing faces were simulated as more positive when the simulator viewed them on winning bodies than on losing bodies. Conversely, winning faces were simulated as more negative when the simulator viewed them on losing bodies than on winning bodies.

The top-down influences on face perception uncovered by Aviezer et al. (2012a) and Hassin and Trope (2000) have important implications for inferences about others' personality dispositions. Physiognomy, the art of reading traits from the face, has been practiced through the centuries since the times of the ancient Greeks (Zebrowitz, 1997). The belief in physiognomy has persisted in modern time. A survey of 535 respondents conducted by Hassin and Trope found that 75% of the respondents believed that it was possible to know an individual's personality traits from his or her face. As described earlier, perceivers' facility with inferring traits from faces is evinced by the extensive research by Jim Uleman's pioneering work on spontaneous trait inferences (see review by Uleman et al., 2008) and the research it has spurred by Alex Todorov (see review by Todorov, 2017) and Jon Freeman (see review by Freeman, 2018) on face perception.

From the present predictive coding perspective, the Aviezer et al. (2012a) findings have unique implications for the fundamental attribution error, the tendency to attribute behavior to the corresponding personal dispositions despite situational inducements present in the context where the behavior occurs (Gilbert & Malone, 1995; Jones, 1979). These implications have been laid out by Trope (1986) in the two-stage dispositional inference model. According to this model, the over-attribution of personal traits may be due to the top-down contextual influences on the perception of people's behavior. For example, a frown may be nondiagnostic of a specific emotion. However, in a dangerous situation frowning will be perceived as fearful facial expression, whereas in an adversarial situation it will be perceived as an angry facial expression. Because people treat these perceptions as real rather than as top-down derivations from their contextual expectations, they may use them as independent evidence about the actor and erroneously infer that he or she is dispositionally anxious or hostile (see e.g., Trope, & Alfieri, 1997; Trope & Cohen, 1989; Trope & Gaunt, 2000; Trope et al., 1988; Trope et al., 1991). These top-down processing effects reflect predictive coding in that they involve situational predictions altering the nature of incoming bottom-up info.

Spontaneous Trait Inferences across Psychological Distance

Spontaneous traits inferences (STIs) represent relatively abstract, high-level construals of others compared with behaviors. Traits are high level in that they are relatively (1) more abstract and stable, (2) more central to the person representation (i.e., changing a person's personality trait changes the

person more fundamentally than changing a specific behavior), and (3) superordinate to behaviors in that traits as internal causes bring about specific behaviors (i.e., being *clever* causes one to *solve the mystery halfway through the book*). Thus, a CLT analysis of STIs suggests that STI formation will be enhanced for psychologically distal versus proximal actors, and research supports this idea. Rim et al. (2009) found that perceivers formed more STIs from behaviors of others who were described as being in a spatially remote versus proximal location (Study 1) and from behaviors believed to have occurred in the distant versus recent past (Study 2); amount and content of actor information were held constant across distances. These effects were not attributable to differences in perceived similarity with the actors (Studies 1 and 2) or to differences in level of familiarity with the distal and proximal locations (Study 1). Rim et al. (2009) speculated that STIs are more functional for representing distant others because the specifics of the immediate situation (e.g., exact behaviors) may not always hold for those individuals. Abstract traits are more invariant across psychological distance and hence are more useful in representing distal people (Moskowitz & Olcaysoy Okten, 2016). This set of studies on STIs provides the most direct evidence that perceivers use abstract traits to represent psychologically distant others. This, in turn, expands the spatiotemporal scope of people perceivers can relate to.

These findings also have interesting implications for the actor–observer effect, the tendency for people to attribute others' behaviors (failing an exam) to dispositional causes (because he is stupid) while attributing one's own behaviors to situational causes (because the exam was unfair). In terms of CLT, others are, by definition, more distant from the self than the self is from the self. Therefore, the fact that others' behaviors are thought of in terms of traits more than the same behaviors by the self is consistent with CLT. However, differences in amount of information and differences in informational salience could account for these effects as well. Thus, an important question is whether psychological distance affects the tendency to give a dispositional attribution for an actor's situationally constrained behavior, controlling for the nature and amount of information given. Nussbaum et al. (2003) found the answer to be affirmative. Participants made stronger correspondent attitude inferences from a constrained essay after they had made judgments regarding the writers' distant versus near future behaviors. Henderson et al. (2006) replicated this effect manipulating spatial distance. That is, perceivers were more likely to ignore situational information and draw correspondent inferences when the actor was believed to be spatially remote versus proximal.

In sum, consistent with the current predictive coding framework, both STIs and more direct measures of traits inferences suggest that perceivers rely on high-level trait construals when called upon to make judgments about an actor in temporally or spatially distant situations. Such construals thus serve to expand the scope of humans' social relations beyond the demands of the immediate situation.

Conclusion

A central tenet of social psychology is that human respond to their subjective construal of the surrounding situation (Lewin, 1951; Ross & Nisbett, 1991; Zimbardo et al., 2003). Predictive coding provides a general framework for understanding the construal process. This framework tells us that our pre-existing knowledge structures give rise to expectancies that in turn profoundly shape the most basic aspects of our perceptions. In this chapter, we have reviewed some research from our lab that illustrates how expectancies, through top-down processing, can drastically alter our visual perception of objects and people. We have highlighted the circumstances that modulate top-down effects. We specifically argued that uncertainty, the concomitant of psychological distance from objects, modulates expectancy effects on their perception and the resulting visual illusions. Correspondingly, we argued that ambiguity of social behavior, in general, and facial expressions, in particular, modulates expectancy effects on their perception and the resulting over-attribution of personality traits. We studied expectancies deriving from the immediate context surrounding the objects and people we observe. However, as social psychologists, we should note, that those expectancies themselves originate in our social cultural knowledge. In this sense, predictive coding is where the social and the psychological meet and shake hands.

References

Amit, E., Mehoudar, E., Trope, Y., & Yovel, G. (2012). Do object-category selective regions in the ventral visual stream represent perceived distance information? *Brain and Cognition*, *80*(2), 201–213.

Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology*, *41*(3), 258–290.

Aviezer, H., Trope, Y., & Todorov, A. (2012a). Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science*, *338*, 2225–2229.

Aviezer, H., Trope, Y., & Todorov, A. (2012b). Holistic person processing: Faces with bodies tell the whole story. *Journal of Personality & Social Psychology*, *12*, 20–37.

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., … et al. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences*, *103*(2), 449–454.

Bar-Anan, Y., Liberman, N., Trope, Y., & Algom, D. (2007). Automatic processing of psychological distance: Evidence from a stroop task. *Journal of Experimental Psychology: General*, *136*(4), 610.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge, England: Cambridge University Press.

Bruner, J. S. (1957). On perceptual readiness. *Psychological Review*, *64*(2), 123.

Cameron, J. A., & Trope, Y. (2004). Stereotype-biased search and processing of information about group members. *Social Cognition*, *22*, 650–672.

Cantor, N., & Mischel, W. (1979). *Advances in experimental social psychology*, (Vol. 12, pp. 3–52). Academic Press.

Casasanto, D., & Boroditsky, L. (2008). Time in the mind: Using space to think about time. *Cognition*, *106*(2), 579–593.

Chabris, C. F., Weinberger, A., Fontaine, M., & Simons, D. J. (2011). You do not talk about Fight Club if you do not notice Fight Club: Inattentional blindness for a simulated real-world assault. *i-Perception*, *2*(2), 150–153.

Chaiken, S., Liberman, A., & Eagly, A. H. (1989). Heuristic and systematic information processing within and beyond the persuasion context. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 212–252). The Guilford Press.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204.

Clark, A. (2015). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford University Press.

Cronbach, L. J. (1955). Processes affecting scores on "understanding of others" and "assumed similarity". *Psychological Bulletin*, *52*(3), 177–193.

Dror, I. E., Charlton, D., & Péron, A. E. (2006). Contextual information renders experts vulnerable to making erroneous identifications. *Forensic Science International*, *156*(1), 74–78.

Freeman, J. B. (2018). Doing psychological science by hand. *Current Directions in Psychological Science*, *27*, 315–323.

Fiske, S. T., Lin, M. H., & Neuberg, S. L. (1999). The continuum model: Ten years later. In S. Chaiken , & Y. Trope (Eds.), *Dual process theories in social psychology*, (pp. 231–254). New York: Guilford.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological sciences*, *360*(1456), 815–836.

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, *117*(1), 21–38.

Gilead, M., Trope, Y., & Liberman, N. (2019). Above and beyond the concrete: The diverse representational substrates of the predictive brain. *Behavioral and Brain Sciences*, *43*, 1–63.

Hansen, J. (2019). Construal level and cross-sensory influences: High-level construal increases the effect of color on drink perception. *Journal of Experimental Psychology: General*, *148*(5), 890.

Hassin, R., & Trope, Y. (2000). Facing faces: Studies on the cognitive aspects of physiognomy. *Journal of Personality and Social Psychology*, *78*, 837–852.

Heckhausen, H. (1986). Why some time out might benefit achievement motivation research. In J. H. L. van den Bercken, T. C. M. Bergen & E. E. J. De Bruyn (Eds.), *Achievement and task motivation* (pp. 7–39). Lisse, The Netherlands: Swets & Zeitlinger.

Heider, F. (1958). *The psychology of interpersonal relations*. John Wiley & Sons.

Helmholtz, H. (1860). *Handbuch der physiologischen Optik*. Leipzig: Voss.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83.

Henderson, M. D., Fujita, K., Trope, Y., & Liberman, N. (2006). Transcending the "Here": The effect of spatial distance on social judgment. *Journal of Personality and Social Psychology*, *91*(5), 845.

Johnson, M. A. (1986). Color vision in the peripheral retina. *American Journal of Optometry and Physiological Optics*, *63*(2), 97–103.

Jones, E. E. (1979). The rocky road from acts to dispositions. *American Psychologist*, *34*(2), 107.

Jones, E. E., & Davis, K. E. (1965). From acts to dispositions the attribution process in person perception. *Advances in Experimental Social Psychology*, *2*, 219–266.

Jost, J. T., & Trope, Y. (2013). Interpersonal expectancies: Where the social meets the psychological. In Slawomir Trusz (Ed.), *Interpersonal expectancy effects: Essential readings*. (pp. 29–32), Warsaw: Wydawnictwo Naukowe Scholar.

Kleinschmidt, A., Sterzer, P., & Rees, G. (2012). Variability of perceptual multistability: From brain state to individual trait. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *367*(1591), 988–1000.

King, D. J., Hodgekins, J., Chouinard, P. A., Chouinard, V. A., & Sperandio, I. (2017). A review of abnormalities in the perception of visual illusions in schizophrenia. *Psychonomic Bulletin & Review*, *24*(3), 734–751.

Kundt, A. (1863). Untersuchungen über Augenmaass und optische Täuschungen. *Annalen der Physik*, *196*(9), 118–158.

Lewin, K. (1951). *Field theory in social science*. New York: Harper.

Levin, Z. (2015). The truth is in the details: Concrete thinking accelerates illusion decrement in visual adjustment of the Müller-Lyer figure. (unpublished MA thesis).

Liberman, N., & Trope, Y. (2014). Traversing psychological distance. *Trends in Cognitive Sciences*, *18*(7), 364–369.

Liberman, N., & Trope, Y. (2008). The psychology of transcending the here and now. *Science*, *322*(5905), 1201–1205.

Lippmann, W. (1922). The world outside and the pictures in our heads. In W. Lippmann (Ed.), *Public Opinion* (pp. 3–32). New York: MacMillan.

Ludwin-Peery, E., & Trope, Y. (2022). Psychological distance increases the strength of visual illusions. *Manuscript in preparation*.

Markus, H. (1977). Self-schemata and processing information about the self. *Journal of Personality and Social Psychology*, *35*(2), 63–78.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*(2), 227–234.

Mikellidou, K., & Thompson, P. (2014). Crossing the line: Estimations of line length in the Oppel-Kundt illusion. *Journal of Vision*, *14*(8), 20–20.

Moskowitz, G. B. (2005). *Social cognition: Understanding self and others*. New York: Guilford Press.

Moskowitz, G. B., & Olcaysoy Okten, I. (2016). Spontaneous goal inference (SGI). *Social and Personality Psychology Compass*, 10(1), 64–80.

Moskowitz, G. B., Olcaysoy Okten, I., & Schneid, E. (in press). The updating of first impressions. In E. Balcetis & G. B. Moskowitz (Eds.), *Handbook of impression formation*. New York: Psychology Press/Taylor and Francis.

Müller-Lyer, F. (1889). Optische Urteilstauschungen. *Archiv fur Anatomie und Physiologie, Physiologische Abteilung*, *2*, 263–270.

Nussbaum, S., Trope, Y., & Liberman, N. (2003). Creeping dispositionism: The temporal dynamics of behavior prediction. *Journal of Personality and Social Psychology*, *84*(3), 485.

Oppel, J. J. (1855). Uber geometrisch-optische Tauschungen. *Jahresbericht des Physikalischen Vereins zu Frankfurt am Main*.

Rim, S., Uleman, J. S., & Trope, Y. (2009). Spontaneous trait inference and construal level theory: Psychological distance increases nonconscious trait thinking. *Journal of Experimental Social Psychology*, *45*(5), 1088–1097.

Ross, L., & Nisbett, R. (1991). *The person and the situation. Perspectives of social psychology*. New York: McGraw-Hill.

Schank, R. C., & Abelson, R. P. *Scripts, plans, goals, and understanding*. Hillsdale, NJ: Erlbaum, 1977.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception*, *28*(9), 1059–1074.

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences*, *1*(7), 261–267.

Stephan, E., Liberman, N., & Trope, Y. (2010). Politeness and psychological distance: A construal level perspective. *Journal of Personality and Social Psychology*, *98*(2), 268–280.

Todorov, A. (2017). *Face value: The irresistible influence of first impressions*. Princeton University Press.

Trope, Y. (1974). Inferential processes in the forced compliance situation: A Bayesian analysis. *Journal of Experimental Social Psychology*, *10*, 1–16.

Trope, Y. (1986). Identification and inferential processes in dispositional attribution. *Psychological Review*, *94*, 237–258.

Trope, Y. (1989). Perceptual and inferential effects of stereotypes. In D. Bar-Tal, C. F. Graumann, A. W. Kruglanski & W. Stroebe (Eds.), *Stereotypes and prejudice: Changing conceptions*. New York: Springer-Verlag.

Trope, Y., & Alfieri, T. (1997). Effortfulness and flexibility of dispositional judgment processes. *Journal of Personality and Social Psychology*, *73*(4), 662.

Trope, Y., & Cohen, O. (1989). Perceptual and inferential determinants of behavior-correspondent attributions. *Journal of Experimental Social Psychology*, *25*(2), 142–158.

Trope, Y., Cohen, O., & Alfieri, T. (1991). Behavior identification as a mediator of dispositional inference. *Journal of Personality and Social Psychology*, *61*(6), 873.

Trope, Y., Cohen, O., & Maoz, Y. (1988). The perceptual and inferential effects of situational inducements on dispositional attribution. *Journal of Personality and Social Psychology*, *55*(2), 165.

Trope, Y., & Gaunt, R. (2000). Processing alternative explanations of behavior: Correction or integration? *Journal of Personality and Social Psychology*, *79*(3), 344.

Trope, Y., & Higgins, E. T. (1993a). *Dispositional inferences from behavior*. New York: Sage Publications.

Trope, Y., & Higgins, E. T. (1993b). The what, how, and when of dispositional inference: New questions and answers. *Personality and Social Psychology Bulletin*, *19*, 493–500.

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review*, *117*(2), 440–463.

Trope, Y., & Thompson, E. P. (1997). Looking for truth in all the wrong places? Asymmetric search of individuating information about stereotyped group members. *Journal of Personality and Social Psychology*, *73*, 229–241.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, *5*(2), 207–232.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science*, *185*(4157), 1124–1131.

Uleman, J. S., Adil Saribay, S., & Gonzalez, C. M. (2008). Spontaneous inferences, implicit impressions, and implicit theories. *Annual Review of Psychology*, *59*, 329–360.

Vallacher, R. R., & Wegner, D. M. (1987). What do people think they're doing? Action identification and human behavior. *Psychological Review*, *94*, 3–15.

Wackermann, J., & Kastner, K. (2010). Determinants of filled/empty optical illusion: Search for the locus of maximal effect. *Acta Neurobiol Exp (Wars)*, *70*(4), 423–434.

Wakslak, C., & Trope, Y. (2009). The effect of construal level on subjective probability estimates. *Psychological Science*, *20*(1), 52–58.

Zebrowitz, L. (1997). *Reading faces: Window to the soul?* Routledge.

Zimbardo, P. G., Weber, A. L., & Johnson, R. L. (2003). *Psychology: Core concepts* (5th ed.). Boston: Allyn & Bacon.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Part II

Impression Formation Processes: Implicit Effects of Inference and Activation

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

9 Reflections on a 30-Year-Long Program of Research Exploring Perceivers' Spontaneous Thoughts about Social Targets

John J. Skowronski and Randy J. McCarthy
Northern Illinois University

Into the decade of the 1970s, theories explaining processes by which people made trait inferences about others insinuated that trait inference-making was an effortful and lengthy cognitive process (i.e., see Fiske & Taylor, 1991, p. 22). As the 1970s waned and the 1980s waxed, several scholars challenged this idea (see Smith, 1989; Uleman, 1989). These scholars claimed that trait inferences could sometimes be made unintentionally (or *spontaneously*, as in ***spontaneous trait inference, or STI***, the term that was adopted to describe this process; see Winter & Uleman, 1984).

One research program involving this chapter's authors (usually in collaboration with other scholars) was profoundly influenced by these ideas. This chapter reflects on this 30-year-long program of research, reviews the origins of the research, and presents findings that have emerged from the research program. Moreover, this chapter also offers suggestions about new research that might be pursued by those who are interested in the spontaneous thoughts that perceivers might generate as they encounter information about social targets. Finally, in addition to providing a description of the research program, we also hope that the chapter content will aid new scholars who wonder what their mentors mean when the mentors say that scholars need to develop a "research program."

The Origin Story: Two Guys in a Mini-Van and the Duck Conferences

The story of this program of research and theory begins with two guys (Donal Carlston and John Skowronski) in a mini-van. Their destination was the Duck Conferences on Social Cognition, a small yearly gathering of scholars who pursued work in social cognition and related areas. For a number of years, Don and John made the cross-country trek to the Duck Conferences on Social Cognition together in Don's Chrysler mini-van. Like the open road before them, their conversations meandered for hours. Considerable amounts of scholarship, including the Don and John–led program of research exploring spontaneous inferences that observers make about actors, emerged

DOI: 10.4324/9781003045687-11

from the interactions that took place in that mini-van, at the places that were visited on the drive, and at the Duck Conferences themselves.

Across several years, the Duck Conference scholars observed many presentations that Don and John jointly made. The interplay between Don and John, and between the audience and the two presenters, was extremely stimulating and beneficial to the presenters, as were the long walks on the beach during which audience members provided additional ideas about the research. In this regard, we note that one frequent Duck attendee was prominent STI scholar Jim Uleman. Among the many scholars who influenced the Don and John research program, it was Jim's commentary that most often prompted Don and John to probe harder and longer in their research.

The Spark That Lit the Fire

In 1984, Winter and Uleman published a manuscript suggesting that perceivers infer traits *spontaneously*—that is, they infer traits even when there were no explicit instructions to form trait inferences (Winter & Uleman, 1984). This was a bold claim. Many scholars thought that inference-making only happened in the presence of explicit inference-making goals (e.g., Hamilton, 1981; for a direct empirical rebuttal to Hamilton, see Uleman & Moskowitz, 1994).

During one of Don and John's van rides, Don offered concerns about the research paradigm that was the basis of the Winter and Uleman claim. To understand the concerns, readers need to know a bit about the paradigm Winter and Uleman (1984) used to make their claims of spontaneously inferred traits. Winter and Uleman's paradigm used the principle of encoding specificity to garner evidence for spontaneous inference making. The encoding specificity principle states that when an effort is made to recall learned material, the cues present when that material was encoded (including self-generated thoughts) serve as effective retrieval cues. Winter and Uleman reasoned that if people make trait inferences when encoding trait-relevant behavioral sentences, the (internally generated) trait and the behavioral sentence would be contiguously presented and, according to encoding specificity, the trait should provide effective retrieval cues for those sentences. For instance, if on reading the behavior description "Mark left a 40% tip" the reader infers that "Mark is generous," then the behavior (actual stimuli) and the trait inference (internally-generated thought) are contiguous during encoding. Thus, Winter and Uleman reasoned that presenting the trait "generous" would be a useful cue to recall the behavior "he left a 40% tip." This is indeed what they found.

The Winter and Uleman research seemed to indicate that this method was viable as a way to detect trait inference-making (for a review of similar results beyond those described by Winter & Uleman, 1984, see Uleman et al., 1996). Moreover, the fact that evidence suggestive of trait inference-making emerged from the method, even when participants were not explicitly told to

make inferences or to understand actors, led Winter and Uleman to conclude that people routinely and *spontaneously* made trait inferences about actors when encoding descriptions of actor behaviors.

However, Don articulated two concerns about the Winter and Uleman (1984) results. For one, Don argued that the Winter and Uleman results might not reflect trait inferences made at behavior encoding. This is because even if perceivers had not made trait inferences when initially exposed to stimulus behaviors, trait terms might nonetheless provide effective retrieval cues because of participants' prior knowledge about traits and behaviors. For example, for most perceivers the cue dishonest is linked in memory to dishonesty-prototypic behaviors, such as lying and stealing. Thus, trait cues that prompt access to these prototypic behaviors could also aid the retrieval of specific stimulus episodes that exemplify lying or stealing. A second issue raised by Don was that trait inferences might have been generated during behavior encoding, but those inferences only described the behaviors, NOT the actors. In other words, the observation that Erika kicked a dog may prompt a perceiver to think "that was a mean behavior," but not necessarily to think "Erika is mean." Indeed, Don noted that the Winter and Uleman data showed only that the trait term cued recall of the behavior, but did not show that the trait term was linked to the ***actor*** who performed the behavior.

The Trait Relearning Paradigm and the Savings Measure: An Overview

The Uleman team eventually came to recognize these concerns with the evidence from the encoding-specificity paradigm and conducted various programs of research to address them (e.g., see Uleman et al., 1996). However, Don and John approached the phenomenon with a different method altogether. Don suggested that the problems with the encoding-specificity paradigm could be addressed via the use of a research finding often attributed to Ebbinghaus (1885/1964). Ebbinghaus (and others; for a review, see Nelson, 1985) showed that it is generally easier to relearn information than to learn it for the first time. This advantage emerges even after intervals of several years and even after the originally learned material can no longer be intentionally recalled. This learning advantage is generally quantified via a "savings" measure. The ***savings effect*** is the decrease in trials or time needed to learn material at re-learning as opposed to at initial learning.

Don conceived of a paradigm that could use the savings measure to assess spontaneous inference-making about actors. The logic of how savings-in-relearning could be applied to STI was straightforward. At Time 1, perceivers would see a photo of a person paired with a behavior that had trait implications. At Time 2, perceivers would see these photos again, now paired with trait words. At Time 3, the extent to which perceivers could recall each trait word when cued by the photo with which it was previously paired at Time 2 would be assessed. If people made trait inferences about photographed

actors during Time 1, those traits would be more easily associated with those actors at Time 2 as evidenced by higher rates of actor photo cued recall at Time 3. Indeed, when Don and John used this paradigm in research, this is what they found. *The savings effect in these studies reflects recall rates in the conceptual relearning condition that exceed recall rates in the control condition.*

The savings-in-relearning paradigm solves both of Don's concerns with the Winter and Uleman (1984) paradigm. In the relearning paradigm, the behaviors are never re-presented. Thus, behavior re-presentation cannot cue generation of a trait that could facilitate behavior recall, as could be the case in the Winter and Uleman paradigm. Moreover, the proposed savings measure in the new paradigm assessed, and found evidence for, a link between the ***person in the photo*** and the trait. Data suggesting the presence of such a link were weak to non-existent in results from studies using the Winter and Uleman cued recall procedure.

The relearning paradigm is not the only trait inference-detection paradigm that has been used in the line of research that emanated from the Don and John van rides. However, we highlight it here because it has been used frequently and is a unique methodological contribution of the research program that emanated from the Don and John van rides. Moreover, this method helped to determine the direction of some of the research in the program. That is, because the savings measure is an ***indirect measure*** of inference-making, many studies in the research program attempted, at least in part, to validate the measure.

Thus, the story of the research program told in this chapter reflects two major themes. One theme is methodological, focused on establishing the validity of the savings measure as an indicator of trait inference-making. A second theme is more substantive and focuses on understanding inferences made about actors. Major issues in this second theme include (1) the extent to which such inferences are made spontaneously, (2) the kinds of inferences that might be made (e.g., trait, evaluative), (3) the conditions under which spontaneous inferences are made, (4) the processes involved in the generation of such inferences, and (5) the mental and behavioral consequences of such inferences.

Thoughts about STI Findings from the Research Program

Evidence of STI Generation as Measured via Savings Is Not Affected by Impression Formation or Memory Goals

Many of these themes were in evidence right at the start of the research program. For example, consider the initial study that looked for evidence of STIs using the relearning paradigm (Carlston & Skowronski, 1994). Participants in the study were given one of three sets of instructions. One group was exposed to the Time 1 behavior-photo pairs without being told much other than "we'll use this stuff later" (the ***uninstructed condition***).

A second group was given rather ***general instructions to form impressions*** of the people in the photos from the Time 1 photo-behavior pairings. A third group was specifically told to ***form trait impressions*** of the people in the photos from the Time 1 photo-behavior pairings.

Their research design indicates an interest in (1) whether the paradigm would yield a savings effect when trait-photo learning at Time 2 conceptually constituted relearning (indicating that it was sensitive to trait inference-making—a measure validation goal); (2) the extent to which a savings effect would appear in the uninstructed condition (providing evidence that trait inferences could be made *spontaneously*), and (3) whether the magnitude of the savings effect would differ across the three instruction conditions (potentially providing evidence that trait inference-making was more prevalent when perceivers were focused on understanding people and/or their traits then when they were not).

The *a priori* expectation was that the study would find evidence that the Winter and Uleman (1984) procedure substantially overstated the evidence indicative of STI generation. That is, it seemed to Don and John that while evidence of inference generation might weakly appear in the uninstructed condition (perhaps because the behaviors were pretested to have strong trait implications), they also expected that evidence for inference generation would be significantly stronger when participants were given the explicit goal (in the general impression condition or the specific trait generation condition) to think about actor trait dispositions. Hence, Don and John's expectation was that they would find evidence both for STI generation, and for the attribution-derived idea that inference generation rates increased when people had an explicit goal to think about others' dispositions.

One of this chapter's authors (JJS) often told his students that he made a career out of being wrong and often used this experiment's results to illustrate that point. In contrast to expectations, the results revealed a very robust savings effect that ***was about the same magnitude across all three instruction conditions***. The robust outcome in the no-instruction condition supported the Winter and Uleman (1984) assertion that people routinely make spontaneous trait inference about others. Importantly, the result added emphasis to the Winter and Uleman assertion by showing that (1) the STI evidence emerged in a task in which the behavior was not re-presented (as in the Winter and Uleman paradigm), so the effect could not be attributed to trait generation during the recall task; (2) the trait was not just linked to the behavior, but was linked to the actor, and (3) evidence of inference generation was not enhanced by a prior perceiver goal to understand actor trait dispositions.

That savings effects emerge in the relearning paradigm when participants are essentially uninstructed, and that the effects are not altered for participants who are given processing goals that might be expected to enhance trait inference-making, have been replicated numerous times (Carlston & Skowronski, 1994; Carlston et al., 1995). Moreover, results from one

additional study (Carlston & Skowronski, Experiment 4) showed that the magnitude of the savings effect was also largely unaffected when participants were given the goal to remember the photos and the stimulus behavior that accompanied each photo. Finally, results from one more attempt to examine effects of encoding conditions on savings effects showed they were similar for participants who encountered either an informant's trait-implicative self-description or an informant who described the behavior of a third-party actor who participants are led to believe is similar to the actor (Carlston et al., 1995, Experiment 4).

A Note of Caution

The results reported so far might lead one to believe that trait inference generation occurs with a high degree of automaticity, so will rarely or never be related to the mental goals that might be in place when participants encounter the behaviors of an actor. Some might also be tempted to claim that trait generation might be the dominant form of spontaneous thought in response to encounters with actor behavior.

However, we encourage readers to eschew such conclusions. One reason is that the studies in this research program have tended to use behaviors that are pretested to have strong and unambiguous trait implications. Clearly, other actor behaviors might not have any trait implications at all, but might instead tend to prompt inferences about actor goals, actor emotional states, or even environmental influences on behavior. One might not find much evidence of STI generation with such behaviors.

Moreover, perceiver processing goals might be expected to affect trait inference-making for behaviors with weaker trait implications or for behaviors that have multiple implications. For example, when behaviors have simultaneous and strong implications for both traits and emotional states, it seems plausible that processing goals in place prior to encountering actor behaviors might alter the nature of the inferences drawn from the behavior. Such effects might also be expected to occur if participants are exposed to training designed to habitualize either trait or emotional state inference-making: Those trained to make trait inferences and those trained to make state inferences might be expected to spontaneously make different inferences from behaviors that have either trait or emotional state implications.

Even more importantly, results from additional studies (e.g., McCarthy & Skowronski, 2011a) in the research program have shown that even when encountering actor behaviors with strong trait implications, some mental goals, such as the goal to detect lying, or the goal to detect certain letter strings in sentences, can interfere with STI generation. Hence, while the process of generating STIs contains elements of automaticity, most notably (and by definition) *spontaneity*, this process is almost certainly ***not fully automatic*** because it is not impervious to other task demands that can be placed

on the perceiver (for additional evidence on this point from other paradigms, see Uleman & Moskowitz, 1994).

This conclusion is bolstered by results from studies showing that the tendency to engage in STIs can vary across conditions. One such set of studies was reported by Crawford et al. (2013). In one study, participants who were uninstructed as to how to think about the stimuli completed the Time 1 task in the relearning paradigm while enacting either an approach behavior (arm flexion) or an avoidance behavior (arm extension). Results from the savings measure indicated that savings was enhanced for negative trait terms when the participants were engaged in the arm extension behavior at Time 1, but for the positive trait terms when participants engaged in the arm flexion behavior at Time 1.

A similar result was produced in Study 2, which employed a false recognition paradigm to detect trait inferences. In the false recognition paradigm, participants sometimes encounter actor behaviors in which a trait term is included in the behavior description (e.g., he was clumsy and slipped on the ice), and sometimes encounter sentences in which the trait is not included. Participants are later given a cue (e.g., actor photo) and are asked to report whether a trait term appeared in the description that accompanied the photo. A false positive response (answering "yes" when the trait was not included) is seen as evidence for STI. The data reported by Crawford et al. (2013) in Study 2 showed that false recognition was especially great for positive traits when participants were exposed to a physically warm stimulus during Time 1, and was especially great for negative traits when participants encountered a physically cold stimulus during Time 1. In addition, in this second study the Jacoby (1991) process dissociation approach was used to calculate the degree to which processing in the false recognition task might be either automatic or controlled. Results from the analyses suggested that the effect of the warm/cold stimuli was to increase the impact of automatic processes on performance in the false recognition task on those trials where the trait and the warm/cold stimulus were congruent. This latter shift should not occur if trait inference making were always fully automatic.

Results from a series of studies reported by McCarthy and Skowronski (2014) also supported the notion that STI generation is not fully automatic by showing that it varies across circumstances. These authors speculated that the tendency to generate STIs might be pushed by the need to interact with others (for a similar idea, see Uleman et al., 1996). They surmised that if this were true, then a manipulation designed to reduce this desire for interaction might reduce STI generation. In their attempt to produce such a reduction, they exposed some participants to photos depicting infectious diseases, which in theory dampens people's desire for social interactions. The manipulation worked: It reduced the extremity of judgments on a trait rating task (Study 1), and on a false trait recognition task (Studies 2 through 5). Additional measures and analyses showed that this reduction was not caused by participants'

feelings of negative affect, and did not generalize to exposure to other threatening stimuli (e.g., weapons).

Results That Help to Validate the Savings Measure as an Indicator of Trait Inferences

There are always reasons to be skeptical of the meaning of an outcome that emerges from research (one of the chapter's authors [JJS] frequently told his students to "beware of drinking the Kool-Aid" with regard to interpretations of research outcomes). For example, it is always possible that a given outcome can occur as a result of a theoretically problematic quirk of research methodology. One useful response to this possibility is for researchers to alter their paradigm, or to use alternative paradigms, to minimize the quirks of an individual paradigm. Often, these paradigm alterations or substitutions also help add validity to the original research results by ruling out alternative theoretical explanations that can be generated for an outcome.

Adding Validity Evidence by Ruling Out Alternative Explanations for Savings Effects

The Don and John–led research program aggressively pursued several forms of evidence that the savings measure is a valid indicator of trait inference-making at Time 1. For example, the original savings paradigm exhibited a photo familiarity confound: Photos used on the relearning trials at Time 2 had been previously seen at Time 1; the photos on control trials had not. This confound did not account for the savings effect: It remained robust, even when the confound was eliminated (Carlston & Skowronski, 1994).

A second idea attacking the meaning of the savings measure was the possibility that presentation of the photo at Time 2 prompted memory for the behavior that was paired at Time 1 with the photo, which facilitated trait-photo learning. This explanation suggests that savings should be stronger when behaviors are recognized. Instead, in a paradigm that assessed behavior recognition (Carlston & Skowronski, 1994, Experiment 3; also see Carlston et al., 1995, Experiment 5), results showed that savings effects were equivalent regardless of whether or not the behavior was correctly remembered or recognized.

A third non-trait inference explanation for saving effects was that exposure to the behaviors ***primed*** traits (e.g., increased their activation level), making them easier to learn at Time 2 than non-primed traits. This implies that savings should occur, even when the to-be-learned trait did not match the trait implied by the behavior with which a photo was paired at Time 1. This did not occur (Carlston & Skowronski, 1994, Experiment 5). Enhanced savings effects emerged only when the trait implied by a behavior in a behavior-photo pairing at Time 1 conceptually matched the trait-photo pairing used at Time 2.

A fourth non-trait inference explanation for saving effects was the possibility that relearning could be facilitated by an evaluative match between the trait generated during behavior encoding and the trait later paired with an actor. Results from Experiment 2 in Carlston et al. (1995) explored this idea by again having participants write a trait word in response to the behavior at Time 1. In one analysis, the trait words that semantically matched the trait word used in the savings task were discarded. No savings effect emerged when the word generated for an actor at encoding only evaluatively matched (but did not semantically match). Hence, the results from this study discounted target evaluations as the source of the enhanced savings effects observed in the main relearning paradigm.

Validating Savings as a Measure of Trait Inference-Making via Convergence across Measures

One other way to validate a measure is to show that it behaves in the same way that other measures thought to measure the same construct behave. Such results emerged from studies in which participants explicitly provided trait ratings of the social targets encountered at Time 1. Unsurprisingly, results from such studies confirm that people rate social targets high on the trait implied by each behavior after having read trait-implicative behaviors (e.g., Skowronski et al., 1998, Experiment 2; Carlston & Skowronski, 2005).

Similar results emerged from Experiment 5 reported by Carlston et al. (1995). They replaced the Time 2 and Time 3 tasks in the relearning paradigm with a trait reporting task in which participants simply saw a photo and tried to report a trait that was prompted by the photo. The traits that were reported tended to match the trait implications of the behaviors with which the photos were paired at Time 1, thus promoting the validity of the savings measure as a measure of trait inference-making.

In a variant of these explicit trait report studies, McCarthy and Skowronski (2011b) explored the behavior predictions that were made by participants after exposure to the photo-behavior pairs. That is, they tested whether seeing a person behave in a trait-implicative manner would lead perceivers to predict that person to behave in other ways consistent with the implied trait. If so, then it seems reasonable to assume the behavior prediction is based on a trait that was inferred from the first behavior. As expected, these predictions were consistent with the implications of the traits implied by the behaviors. Additional validation evidence came from two more findings. First, these behavior predictions were not related to behavior recall, a result suggesting that participants were not generating predictions by generalizing from their explicitly recalled behavior memories. In addition, the extremity of the predictions was much greater when there was an exact trait match between the original behavior and the predicted behavior than when the match did not match in trait but did match in evaluative direction.

There were a couple of other noteworthy findings reported in the McCarthy and Skowronski (2011b) article. The first was that the behavior predictions were more extreme for those participants who were uninstructed than for those who were given an explicit impression formation goal. This varies from the usual pattern showing that the savings data from these two conditions are equivalent. One avenue for future research is to understand why the effect of these instructions on the data yielded by these two measures differs across the measures.

A second noteworthy finding is that this tendency to make behavior predictions that matched the trait implications of previously encountered actor behaviors was mostly eliminated by telling participants prior to encoding the initial behaviors that some informants might be lying. One thrust of future research could be to better understand why this instruction had such an effect: Did it suppress trait inferences or did it inhibit behavior predictions made from trait inferences (or both)? One way to answer this question is to see whether such lie detection instructions suppresses evidence of trait inference-making as derived from the savings measure. Regrettably, to our knowledge no one has published results from such a study.

A third noteworthy finding from McCarthy and Skowronski (2011b) explored the extent to which the behavior predictions reflected automatic processes and the extent to which they reflected controlled processes. This was accomplished by explicitly asking participants on some trials to use the trait implications of the prior behavior in their predictions, and by asking on other trials to ***explicitly avoid using*** the trait implication of the prior behaviors. The data from these two trial types were entered into analyses specified by Jacoby's (1991) process dissociation model. The results indicated that even when trying to avoid doing so, participants often made behavior predictions that matched the trait implications of an actor's prior behavior. Thus, unsurprisingly, the process dissociation analyses indicated that spontaneously inferred trait information influenced participants responding outside of their control.

More validation of the savings measure as an index of trait inference-making comes from results of other research program studies that used another indirect measure, the false recognition measure, to detect inference-making. As noted earlier in this chapter, in this paradigm (e.g., Olcaysoy Okten & Moskowitz, 2020; Todorov & Uleman, 2002), people see trait-implicative sentences paired with actor photos. Sometimes a trait term that matches the trait implications of the behavior is included in the sentence (e.g., *he was generous and left a 40% tip*) and sometimes the trait is merely implied by the behavior (e.g., *he left a 40% tip*). Later, participants are shown a photo and are asked to report whether the trait term (e.g., *generous*) was originally in the sentence that accompanied the photo. False recognition errors—erroneously responding that a trait actually appeared in a previously-shown behavior that only implied the trait—are taken as evidence of trait inference-making, an outcome that has been repeatedly observed (e.g.,

Shimizu et al., 2017). The beauty of the false recognition paradigm is that STIs formed during encoding lead to more false recognitions, whereas correct recall for the wording of the behaviors would lead to fewer false recognitions. Thus, this paradigm provides strong evidence that trait inferences are formed during encoding.

One such set of observations was reported by McCarthy and Skowronski (2011b). Consistent with results from the savings measure, they found that the evidence for trait inference-making was equivalently produced by uninstructed participants and those explicitly instructed to make trait inferences. However, as in the behavior prediction studies, evidence for inference-making could be reduced by altering participants' goals during encoding. The evidence for trait inference-making was stronger in the uninstructed condition and in the explicit impression condition relative to a condition in which participants were instructed to search for specific letter strings in the sentences (e.g., the letter combination *ch*). Because to our knowledge it has never been tried, it is unclear whether a similar effect would emerge in studies using the savings measure. Hence, such studies reflect another direction for future research.

As in the behavior prediction studies, McCarthy and Skowronski (2011b) also used the Jacoby (1991) process dissociation methods to explore the extent to which the false recognition measure was influenced by automatic processing and the extent to which it was influenced by controlled processing. There was strong evidence of automatic processing in those conditions (uninstructed, intentional inference) in which people were expected to make trait inferences, but not in the letter sequence search (grapheme) condition (suggesting that focusing on letter strings interfered with trait inference-making). Again, this shift in the estimated contribution of automatic processes to responding is inconsistent with the thesis that trait generation is fully automatic.

Results from additional studies show that the contribution of automatic processes and controlled processes to STIs might vary across both stimulus type and perceiver. For example, a study reported by McCarthy et al. (2013) explored these automatic processes and controlled processes as they were exhibited in parents who were measured to vary in their risk of child abuse. One idea behind the research was that parents who were at low risk of child abuse might be less likely than high-risk parents to automatically make negative trait inferences about misbehaving children, especially when the child behaviors were ambiguous.

The study that was conducted to explore this idea used the false recognition method of inference detection. That is, parents in the study completed a false-recognition task. In the false recognition paradigm used in the McCarthy et al. study, parent participants first viewed behavior descriptions paired with child photographs. The behavior descriptions either vaguely or strongly implied a trait, and the traits implied were sometimes positive and sometimes negative. Descriptions sometimes included the trait

term implied by the behavior and sometimes not. Participants later completed a task in which they saw a photo and were asked if a trait term appeared in the behavior that was paired with the photo. STI generation is inferred from the false recognition rate observed in these judgments: The more often a trait is falsely recognized as having been present at Time 1, the greater the evidence for STI generation at Time 1.

Results from the study showed that low CPA risk parents were significantly less likely to indicate negative traits were present in behavioral descriptions of children when negative traits were vaguely (compared to strongly) implied. In contrast, high CPA risk parents were equally likely to indicate negative traits were present regardless of whether the traits were vaguely or strongly implied. More relevant to the notion that STIs may be partially automatic is that process dissociation analyses can be applied to the false recognition data to tease apart the contributions of automatic processes and controlled processes to the response patterns. Results from these analyses showed that for parents who were at low risk of engaging in child abuse, automatic processes contributed significantly less to task performance when negative traits were vaguely implied compared to when the same traits were strongly implied. This pattern did not emerge for parents who were at high risk of engaging in physical child abuse.

For now, the key take-away point that we want to emphasize from the study is that people differ in the extent to which automatic processes and controlled processes contribute to trait inference-making, and they do so in combination with the characteristics of the stimuli that are encountered. However, we also want to argue that findings such as these have potentially important practical implications. Those shall be discussed when we return to this study at the end of this chapter.

Is There Anything Unique about the Savings Measure? A Note on Discriminant Validity

Given the typical convergence among measures in this research program, one might wonder whether there is any need for the savings measure (or any other indirect measures of inference, such as the false recognition measure). One easy answer to the question lies in the reminder that one advantage of the indirect measures, such as savings, is that they can detect evidence of inference making even when participants are never asked to make inferences about actors. This is not the case in explicit trait reports, because the act of asking for an inference can cause an inference to be generated in response to the probe. This problem is colloquially referred to by this chapter's authors as the refrigerator light problem: If one wants to know if the refrigerator light stays on when the door is closed, it seems obvious to open the door and check. But opening the door initiates a process that causes the lights to go on. Similarly, with STI research, we often want to detect trait inferences [whether the light stays on] without actually asking people to report on their

trait inferences [without opening the refrigerator door]). Hence, it is not clear that such explicit trait reports reflect ***spontaneous*** inference-making in response to behavior observation.

Moreover, despite the fact that there might be considerable convergence in results provided by implicit measures and explicit measures, this need not always be the case. Indeed, there is evidence that measures such as the savings measure may reflect implicit elements of inference making that are not captured via explicit trait reports. For example, results reported by Carlston et al. (1995) directly linked trait generation performance to the savings task data. In one condition of these experiments, participants were explicitly asked to generate an actor trait in response to a photo two days after seeing the behavior-photo pairing. They also completed the usual Time 2 and Time 3 tasks of the relearning paradigm. Again suggesting the validity of the savings measure as an index of trait generation, analysis of the results from this forced/recorded inference condition showed that recall was best on relearning trials when the trait word generated initially matched the trait word presented in the relearning task. However, results from the study also revealed a smaller, but significant, savings effect on relearning trials when the trait explicitly generated did not match the trait used on the relearning task. In conjunction with the low trait photo-cued generation rate observed in Experiment 5 of Carlston et al. (1995), the authors saw these data as evidence for ***implicit trait knowledge***: Perceivers could have extracted and stored trait information about an actor that might influence their subsequent responding, even though that trait information might not be immediately accessible or consciously reportable. Thus, the savings measure (and, by extension, other indirect measures of trait inference-making) may assess elements of, or after-effects of, trait inference generation that go beyond the elements assessed by direct explicit measures.

Enter Spontaneous Trait Transference (STT)

One other attempt to validate the savings measure as an index of trait inference making in the trait relearning paradigm, one that turned out to be quite influential in the direction of the research program, was first reported by Carlston et al. (1995, Experiment 3). They used a variant of the relearning paradigm in which some participants believed from the first-person wording of the behaviors that the actors in the photos were describing their own behaviors (***self-informant condition***). Other participants believed from the third-person wording of the behaviors that the people in the photos were informants who were describing the behavior of others (***third-party informant condition***). The idea behind the experiment was to validate the savings measure as an index of trait inference-making by showing that savings emerged only when the behavior was described by the actor, and not by a third-party informant.

However, the experiment's results showed that a savings effect emerged even when the informant described a third-party's behavior. Although not

labeled as such at the time, this finding is now known by the term ***spontaneous trait transference (STT)***. We note that in this initial study the magnitude of the savings effect was somewhat larger when people in the photos were said to describe themselves than when they were said to describe a third-party, but the difference in the savings effect across the conditions was not statistically different. Thus, this initial result posed a significant challenge to the validity of the savings measure as an index of trait inference making about an actor.

However, as is now known from the results of many other studies, the initial Carlston et al. (1995) result somewhat overstated the magnitude of this challenge: In most subsequent studies, the magnitude of the savings effects observed is ***significantly*** larger when perceivers believe that informants described themselves than when they believe that informants described third parties. Hence, when using the savings paradigm, the usual outcome is that in conditions in which an informant is describing their own behavior savings effects are greater than when the informant is describing the behavior of a third party (STI > STT).

Nonetheless, regardless of this nuance, it is clear from the results of the entire corpus of research that used this third-party description condition (our shorthand for this going forward will be "the STT condition") that informants who describe third parties often become linked to the traits implied by the behaviors that the informants describe. Moreover, additional research results show that the trait content of informant third-party descriptions has implications for how the third-party informants are ***judged***. That is, in some studies, after encountering third-party informants who described the trait-implicative behavior of actors, participants judged the informants to be more extreme on the traits implied by the behaviors than when they judged targets who had not provided such descriptions (e.g., see Skowronski et al., 1998, Experiment 3). Thus, if participants read Jim's description of Donald acting dishonestly, Jim is rated as slightly more dishonest than if he had not provided such a description.

These findings potentially have far-reaching implications. For example, consider this finding in the context of those people who have the job of describing the negative behaviors of others (newscasters, lawyers). Given the STT findings, who would want a job in which one is in danger of being perceived negatively simply because one must describe the negative behavior of others? However, it is easy to generate examples of communicators for whom these effects did not seemingly occur. For example, surveys at the time suggested that Walter Cronkite was a highly trusted source, despite all the negative acts he had to describe in his newscaster job. Perhaps factors such as the circumstances in which people describe the behavior of others may mitigate against STT effects.

Indeed, research results suggest that there are a number of conditions that reduce or eliminate the emergence of STT effects. For example, Experiment 4 in Carlston et al. (1995) found that savings effects were equivalently strong

in response to descriptions of actors who described the behavior of others in uninstructed conditions and in response to descriptions provided by third-party informants in conditions in which the informants were said to be similar to the actors who performed the behaviors. However, the savings effects observed were reduced when participants were either given the goal of forming impressions of the third-party informant, or were given the explicit goal of forming impressions of the actor described by the informant. One way to understand the first of these reductions is via the idea that traits explicitly formed about informants from their act of description (e.g., "tattle-tale") might differ from the trait implied by the described behavior, and hence, might interfere with linking traits implied by the behavior to third-party informants. One way to understand the second of these reductions is that enhancing the attention given to understanding the enactor of the behavior might interfere with the linking of the traits implied by the behavior to third-party informants. Importantly, then, these latter findings suggest that perceiver formation of linkages between third-party informants and traits implied by their descriptions is ***not inevitable.***

This latter point is emphasized by additional results reported by Matt Crawford and his team. Results from some of these studies (Crawford et al., 2007, Study 1; Crawford et al., 2008, Study 1) found that STT effects dissipated when the third-party (depicted in a photo) informant's description of an actor was accompanied by an actor photo. This dissipation occurred both on savings measures (Crawford et al., 2007, Study 1, Crawford et al., 2008, Study 2) and on trait judgment measures (Crawford et al., 2007, Study 2; Crawford et al., 2008, Study 2). Both Crawford teams suggested that this elimination of STT occurred because the presence of the actor photo prompted participants to make inferences about the actors, and that this process inhibited the mental work needed for the formation of associations between third-party informants and the traits implied by their behavior descriptions. Simple attention to the actor photo was not enough to explain the elimination of the STT effect: Eye tracking data obtained during one of the studies (Crawford et al., 2008, Study 2) suggested that the elimination of the STT effect by the inclusion of an actor photo was not simply a function of the focus of participant attention during the encoding task.

STI vs. STT: Implications for Theory, Measurement, and Research

Theoretical Musings

These kinds of STT findings prompted extensive thought about the mental processes that were in play in the STT and STI conditions and the mental consequences of those processes. This was the case because the STT findings muddled the meaning of the results derived from the indirect paradigms being used in the STI research. For example, in the relearning paradigm,

savings might be produced from a process in which observers made inferences about behaviors, and those inferences became attached in memory to informants regardless of whether informants described themselves or a third party. However, the STT finding could also attest to the power of observers' tendency to make trait inferences. That is, an observer might tend to make an inference about an informant who described a third party because the observer routinely assumed that the informant was in some way similar to the described target (i.e., had similar traits). Attribution theory suggests that attention to circumstances might be why perceivers could avoid ascribing traits to third-party describers such as Walter Cronkite. Perceiver awareness of stereotypes or roles could work against the assumption that informants are similar to those they describe, so the STT effect can be avoided for such communicators.

These kinds of ideas contributed to theorizing that first appeared in Skowronski et al. (1998). These authors argued that the processes that typically occurred in the STI case and in the STT case differed (not all accept this assumption; see Orghian et al., 2015). In the STI case, the authors argued that inferences ***about the actor*** were made at encoding and produced a mental representation of the actor that included information that the trait was a property of the actor. In contrast, the authors argued that in the STT case the trait implied by the behavior was ***activated*** when encoding the behavior, but ***at that time an inference was not made about the informant***. Instead, in the STT case the mental representation of the informant included only a ***non-inferential association between the trait and other informant knowledge***. This trait-person mental association is detectable with paradigms such as the savings-in-relearning paradigm and might affect later thoughts about the informant (including future inferences), but was not initially stored in memory in the form of an inference.

An example is illustrative. We begin with what we believe to be occurring in the STI case. Assume that a perceiver listens to Jim describe how he helped a colleague with a research problem. The perceiver might immediately conclude that "Jim is kind," and store that knowledge in memory. Things differ in the STT case. When a perceiver listens to Jim describe how Bettina helped a colleague with a research problem, the perceiver might think "that was a kind thing to do." In both of these cases, the trait term "kind" is associated to the mental representation of "Jim." If measured, this mental association would exert a measurable influence on subsequent responding (e.g., savings-in-relearning, explicit trait inferences, etc.).

Implications for the Measurement of Trait Inference-Making

This theorizing has implications for the meaning of the savings measure and its validity as an index of trait inference-making. Our current view is that the savings measure can probably best be thought of as a measure of the extent to which a person and a trait are ***associated***. Trait inference-making might be

one source of such associations (so the measure has some degree of validity), but there may be others. For example, the trait of "dishonesty" activated at the same time an actor arrives on the scene of a robbery might become associated with the actor, ***even though the trait is not about the actor*** (as it is in trait inferences). Hence, though the savings measure can reflect trait inference-making at the time of behavior encoding, it is likely not a "pure" measure of such inference-making.

However, this is not a fatal flaw in the savings measure. We note that the same kind of conceptual impurity probably applies to other indirect measures of trait inference-making (e.g., false memory; response latencies), and even those that are supposedly more direct, such as trait judgments. For example, in this latter case, though trait judgments may directly reflect trait inferences made about an informant at encoding (e.g., as can happen in STI conditions), in the Skowronski et al. (1998) view, they can also reflect the presence of informant-trait associations that are later used to make trait inferences sometime after initial encoding has occurred (e.g., when the observer is asked to make a trait judgment).

In our view, the best approach in the face of this kind of conceptual impurity is to use across studies different (and complementary) dependent measures that are all thought to reflect the influence of the construct of interest. By examining results across studies, one can hope to see evidence of trait inference-making across all the measures. As we have noted previously, this is one important element of the research program described in this chapter: In addition to the savings measure, the dependent measures used in the research program included trait judgments, behavior predictions, and other indirect measures (e.g., false memory).

Pitting STI vs. STT

The theorizing of Skowronski et al. (1998) prompted a plethora of studies designed to compare STIs and STT. One important question in these studies was to determine whether STT effects and STI effects responded in the same way, or in different ways, in different circumstances and on different measures. The emergence of such differences would support the idea that different cognitive processes were involved in STI and STT and they produced different mental consequences.

One such early attempt appeared in Skowronski et al. (1998, Study 1). Results from the study showed that savings effects emerged for both self-informants (STI condition) and third-party informants (STT condition) but were stronger in the STI case than the STT case (note that this same ordering of means appeared in Carlston et al. (1995), but was not statistically reliable there). Additional data reported by Skowronski et al. (1998, Studies 2 & 4) also showed that trait judgments made about informants followed this same pattern: Significant in both cases, but stronger in the STI condition than in the STT condition. Results from other studies (Carlston & Skowronski, 2005)

showed that this STI > STT pattern in judgments of traits implied by the behaviors was unaffected by whether participants had 10s or 20s to read behaviors at Time 1 (Study 1), and was not caused in STT conditions by misidentification of third-party informants as self-informants (Studies 2 and 3). Hence, across studies, one reliable difference between results obtained from STI conditions and STT conditions is that the results from STI conditions, on both savings measures and judgment measures, tend to be larger than results from STT conditions.

One other consistent STI vs. STT difference also emerges on trait judgment measures. In Carlston and Skowronski (2005, Experiments 1, 2, & 3), in the STI condition but not the STT condition an evaluative-congruency effect occurred for judgments of traits that were not directly implied by the behavior (increased ratings on traits that evaluatively matched the trait implied by the behavior; decreased ratings on traits that did not evaluatively match the behavior). This finding also consistently emerged in the results reported by Wells et al. (2011; also see results for the STT condition in Studies 2 and 4, Skowronski et al., 1998). Hence, across studies, the occurrence of evaluatively-congruent halo effects in judgments of traits not directly implied by behaviors reliably distinguishes between the trait judgments derived from inferential processes (e.g., STI conditions) and those derived from associative processes (e.g., STT condition).

Though convergence of results has been typical in this line of research, we note one point of inconsistency in the STI > STT results reported across studies. One claim made by Carlston and Skowronski (2005) is that there are stronger effects in STI conditions than STT conditions in negative trait judgments than on positive trait judgments. However, this finding has not consistently emerged across studies. Hence, in our opinion this finding should be considered as tentative and should be a topic of future research.

Other STT vs. STI studies yielded evidence that they sometimes respond similarly to different manipulations or behave similarly across circumstances. For example, in four studies, Wells et al. (2011) examined the extent to which STT effects and STI effects were both dependent on an individual's cognitive capacity at behavior encoding and whether this dependency differed between STT conditions and STI conditions. The results from all four studies reported by Wells et al. suggested that manipulations (e.g., requesting performance of additional tasks) that reduced available cognitive capacity during behavior encoding reduced both STI and STT effects. These reductions emerged both on the savings measure (Studies 1 and 3) and on explicit trait rating measures (Studies 2 and 4). Similar results indicating that both STT and STI were reliant on cognitive capacity came from results of one study (Study 4) that used an individual difference measure of cognitive capacity. Thus, even though STIs and STTs are believed to be caused by different cognitive processes, there seems to be similarities in the conditions under which those processes operate (e.g., they both require some cognitive processing resources during encoding of the behaviors).

The comparison of STI to STT also characterized studies reported by McCarthy et al. (2018). In a repeated behavior-presentation paradigm, they conducted studies that varied both the type and amount of behavioral information that perceivers encountered in either STI or STT conditions. Using a modified savings-in-relearning paradigm, results from Experiments 1a and 1b demonstrated that repeated presentations of an individual and a behavior description increased the strength of association between the target and implied trait, and this effect did not depend on whether the repeated presentations involved redundant information or new information. In comparison, Experiments 2a and 2b used a trait ratings dependent variable and demonstrated that the effects of behavior repetition were stronger for STI, but not STT. However, this effect emerged only when behaviors added new information to the previously encountered information; the difference did not emerge when the new information was redundant with the previously encountered information. This result suggests that trait rating measures might sometimes discriminate between the processes thought to underlie STI and STT, even when the savings measure is unable to do so.

Equivalence between STT and STI conditions emerged from results reported by Zengel et al. (2017). In the study they described, informants read behaviors ostensibly reported by either previously known informants (1) who were positive (e.g., Abraham Lincoln), (2) neutral (e.g., Jay Leno), or (3) negative (e.g., Adolf Hitler), or by previously unknown informants. As in past studies, the behaviors described were either trait-implicative positive behaviors, trait-implicative negative behaviors, or neutral behaviors. As in past STT vs. STI studies, these descriptions were framed as either the behavior of the informant or the behavior of another person as described by the informant. Results from a savings measure yielded the usual pattern: For trait implicative behaviors there were significant savings effects in both STT conditions (conceptually replicating results reported by Mae et al., 1999) and STI conditions, but the effect differed in magnitude (STI > STT). However, the emergence of these STT effects and STI effects was unaffected by whether the informant was known or unknown, or the prior evaluation of the informant. Hence, prior knowledge about the informants did not seem to alter the processes that were involved in either STT or STI, at least when these were assessed via a savings measure.

Zengel et al. (2017) never assessed the impact of informant familiarity on other measures (trait judgments, behavior predictions). It seems possible to us that the impact of prior informant knowledge on these measures might dissociate, showing different patterns in the STI condition and the STT condition. This represents one interesting direction for future research.

Beyond Trait Inferences

As we noted earlier in this chapter, early STI scholarship (especially Jim Uleman's) had a huge influence on the exploration of spontaneous thoughts

about actors. One ironic consequence of that influence may be that it restricted the study of spontaneous thoughts to the study of trait inferences—because that's where Jim started. The study of other kinds of spontaneous thoughts has emerged only gradually. These have come to include the study of spontaneous thoughts about actor emotional states, actor goals, and how situations influenced actor behaviors.

Just as with the rest of the field, it took a while for our own research program to begin to study some of these other kinds of spontaneous thoughts and their effects. One example comes from studies originally concerned with assessing the validity of the savings measure as an index of trait generation.

In one experiment, Schneid et al. (2015b) used a standard savings-in-relearning paradigm to explore whether exposure to trait-implicative behavior descriptions facilitated the learning of evaluatively congruent, as well as behavior-implied, personality traits. Evidence for the facilitated learning of evaluatively congruent traits was not obtained. This result enhanced the validity of the trait-based savings measure as an index of trait generation about an actor.

However, given that other research and theory has suggested that evaluations of stimuli are ubiquitous, Schneid et al. (2015b) wondered whether their savings task could be modified to capture spontaneous evaluations of the actors. This led to a second experiment in which the savings-in-relearning paradigm was altered to directly assess spontaneous evaluative inference (SEI) generation via savings in relearning of evaluative words (good/bad). The results found exactly such an effect. These results not only revealed a new method for exploring evaluations made about actors, they also provided additional validity for the trait-based savings measure by showing that evaluations did not seem to influence responding on the trait-based savings task, even though these spontaneously-produced actor evaluations could be detected and these evaluations were produced at the same time as the trait inferences.

Schneid et al. (2015b) tried to produce additional measure validity evidence by looking in a third study for similar results from evaluation-based and trait-based versions of the false recognition task. They found such evidence, but with one qualification. Their results showed that, in contrast to the data indicating that the trait-based savings measure was relatively immune to evaluative congruency effects, false recognition of trait terms might be influenced by evaluative congruency. This suggests that, though both measures have been useful in the study of trait inferences, the savings measure might be preferable to the false recognition measure because the trait-based savings measure is relatively immune to the effects of spontaneous evaluations, whereas the false recognition measure is not.

One additional set of studies (Schneid et al., 2015a) also explored the spontaneous evaluations that perceivers might generate about actors. To search for SEIs, Experiments 1 and 2 in this article again used modified versions of the relearning paradigm in which trait terms were replaced with

evaluative terms (good/bad) in the relearning task. Savings measure results yielded evidence of the production of evaluative inferences, and also indicated that such inferences emerged equally regardless of whether participants were instructed to form trait impressions, evaluative impressions, or neither. Results from the experiments also showed that SEIs were not dependent on trait inferences: Evidence for SEIs occurred regardless of whether there was explicit recall for the trait implications of the stimuli.

Experiment 3 in Schneid et al. (2015a) pursued this latter distinction between STIs and SEIs. Results from the behavior-prediction task used in that study showed that new behaviors that implied the same trait as previously-encountered behaviors were seen as less likely to be performed by the actor in the future when participants had been asked to detect lies in the descriptions (potentially inhibiting trait inferences) than in uninstructed (e.g., STI) conditions. The same pattern generally occurred for predictions made about behaviors that did not match the original behavior's trait implication, but that matched the behavior's valence. In contrast, this pattern did not emerge on a measure that assessed whether an observer would approach or avoid an actor. That is, the approach/avoid judgments were unaffected by instructions, and instead were determined totally by the evaluative implications of the actor's original behavior. This dissociation between measures was explained by assuming that the behavior predictions were largely linked to the semantic (trait) implications of the behaviors, while the approach/avoid judgments were determined by the evaluative implications of the behaviors. This idea was linked to the claim of many theorists that the extraction of evaluative information from a stimulus might be especially difficult to disrupt. Hence, a manipulation that might disrupt inference-making (a lie detection goal) might not disrupt evaluative inference generation, and it is this difference in inference disruptability that was thought to be responsible for the dissociation in results between the two measures.

Why Should Readers Who Are Not Invested in Social Cognition Give a Hoot?

To some, the study of STIs seems awfully esoteric, especially given that evidence for STIs has been pursued largely by scholars who have a passion for understanding social cognition. However, it should not be overlooked that one reason to try to understand STI generation is that the inferences that people generate might significantly alter their behavior (as Susan Fiske once wrote [echoing both Gordon Allport and William James], "thinking is for doing").

To illustrate this point, we return to a study that we described earlier—that parents who vary in child abuse risk exhibit different STI tendencies as assessed in a false recognition paradigm (McCarthy et al., 2013). We remind our readers that results from the study showed that low

CPA risk parents were significantly less likely to indicate negative traits were present in behavioral descriptions of children when negative traits were vaguely (compared to strongly) implied in behaviors. In contrast, high CPA risk parents were equally likely to indicate negative traits were present regardless of whether the traits were vaguely or strongly implied.

One can extrapolate from this pattern and infer that high CPA risk parents might be especially likely to abuse their kids in that they often perceive their kids to be engaging in behaviors that reflect negative child traits. More importantly, because thinking can lead to doing, this pattern of inferences can lead to enhanced child abuse in high CPA risk parents. For example, when little Brett innocuously scrunches his nose in response to a parent's request, a high-risk parent might see that as a reflection of little Brett's rebellious nature and act punitively in response to the inference. Moreover, because these inferences occur automatically during encoding, the parent feels as if they are merely seeing the behavior for what it is, they do not feel as if they are effortfully interpreting their child's behavior. Hence, one reason to understand the inferences that perceivers draw from their observations of actor behavior, and when they are (or are not) drawn, is that such inferences can help to explain the subsequent behavior of the perceiver toward the actor.

Acknowledgement

We express our thanks to both Don Carlston and Gordon Moskowitz for reviewing an earlier draft of this chapter and for providing helpful comments that guided our revision.

References

Carlston, D. E., & Skowronski, J. J. (1994). Savings in the relearning of trait information as evidence for spontaneous inference generation. *Journal of Personality and Social Psychology*, 66(5), 840–856. doi: 10.1037/0022-3514.66.5.840

Carlston, D. E., & Skowronski, J. J. (2005). Linking versus thinking: Evidence for the different associative and attributional bases of spontaneous trait transference and spontaneous trait inference. *Journal of Personality and Social Psychology*, 89(6), 884–898. doi: 10.1037/0022-3514.89.6.884

Carlston, D. E., Skowronski, J. J., & Sparks, C. (1995). Savings in relearning: II. On the formation of behavior-based trait associations and inferences. *Journal of Personality and Social Psychology*, 69(3), 420–436. doi: 10.1037/0022-3514.69.3.429

Crawford, M. T., McCarthy, R. J., Kjaerstad, H. L., & Skowronski, J. J. (2013). Inferences are for doing: The impact of approach and avoidance states on the generation of spontaneous trait inferences. *Personality and Social Psychology Bulletin*, 39(3), 267–278. doi: 10.1177/0146167212473158

Crawford, M. T., Skowronski, J. J., & Stiff, C. (2007). Limiting the spread of spontaneous trait transference. *Journal of Experimental Social Psychology*, 43(3), 466–472. doi: 10.1016/j.jesp.2006.04.003

Crawford, M. T., Skowronski, J. J., Stiff, C., & Leonards, U. (2008). Seeing, but not thinking: Limiting the spread of spontaneous trait transference II. *Journal of Experimental Social Psychology*, *44*(3), 840–847. doi: 10.1016/j.jesp.2007.08.001

Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology*. New York: Dover. (Original work published 1885).

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition*. New York: McGraw-Hill.

Hamilton, D. L. (1981). Cognitive representations of persons. In E. T. Higgins, C. P. Herman & M. P. Zanna (Eds.), *Social cognition: The Ontario symposium* (Vol. 1, pp. 135–159). Hillsdale, NJ: Erlbaum.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513–541. doi: 10.1016/0749-596x(91)90025-F

Mae, L., Carlston, D. E., & Skowronski, J. J. (1999). Spontaneous trait transference to familiar communications: Is a little knowledge a dangerous thing? *Journal of Personality and Social Psychology*, *77*(2), 233–246. doi: 10.1037/0022-3514.77.2.233

McCarthy, R. J., Crouch, J. L., Skowronski, J. J., Milner, J. S., Hiraoka, R., Rutledge, E., & Jenkins, J. (2013). Child physical abuse risk moderates spontaneously inferred traits from ambiguous child behaviors. *Child Abuse & Neglect*, *37*(12), 1142–1151. doi: 10.1016/j.chiabu.2013.05.003

McCarthy, R. J., & Skowronski, J. J. (2011a). The interplay of controlled and automatic processing in the expression of spontaneously inferred traits: A PDP analysis. *Journal of Personality and Social Psychology*, *100*(2), 229–240. doi: 10.1037/a0021991

McCarthy, R. J., & Skowronski, J. J. (2011b). What will Phil do next? Spontaneously inferred traits influence predictions of behavior. *Journal of Experimental Social Psychology*, *47*(2), 321–332. doi: 10.1016/j.jesp.2010.10.015

McCarthy, R. J., & Skowronski, J. J. (2014). Disease avoidance cues interfere with spontaneous trait inferences. *Evolutionary Behavioral Sciences*, *8*(4), 289–302. doi: 10.1037/h0099105

McCarthy, R. J., Wells, B. M., Skowronski, J. J., & Carlston, D. E. (2018). Multiple behavior descriptions affect the acquisitions of STI and STT. *Psychological Reports*, *121*(4), 615–634. doi: 10.1177/0033294117736317

Nelson, T. O. (1985). Ebbinghaus's contribution to the measurement of retention: Savings during relearning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(3), 472–479. doi: 10.1037/0278-7393.11.3.472

Orghian, D., Garcia-Marques, L., Uleman, J. S., & Heinke, D. (2015). A connectionist model of spontaneous trait inference and spontaneous trait transference: Do they have the same underlying processes? *Social Cognition*, *33*(1), 20–66. doi: 10.1521/soco.2015.33.1.20

Olcaysoy Okten, I., & Moskowitz, G. B. (2020). Easy to make, hard to revise: Updating spontaneous trait inferences in the presence of trait-inconsistent information. *Social Cognition*, *38*(6), 571–625. doi: 10.1521/soco.2020.38.6.571

Schneid, E. D., Carlston, D. E., & Skowronski, J. J. (2015a). Spontaneous evaluative inferences and their relationship to spontaneous trait inferences. *Journal of Personality and Social Psychology*, *108*(5), 2015, 681–696. doi: 10.1037/a0039118

Schneid, E. D., Crawford, M. T., Skowronski, J. J., Irwin, L. M., & Carlston, D. E. (2015b). Thinking about other people: Spontaneous trait inferences and spontaneous evaluations. *Social Psychology*, *46*(1), 24–35. doi: 10.1027/1864-9335/a000218

Shimizu, Y., Lee, H., & Uleman, J. S. (2017). Culture as automatic processes for making meaning: Spontaneous trait inferences. *Journal of Experimental Social Psychology*, *69*, 279–285. doi: 10.1016/j.jesp.2016.08.003

Skowronski, J. J., Carlston, D. E., Mae, L., & Crawford, M. T. (1998). Spontaneous trait transference: Communicators take on the qualities they describe in others. *Journal of Personality and Social Psychology*, *74*(4), 837–848. doi: 10.1037/0022-3514.74.4.837

Smith, E. R. (1989). Procedural efficiency: General and specific components and effects on social judgment. *Journal of Experimental Social Psychology*, *25*, 500–523. doi: 10.1016/0022-1031(89)90003-6

Todorov, A., & Uleman, J. S. (2002). Spontaneous trait inferences are bound to actors: Evidence from false recognition. *Journal of Personality and Social Psychology*, *83*, 1051–1065. doi: 10.1037/0022-3514.83.5.1051

Uleman, J. S., & Moskowitz, G. B. (1994). Unintended effects of goals on unintended inferences. *Journal of Personality and Social Psychology*, 66(3), 490–501. doi: 10.1037/0022-3514.66.3.490

Uleman, J. S., Newman, L. S., & Moskowitz, G. B. (1996). People as flexible interpreters: Evidence and issues from spontaneous trait inference. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 28, pp. 211–279). San Diego, CA, US: Academic Press.

Uleman, J. S. (1989). A framework for thinking intentionally about unintended thoughts. In J. S., Uleman, & J. A., Bargh (Eds), *Unintended thought*, (pp. 425–449). New York, NY: The Guilford Press.

Wells, B. M., Skowronski, J. J., Crawford, M. T., Scherer, C. R., & Carlston, D. E. (2011). Inference making and linking both require thinking: Spontaneous trait inference and spontaneous trait transference both rely on working memory capacity. *Journal of Experimental Social Psychology*, *47*(6), 1116–1126. doi: 10.1016/j.jesp.2011.05.013

Winter, L., & Uleman, J. S. (1984). When are social judgments made? Evidence for the spontaneousness of trait inferences. *Journal of Personality and Social Psychology*, *47*(2), 237–252. (See correction in (1986) *Journal of Personality and Social Psychology, 50*(2) 355, doi:10.1037/h0090437).

Zengel, B., Ambler, J. K., McCarthy, R. J., & Skowronski, J. J. (2017). Spontaneous trait inference and spontaneous trait transference are both unaffected by prior evaluations of informants. *The Journal of Social Psychology*, *157*(3), 382–387. doi: 10.1080/00224545.2016.1192099

10 Impression Formation, Right Side Up

David E. Melnikoff[1] *and John A. Bargh*[2]
[1]*Northeastern University*
[2]*Yale University*

On October 18, 1961, New York's Museum of Modern Art unveiled a new exhibition celebrating the works of Henri Matisse, not realizing that one of the pieces, Matisse's *Le Bateau*, was hung upside down. For 47 days, the error escaped the notice of curators, museum staff, and some 115,000 visitors.

In science, there exists an unknown number of ideas that, like *Le Bateau*, are 180° from the truth, perfect inversions of reality hanging upside down in plain sight. Noticing these errors, and turning them right-side up, can be paradigm shifting. Take Darwin's theory of evolution by natural selection (Darwin, 1859). Before Darwin, the dominant theory of creation drew a causal path from intelligence, in the form of a wise and omniscient God, to the basic principles of biology. Darwin turned this view on its head by showing that biological principles give rise to intelligence, not the other way around. Robert MacKenzie (1868), one of Darwin's 19th-century critics, called this a "strange inversion of reasoning" (see Dennett, 2009).

Psychology has its own strange inversions. James (1884) inverted the causal path from emotion to behavior—fear, for instance, is not a cause of fleeing, but an effect of fleeing. Festinger (1957) inverted the path from attitudes to behavior; Schachter and Singer (1962), the causal path from emotion to physiological arousal; Zajonc (1980), the path from thinking to feeling; Haidt (2001), the path from moral reasoning to moral judgment. And Uleman (1999), by showing that humans make social inferences spontaneously, inverted our understanding of automaticity.

At first glance, the picture that emerges from Uleman's research appears similar to the original, but close inspection reveals that the canvas is flipped. The original view of automaticity depicts a bottom-up process whereby low-level sensory inputs (e.g., visual cues) directly activate high-level concepts (e.g., character traits). Yet social inference proceeds in the opposite direction. High-level concepts are used to issue predictions about low-level sensory input, which are compared to the actual input to infer the state of the world. By discovering that social inference proceeds automatically, Uleman's work (e.g., Uleman, 1999; Uleman et al., 1996; Winter & Uleman, 1984) suggests that automatic processes need not be simple, bottom-up mappings from stimulus to response, but can be sophisticated, top-down processes

DOI: 10.4324/9781003045687-12

typically associated with intentionality, control, effort, and consciousness (e.g., Bargh, 1989, 1994; Keren & Schul, 2009; Melnikoff & Bargh, 2018). This is Uleman's strange inversion.

The Inference Process

The topsy-turvy implications of spontaneous trait inference become clear when we consider the computations underlying inference in humans. The emerging consensus is that these computations approximate Bayesian principles (Chater et al., 2006; Clark, 2013; Dayan et al., 1995; Griffiths et al., 2007; Knill & Pouget, 2004; Lee & Mumford, 2003). What does this mean? Three things. First, it means that humans construct generative mental models of the world (Clark, 2013; Friston, 2010; Rao & Ballard, 1999). Such models map causal paths from hidden states of the world (i.e., states we cannot observe directly) to our observations of the world (i.e., percepts). For instance, you may have a generative model that says friendliness causes smiling. This model is generative in the sense that it allows you to *generate* predictions of what you would observe (e.g., the presence or absence of smiling) under different assumptions about hidden states of the world (e.g., whether or not someone is friendly).

The second implication of modeling human inference as Bayesian inference is that the generative models we construct are probabilistic. That is, for any observation d and hidden state h, we represent the probability of d if h were the true state of the world, $P(d \mid h)$, as well as the probability that h is the true state of the world, $P(h)$. In the language of Bayesian inference, $P(h)$ is a "prior" and $P(d \mid h)$ a "likelihood." A generative model that says "friendliness causes smiling" would include the prior probability of friendliness, $P(\textit{friendly})$, and the likelihood of smiling among people who are friendly, $P(\textit{smile} \mid \textit{friendly})$, versus unfriendly, $P(\textit{smile} \mid \textit{unfriendly})$. Mentally representing likelihoods lets us think things like, "Friendly people usually smile" and "Unfriendly people rarely smile." Mentally representing priors lets us think things like, "Most people are friendly" or "Few people are friendly."

The third element of the Bayesian framework involves using generative models to infer the probability that h is the true state of the world after observing d—a quantity called a "posterior," denoted as $P(h \mid d)$. For instance, if you observe someone smile, you might use your generative model ("friendliness causes smiling") to infer the probability that the person is friendly: $P(\textit{friendly} \mid \textit{smile})$. How does this work? Basically, it involves multiplying the likelihood and prior:

$$P(h \mid d) \propto P(d \mid h) P(h) \tag{10.1}$$

This equation says that the posterior, $P(h \mid d)$, is proportional to the product of the likelihood, $P(d \mid h)$, and the prior, $P(h)$. Consider what this means for

inferring friendliness from smiles. For one, it means that, all else being equal, the likelier you think a friendly person is to smile (i.e., the greater the likelihood $P(smile \mid friendly)$), the more certain you'd be that someone who smiled is friendly (i.e., the greater the posterior $P(friendly \mid smile)$). This makes intuitive sense. Holding all else constant, the more common smiling is among friendly people, the more indicative smiling is of friendliness. Another implication of Equation 10.1 is that posteriors are weighted by priors: the greater the prior probability of friendliness, the greater the posterior probability of friendliness. This aligns with the intuition that if friendliness is very rare, then, even if we observe someone smile, we should remain skeptical that the person is friendly. Conversely, if friendliness is extremely common, an absence of smiling should not lead us to infer an absence of friendliness. The logic of Bayesian inference seamlessly extends to more complex generative models, including those with continuous variables and hierarchical structure (e.g., Moskowitz & Olcaysoy Okten, 2016; Van Overwalle et al., 2012).

The contemporary picture of spontaneous trait inference depicts a Bayesian process, mirroring decades of theoretical and empirical work on intentional forms of social perception (e.g., Ajzen & Fishbein, 1975; Anderson, 1974; Atzil et al., 2018; Darley & Fazio, 1980; Ginossar & Trope, 1987; Tamir & Thornton, 2018; Trope, 1986). The "strange inversion" of this picture becomes apparent when Bayesian inference is compared to traditional accounts of automatic processing. Traditional accounts of automaticity replace the concept of generative models with the concept of stimulus-response mappings. Such mappings allow observations to activate mental representations of hidden states in a purely bottom-up manner. For instance, a stimulus-response mapping from "smile" to "friendly" would allow observations of smiles to activate the mental representation of friendliness. It has been argued that the simplicity of stimulus-response mappings makes them uniquely suited to automatic processing.

Stimulus-response mappings are mirror images of generative models. Whereas generative models draw causal paths from hidden states to observations, stimulus-response mappings draw causal paths from observations to hidden states. This is the essence of Uleman's strange inversion. Winter and Uleman (1984) took what we thought was the basic mechanism of automaticity—the stimulus-response mapping—and replaced it with its inverse, the generative model.

It is easy to see that this about-face is a move in the right direction. The literature is replete with findings suggesting that generative models, rather than stimulus-response mappings, underlie spontaneous trait inference (Carlston & Skowronski, 2005; Crawford et al., 2007; Crawford et al., 2008; Goren & Todorov, 2009; Kressel & Uleman, 2010; Mae et al., 2004; Skowronski et al., 1998; Todorov & Uleman, 2004; Wells et al., 2011). Consider the findings of Kressel and Uleman (2010). These researchers explored how the order in which trait-words and action-words are presented

affects the speed with which people recognize these words as causally related. They had participants rapidly judge causal relations between 128 word pairs, including 32 trait-action pairs. Some of the trait-action pairs began with a trait (e.g., silly-giggle), and the rest began with an action (e.g., giggle-silly). If people have generative models that map from hidden states (e.g., traits) to observations (e.g., actions), then judgment should be faster when the trait precedes the action, as this sequence would better match how these concepts are mentally organized. By the same logic, judgment should be faster when the action precedes the trait if people map directly from stimuli to responses. The findings of Kressel and Uleman (2010) clearly indicate that the mental representations linking traits to actions take the form of generative models rather than stimulus-response mappings: Participants were faster to identify trait-action word pairs as causally related when the trait came first.

Further evidence that generative models implement spontaneous trait inference comes from research showing that spontaneous trait inferences are sensitive to mental representations of base rates. Recall that generative models, but not stimulus-response mappings, contain information about the prior probability, or base rates, of hidden states. It follows that sensitivity to base rates is diagnostic of generative models. With this in mind, consider the results of Wigboldus et al. (2003). These researchers found that spontaneous trait inferences are weaker for counterstereotypic actions relative to stereotype-consistent actions. For instance, participants who learned that a garbageman won a science quiz spontaneously inferred the trait *smart* less often than participants who learned that a professor did the same. This result follows naturally from the logic of Bayesian inference. Most people believe—wrongly, perhaps—that the trait *smart* is less common in garbagemen then professors. This belief would lead a Bayesian reasoner to the following conclusion: A science-quiz-winning garbageman is less likely to be smart than a science-quiz-winning professor. Indeed, Bayesian reasoners believe that, all else being equal, the greater the prior probability of a hidden state, the greater the posterior probability of a hidden state (see Equation 10.1). This logic is difficult to articulate in the language of stimulus-response mappings, which do not explicitly account for the prior probability of hidden states.

A New Look at Automaticity

The transformative implications of spontaneous trait inference become clear when placed in historical context. When Winter and Uleman (1984) first demonstrated spontaneous trait inference, there was a general movement in social psychology away from motivational explanations for basic phenomena, and towards more cognitive explanations. In the 1970s for example, there were dozens of experiments attempting to show that cognitive dissonance effects were entirely cognitive and not motivated at all (e.g., Bem, 1972). Stereotyping and prejudice were shown to not require motivations and biases

to occur as they could be produced by purely natural attention and memory processes: Taylor and Fiske (1978) and McArthur (1980) emphasized visual salience that drove greater attention to statistically infrequent types of people, resulting in more available memories of their behavior, and thus overestimation of their causal role in group outcomes. Hamilton and Gifford (1976) showed that the "illusory correlation" between minorities and negative social behavior is caused by the "double whammy" of greater attention paid to both minority social groups and (relatively infrequent) negative social behavior, causing an overestimation in later recall of those behaviors. Finally, Nisbett and Ross (1980) put the capstone on this trend by making a strong case that many of the biases and errors in social judgment were attributable to the limits and constraints on normal human cognitive functioning, not to deliberate, motivated reasoning.

In that *Zeitgeist* of purely cognitive, nonmotivational accounts of classic social psychological phenomena, Winter and Uleman's (1984) spontaneous trait inferences fit right in. They showed that participants naturally understood the behaviors of a target person in personality trait terms, even when they had no goal or motive to form an impression of the actor, because their assigned experimental task had nothing to do with impression formation. Impressions of others (or at least, others' behaviors) were formed automatically and in the absence of any goal to do so. Winter and Uleman (1984) developed a clever paradigm of cued recall to test whether social-behavior sentences were more likely to be later recalled if a trait word cue was presented that was related to but not contained within that behavior description. This latter point was key, because the growing body of "implicit memory" effects of that era (see Bargh & Hassin, 2021) all involved the use of the actual words that had been shown in an earlier experiment—as free associates, for example—even though participants could not explicitly recall those words as having been shown. To show participants had gone beyond the actual stimulus event, Winter and Uleman (1984) had to show the benefits of cue trait terms that had *not* been presented in the original behavior descriptions. This they did.

Because the cue word had never been presented in the acquisition phase of the study, its efficacy in improving recall of the behavior had to be because it had been generated and stored automatically, unconsciously, in the episodic memory trace of that behavior description stimulus at the time the participant read that behavior description. This encoding effect has been replicated many times and the summarizing of complex social behavior in simple trait terms is now considered a basic mechanism supporting impression formation and social judgment. It supports the common social communication experience of easily describing someone you recently met to others as kind, generous, sneaky, unpleasant, distant, etc. often without much memory for the particular behaviors that generated those summary judgments. Trait encodings are the shorthand parlance we use to describe each other—it is relatively simple, efficient, and gives the ability to predict behavior in new

future situations never previously encountered. The same principle has recently been applied in understanding the utility of "emotion words" as compared to emotional expressions—within a given emotion type, the latter are highly variable and unreliably classified, and so the former become valuable as stable shorthand summaries to encode into person memory (Doyle & Lindquist, 2018; Lindquist & Gendron, 2013).

At first glance, the spontaneous trait inference effect fit nicely into the "nonmotivational" *Zeitgeist* of early social cognition research, as another example of an automatic (unintended) mental reaction to the current social environment. Automatic effects did not require the participant's conscious intention to occur, such as an assigned experimental task goal to form impressions or make social judgments. Basic social psychological phenomena were discovered to have these automatic components: with the self-concept (Bargh, 1982), important attitudes (Fazio et al., 1986), and stereotypes of others (Devine, 1989) all shown to become active "automatically." That trait judgments were also directly and unintentionally generated by merely reading behavioral descriptions fit right in to this rolling tide. At least, that is how it was generally understood at the time. Upon closer examination, however, it didn't really fit at all.

To see why, we must go back to the early days of cognitive psychology. Perhaps influenced by Freud's "separate mind" hypothesis in which an unconscious mind first filters and censors information and experience before it enters conscious awareness (see Bargh & Hassin, 2021; Erdelyi, 1974), one hot topic of the 1960s concerned the extent to which information in the environment was "preconsciously" or "preattentively" analyzed for meaning and importance, before the results of this analysis were provided to our conscious awareness (Neisser, 1967; Treisman, 1960). There were two camps in this debate, the "early" and the "late" selection models (Deutsch & Deutsch, 1963; Marcel, 1983; Neisser, 1967; Norman, 1968). Social cognition research continued this tradition by studying the extent to which higher mental processes were put into motion directly by events in the external environment (Bargh & Ferguson, 2000; Bargh & Gollwitzer, 1994; Uleman & Bargh, 1989; Weingarten et al., 2016).

But all of this work—on the self-concept, on stereotypes, on attitudes, on goals themselves—involved the activation of information and mental representations *already stored in memory*. Stereotypes were activated by physical features of a person that were strongly (frequently and consistently) associated with a certain group membership; these stereotypes were learned—via parents, culture, peers, media—and became associated in memory with the diagnostic physical group features (Devine, 1989). Attitudes (in the form of global "good" versus "bad" evaluations) became tightly associated in memory with the representation of the attitude object because they were frequently and consistently generated in past experience with that object (Fazio et al., 1986). And so on. It was all externally driven, and all following standard associative logic (Hebb, 1949): mental contents active at the same time

tended to form associative bonds, which grew stronger the more frequently and consistently they were co-active. They become part of the preconscious analysis of meaning of the environment prior to the information becoming available as inputs into conscious thought and awareness. All of these automatic effects—in cognitive psychology as well (e.g., Corteen & Wood, 1972; Neely, 1977) conformed to the simple definition given in the first ever *Annual Review of Psychology* chapter on social cognition—"the automatic use of stored information" (Higgins & Bargh, 1987, p. 397). The social perceiver when stereotyping an individual is "going beyond the information given" in the current environment, and also when having immediate feelings of liking or disliking upon perceiving an attitude object, but in neither case is that perceiver going beyond the information already stored in memory about that person or object.

In the literature on spontaneous trait inference, something different was happening, unlike all the rest of automatic phenomena in social cognition. Not only were participants going beyond the information given in the behavior descriptions, they were going beyond anything previously stored in memory about them too. They were making, as Winter and Uleman (1984) termed it from the beginning, an *inference*, and by demonstrating this novel effect, they were "inverting" the traditional model of causality attributed to automatic processing. Research on spontaneous trait inference had shown that automaticity was far more sophisticated than we gave it credit for.

Applied work since has validated the spontaneous trait inference effect. For example, Foulk et al. (2016) showed that witnessing of rude behavior activated the concept of rudeness and so increased rudeness-related word fragment completions. The behavior was staged as spontaneous at the start of a class and there were no task instructions on paying attention to it or forming impressions etc. Still, the spontaneous inference of "rude" based on the rude behavior showed its effects in the subsequent word fragment completions. Kawada et al. (2004) similarly showed that subtly inducing a person to act in a given way, such as "nosy" or "helpful," outside of the actual experimental session, caused them to spontaneously encode their own behavior in these ways and consequently become more likely to interpret another person's behavior in these terms (an unconscious form of projection). The mechanism behind these "self-priming" effects had been discovered years earlier by Moskowitz and Roman (1992). In their experiments, participants incidentally formed spontaneous trait inferences when reading behavior descriptions in a first task, and were then found in a subsequent task to be more likely than a control condition to interpret subsequent behaviors into those same trait categories.

The closest historical parallel to spontaneous (unintended, automatic, unconscious) trait inference is Helmholtz's notion of unconscious inferences in visual perception. These occur when we infer, and actually see, environmental features that are literally not actually there. A well-known example is the "dots illusion" created by V. S. Ramachandran (1993; see Kruglanski &

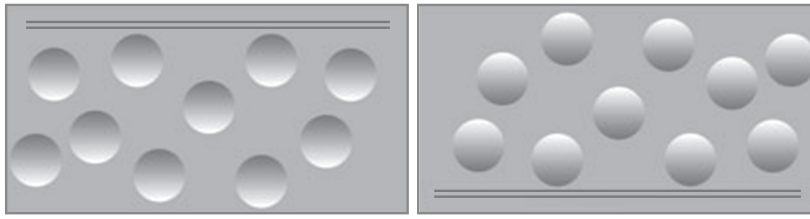

Figure 10.1 Unconscious inferences by a simple heuristic: convex and concave perceptions as function of shading.

Gigerenzer, 2011, for additional examples of sophisticated preconscious inferences), shown in Figure 10.1.

In the left picture, the dots appear concave, pressed into the surface. In the right picture, the dots appear convex, raised up from the surface. The dots in the left picture appear concave, receding into the surface away from the observer, while those on the right side appear convex, curved towards the observer. But in fact, the two pictures are identical, just rotated 180° from each other. If you turn the page upside down, now the formerly right-side dots appear concave, and the formerly left-side ones appear convex.

We "see" these two pictures quite differently because our brain is making unconscious, "spontaneous" inferences based on but certainly *going beyond* the way the shadows cue us into the location of the light source. We don't consciously intend to make this inference, we don't even know we are making it. The same is true of spontaneous trait inference, in which we understand social behavior in terms of personality trait terms, terms that did not appear in the behavioral description itself. This constitutes a higher-order mental process than all of the other automaticity phenomena in social cognition, which involve the direct activation of internally stored information associatively tied to the currently present stimulus information.

The original demonstrations of spontaneous trait inference used verbal descriptions of behavior. Roughly 20 years later, Uleman's former graduate student, Alex Todorov, extended the scope of spontaneous trait inferences to the domain of *faces*. In an extensive and provocative series of studies, Todorov and colleagues showed that participants, upon presentation of a target face photograph, quickly (in as little as 100 ms) infer personality characteristics of that target person, such as trustworthiness, competence, and aggressiveness. There is nothing inherent in the person's physical appearance that should lead to this immediate inference and indeed, these inferences are not diagnostic of the person's actual personality (see Todorov, 2017 for a review). While in most of Todorov's studies the participant is instructed to make trait judgments about the target person, other research has shown that these trait inferences are indeed spontaneous as they occur even when the task goal is unrelated to personality assessment of the target

individual—such as when classifying photographs as to whether they depict houses or people (Slepian et al., 2012).

The Future of Automaticity

What is next for the study of automaticity? If Uleman's strange inversion is anything like those of its predecessors, it is just the beginning of a much broader shift in our understanding of automatic processing. Uleman showed us that we had the causal path from behavioral observations to trait concepts backwards. But what about other causal paths? The paths from objects to evaluations (Fazio et al., 1986), words to concepts (LaBerge & Samuels, 1974), stimuli to goals (Bargh et al., 2001), percepts to actions (Bargh et al., 1996; Wood & Rünger, 2016)—perhaps these are backwards too. Recent work has hinted in this direction. An emerging theme in the literatures on automatic evaluation (Cone et al., 2017; De Houwer, 2014; Kurdi & Banaji, 2017; Melnikoff & Bailey, 2018; Melnikoff et al., 2020; Van Dessel et al., 2018) and habit (Buabang et al., 2021; da Silva & Hare, 2020; de Wit et al., 2018) is that these processes may involve representations far more sophisticated than simple stimulus-response associations—representations akin to the probabilistic generative models underlying spontaneous trait inference. As evidence of this sort accumulates, our picture of automaticity grows increasingly reminiscent of Matisse's *Le Bateau*. Observed right-side up, the artwork at first appears so simple as to seem plain, like scribbles on a page, until the viewer, through an inferential leap, transforms the scribbles into an intricate mental image: a sailboat scuttling across the sea on a breezy day, a sky full of billowing clouds, a reflection of the scene in the water's surface. Subtle in its complexity and masterful in execution, it is an image befitting Jim Uleman's remarkable career.

References

Ajzen, I., & Fishbein, M. (1975). A Bayesian analysis of attribution processes. *Psychological Bulletin*, *82*(2), 261–277.

Anderson, N. H. (1974). Cognitive algebra: Integration theory applied to social attribution. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 7, pp. 1–101). Academic Press.

Atzil, S., Gao, W., Fradkin, I., & Barrett, L. F. (2018). Growing a social brain. *Nature Human Behaviour*, *2*, 624–636.

Bargh, J. A. (1982). Attention and automaticity in the processing of self-relevant information. *Journal of Personality and Social Psychology*, *43*(3), 425–436.

Bargh, J. A. (1989). Conditional automaticity: Varieties of automatic influence in social perception and cognition. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 3–51). Guilford Press.

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. Wyer & T. K. Srull (Eds.),

Handbook of social cognition: Basic processes; Applications (pp. 1–40). Lawrence Erlbaum Associates, Inc.

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, *71*(2), 230–244.

Bargh, J. A., & Ferguson, M. J. (2000). Beyond behaviorism: On the automaticity of higher mental processes. *Psychological Bulletin*, *126*(6), 925–945.

Bargh, J. A., & Gollwitzer, P. M. (1994). Environmental control of goal-directed action: Automatic and strategic contingencies between situations and behavior. In W. D. Spaulding (Ed.), *Nebraska symposium on motivation* (pp. 71–124). University of Nebraska Press.

Bargh, J. A., Gollwitzer, P. M., Lee-Chai, A., Barndollar, K., & Trötschel, R. (2001). The automated will: Nonconscious activation and pursuit of behavioral goals. *Journal of Personality and Social Psychology*, *81*(6), 1014–1027.

Bargh, J. A., & Hassin, R. R. (2021). Human unconscious processes in situ: The kind of awareness that really matters. In A. Reber & A. R. (Eds.), *The cognitive unconscious*. Oxford University Press.

Bem, D. J. (1972). Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 1–63). Academic Press.

Buabang, E. K., Boddez, Y., De Houwer, J., & Moors, A. (2021). Don't make a habit out of it: Impaired learning conditions can make goal-directed behavior seem habitual. *Motivation Science*, *7*, 252–263.

Carlston, D. E., & Skowronski, J. J. (2005). Linking versus thinking: Evidence for the different associative and attributional bases of spontaneous trait transference and spontaneous trait inference. *Journal of Personality and Social Psychology*, *89*(6), 884–898.

Chater, N., Tenenbaum, J. B., & Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Sciences*, *10*(7), 287–291.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204.

Cone, J., Mann, T. C., & Ferguson, M. J. (2017). Changing our implicit minds: How, when, and why implicit evaluations can be rapidly revised. In J. M. Olson (Ed.), *Advances in experimental social psychology* (pp. 131–199). Elsevier Academic Press.

Corteen, R. S., & Wood, B. (1972). Autonomic responses to shock-associated words in an unattended channel. *Journal of Experimental Psychology*, *94*(3), 308–313.

Crawford, M. T., Skowronski, J. J., Stiff, C., & Leonards, U. (2008). Seeing, but not thinking: Limiting the spread of spontaneous trait transference II. *Journal of Experimental Social Psychology*, *44*(3), 840–847.

Crawford, M. T., Skowronski, J. J., Stiff, C., & Scherer, C. R. (2007). Interfering with inferential, but not associative, processes underlying spontaneous trait inference. *Personality and Social Psychology Bulletin*, *33*(5), 677–690.

Darley, J. M., & Fazio, R. H. (1980). Expectancy confirmation processes arising in the social interaction sequence. *American Psychologist*, *35*(10), 867–881.

Darwin, C. (1859). *On the origin of species by means of natural selection, or, the preservation of favoured races in the struggle for life*. Murray.

da Silva, C. F., & Hare, T. A. (2020). Humans primarily use model-based inference in the two-stage task. *Nature Human Behaviour*, *4*(10), 1053–1066.

Dayan, P., Hinton, G. E., Neal, R. M., & Zemel, R. S. (1995). The Helmholtz machine. *Neural Computation*, *7*(5), 889–904.

De Houwer, J. (2014). A propositional model of implicit evaluation. *Social and Personality Psychology Compass*, *8*(7), 342–353.

de Wit, S., Kindt, M., Knot, S. L., Verhoeven, A. A., Robbins, T. W., Gasull-Camos, J., Evans, M., Mirza, H., & Gillan, C. M. (2018). Shifting the balance between goals and habits: Five failures in experimental habit induction. *Journal of Experimental Psychology: General*, *147*(7), 1043–1065.

Dennett, D. (2009). Darwin's "strange inversion of reasoning". *Proceedings of the National Academy of Sciences*, *106*(Supplement 1), 10061–10065.

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, *70*(1), 80–90.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*(1), 5–18.

Doyle, C. M., & Lindquist, K. A. (2018). When a word is worth a thousand pictures: Language shapes perceptual memory for emotion. *Journal of Experimental Psychology: General*, *147*(1), 62–73.

Erdelyi, M. H. (1974). A new look at the New Look: Perceptual defense and vigilance. *Psychological Review*, *81*(1), 1–25.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, *50*(2), 229–238.

Festinger, L. (1957). *A theory of cognitive dissonance*. Row Peterson.

Foulk, T., Woolum, A., & Erez, A. (2016). Catching rudeness is like catching a cold: The contagion effects of low-intensity negative behaviors. *Journal of Applied Psychology*, *101*(1), 50–67.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138.

Ginossar, Z., & Trope, Y. (1987). Problem solving in judgment under uncertainty. *Journal of Personality and Social Psychology*, *52*(3), 464–474.

Goren, A., & Todorov, A. (2009). Two faces are better than one: Eliminating false trait associations with faces. *Social Cognition*, *27*(2), 222–248.

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review*, *114*(2), 211–244.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814–834.

Hamilton, D. L., & Gifford, R. K. (1976). Illusory correlation in interpersonal perception: A cognitive basis of stereotypic judgments. *Journal of Experimental Social Psychology*, *12*(4), 392–407.

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. Wiley.

Higgins, E. T., & Bargh, J. A. (1987). Social cognition and social perception. *Annual Review of Psychology*, *38*(1), 369–425.

James, W. (1884). What is an emotion? *Mind*, *9*, 188–205.

Kawada, C. L., Oettingen, G., Gollwitzer, P. M., & Bargh, J. A. (2004). The projection of implicit and explicit goals. *Journal of Personality and Social Psychology*, *86*(4), 545–559.

Keren, G., & Schul, Y. (2009). Two is not always better than one: A critical evaluation of two-system theories. *Perspectives on Psychological Science*, *4*(6), 533–550.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*(12), 712–719.

Kressel, L. M., & Uleman, J. S. (2010). Personality traits function as causal concepts. *Journal of Experimental Social Psychology*, *46*(1), 213–216.

Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberate judgments are based on common principles. *Psychological Review*, *118*(1), 97–109.

Kurdi, B., & Banaji, M. R. (2017). Repeated evaluative pairings and evaluative statements: How effectively do they shift implicit attitudes? *Journal of Experimental Psychology: General*, *146*(2), 194–213.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*(2), 293–323.

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A*, *20*(7), 1434–1448.

Lindquist, K. A., & Gendron, M. (2013). What's in a word? Language constructs emotion perception. *Emotion Review*, *5*(1), 66–71.

MacKenzie, R. B. (1868). *The Darwinian theory of the transmutation of species examined*. Nisbet & Co.

Mae, L., McMorris, L. E., & Hendry, J. L. (2004). Spontaneous trait transference from dogs to owners. *Anthrozoös*, *17*(3), 225–243.

Marcel, A. J. (1983). Conscious and unconscious perception: An approach to the relations between phenomenal experience and perceptual processes. *Cognitive Psychology*, *15*(2), 238–300.

McArthur, L. Z. (1980). Illusory causation and illusory correlation: Two epistemological accounts. *Personality and Social Psychology Bulletin*, *6*(4), 507–519.

Melnikoff, D. E., & Bailey, A. H. (2018). Preferences for moral vs. immoral traits in others are conditional. *Proceedings of the National Academy of Sciences*, 201714945.

Melnikoff, D. E., & Bargh, J. A. (2018). The mythical number two. *Trends in Cognitive Sciences*, *22*(4), 280–293.

Melnikoff, D. E., Lambert, R., & Bargh, J. A. (2020). Attitudes as prepared reflexes. *Journal of Experimental Social Psychology*, *88*, 103950.

Moskowitz, G. B., & Olcaysoy Okten, I. (2016). Spontaneous goal inference (SGI). *Social and Personality Psychology Compass*, *10*(1), 64–80.

Moskowitz, G. B., & Roman, R. J. (1992). Spontaneous trait inferences as self-generated primes: Implications for conscious social judgment. *Journal of Personality and Social Psychology*, *62*(5), 728–738.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, *106*(3), 226–254.

Neisser, U. (1967). *Cognitive psychology*. Appleton-Century-Crofts.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Prentice-Hall.

Norman, D. A. (1968). Toward a theory of memory and attention. *Psychological Review*, *75*(6), 522–536.

Ramachandran, V. S. (1993). Behavioral and magnetoencephalographic correlates of plasticity in the adult human brain. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(22), 10413.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87.

Schachter, S., & Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review*, *69*(5), 379–399.

Skowronski, J. J., Carlston, D. E., Mae, L., & Crawford, M. T. (1998). Spontaneous trait transference: Communicators take on the qualities they describe in others. *Journal of Personality and Social Psychology*, *74*(4), 837–848.

Slepian, M. L., Young, S. G., Rule, N. O., Weisbuch, M., & Ambady, N. (2012). Embodied impression formation: Social judgments and motor cues to approach and avoidance. *Social Cognition*, *30*(2), 232–240.

Tamir, D. I., & Thornton, M. A. (2018). Modeling the predictive social mind. *Trends in Cognitive Sciences*, *22*(3), 201–212.

Taylor, S. E., & Fiske, S. T. (1978). Salience, attention, and attribution: Top of the head phenomena. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 11, pp. 249–288). Academic Press.

Todorov, A. (2017). *Face value: The irresistible influence of first impressions*. Princeton University Press.

Todorov, A., & Uleman, J. S. (2004). The person reference process in spontaneous trait inferences. *Journal of Personality and Social Psychology*, *87*(4), 482–493.

Treisman, A. M. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, *12*(4), 242–248.

Trope, Y. (1986). Identification and inferential processes in dispositional attribution. *Psychological Review*, *93*(3), 239–257.

Uleman, J. S. (1999). Spontaneous versus intentional inferences in impression formation. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 141–160). The Guilford Press.

Uleman, J. S., & Bargh, J. A. (1989). *Unintended thought*. Guilford Press.

Uleman, J. S., Newman, L. S., & Moskowitz, G. B. (1996). People as flexible interpreters: Evidence and issues from spontaneous trait inference. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 28). Academic Press.

Van Dessel, P., Hughes, S., & De Houwer, J. (2018). How do actions influence attitudes? An inferential account of the impact of action performance on stimulus evaluation. *Personality and Social Psychology Review*, *23*(3), 267–284.

Van Overwalle, F., Van Duynslaeger, M., Coomans, D., & Timmermans, B. (2012). Spontaneous goal inferences are often inferred faster than spontaneous trait inferences. *Journal of Experimental Social Psychology*, *48*(1), 13–18.

Weingarten, E., Chen, Q., McAdams, M., Yi, J., Hepler, J., & Albarracín, D. (2016). From primed concepts to action: A meta-analysis of the behavioral effects of incidentally presented words. *Psychological Bulletin*, *142*(5), 472–497.

Wells, B. M., Skowronski, J. J., Crawford, M. T., Scherer, C. R., & Carlston, D. E. (2011). Inference making and linking both require thinking: Spontaneous trait inference and spontaneous trait transference both rely on working memory capacity. *Journal of Experimental Social Psychology*, *47*(6), 1116–1126.

Wigboldus, D. H., Dijksterhuis, A., & van Knippenberg, A. (2003). When stereotypes get in the way: Stereotypes obstruct stereotype-inconsistent trait inferences. *Journal of Personality and Social Psychology*, *84*(3), 470–484.

Winter, L., & Uleman, J. S. (1984). When are social judgments made? Evidence for the spontaneousness of trait inferences. *Journal of Personality and Social Psychology*, *47*(2), 237–252.

Wood, W., & Rünger, D. (2016). Psychology of habit. *Annual Review of Psychology*, *67*, 289–314.

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, *35*(2), 151–175.

11 Unintentional Influences in Intentional Impression Formation[1]

Bertram Gawronski, Skylar M. Brannon, and Dillon M. Luke
University of Texas at Austin

A substantial body of research suggests that perceivers spontaneously draw inferences from observed behaviors even when they do not have the intention to form a social impression. Such unintentional inferences have been found to give rise to impressions of other people's traits (i.e., spontaneous trait inference; see Uleman et al., 1996) and goals (i.e., spontaneous goal inference; see Moskowitz & Olcaysoy Okten, 2016). For example, when learning that Avery received an A on a math exam, people may spontaneously infer that Avery is smart; and when learning that Alex donated $100 to a local food bank, people may spontaneously infer that Alex had the goal to help. Although these impressions can be the result of intentional processes, the notion of spontaneous inference suggests that they may also arise from unintentional processes.

The current chapter reviews research on a related, yet conceptually distinct phenomenon: unintentional influences in intentional impression formation. The central focus of our review is on the finding that mere co-occurrence of stimuli can produce evaluative responses that are diametrically opposite to intentionally formed impressions based on the particular relation between the co-occurring stimuli. This phenomenon is similar to the concept of spontaneous inference, in that it involves unintentional effects in impression formation. However, it is different from the concept of spontaneous inference, in that it arises in contexts where people do have the intention to form an impression. Another important difference is that, while prior research on spontaneous inference has predominantly focused on impressions with specific semantic content (e.g., intelligent vs. unintelligent), evidence for unintentional influences in intentional impression formation is primarily coming from studies on broad evaluative impressions (e.g., good vs. bad).[2]

In the first part of this chapter, we illustrate the differential effects of mere co-occurrence and relational information in impression formation. Expanding on this distinction, the second part reviews evidence for unintentional influences in intentional impression formation, as reflected in dissociative effects of mere co-occurrence and relational information on implicit and explicit measures. The third part describes a novel approach to identify effects of mere

DOI: 10.4324/9781003045687-13

co-occurrence and relational information via formal modeling. In the fourth part, we discuss competing theoretical explanations for unintentional influences in intentional impression formation and evidence regarding the impact of theoretically derived moderators that make such influences more or less likely to occur. In the final part, we discuss broader implications of the reviewed research for impression formation.

Effects of Mere Co-occurrence and Relational Information

Unintentional influences in intentional impression formation can occur in various forms, as demonstrated by classic research on halo and priming effects in impression formation. In the current chapter, we focus on a more recent line of work suggesting that evaluative responses to an object may be jointly influenced by (1) the mere co-occurrence of the object with a pleasant or unpleasant stimulus (e.g., mere co-occurrence of object A and negative event B) and (2) the object's particular relation to the co-occurring stimulus (e.g., object A starts vs. stops negative event B). To illustrate the difference between mere co-occurrence and relational information, imagine a hypothetical health campaign that aims to promote the use of sunscreen with the message that sunscreen protects against skin cancer. To the extent that people understand and accept this message, the presented information about the relation between sunscreen and skin cancer should lead to a positive response to sunscreen. Yet, in line with the notion of *evaluative conditioning* (EC), the same message could also lead to a negative response to sunscreen due to the mere co-occurrence of *sunscreen* with the negative concept *skin cancer* in the message. EC is commonly defined as the change in the evaluation of a conditioned stimulus (CS) due to its pairing with a positive or negative unconditioned stimulus (US; see De Houwer, 2007). In our thematic example, the mere pairing of sunscreen (CS) and skin cancer (US) in the message may produce an EC effect on evaluative responses to sunscreen that is diametrically opposite to the effect that can be expected if recipients comprehend and accept the causal relation of sunscreen and skin cancer described in the message. Whereas mere co-occurrence should lead to a negative response to sunscreen, relational information should lead to a positive response to sunscreen.

Conceptually, the relation of an object and a co-occurring stimulus can be described as *assimilative* when it suggests an evaluative response to the object that is in line with the valence of the co-occurring stimulus (e.g., smoking causes lung cancer). Conversely, the relation of an object and a co-occurring stimulus can be described as *contrastive* when it suggests an evaluative response to the object that is opposite to the valence of the co-occurring stimulus (e.g., sunscreen prevents skin cancer). At the operational level, unintentional influences in intentional impression formation can be inferred when the following three conditions are met: (1) a given object has a contrastive relation to a positive or negative stimulus, (2) people intentionally use the object's contrastive relation to the co-occurring stimulus in forming

an impression of the object, and (3) evaluative responses to the object are nevertheless influenced by its mere co-occurrence with the stimulus. To the extent that all three conditions are met, the effect under Point 3 can be interpreted as unintentional influence in intentional impression formation. For example, a message stating that sunscreen protects against skin cancer can be said to have an unintentional influence in intentional impression formation when message recipients intentionally form a positive impression of sunscreen in response to the message, but nevertheless show a negative response to sunscreen due to the mere co-occurrence of sunscreen and skin cancer in the message. In the following sections, we review empirical evidence for unintentional influences in intentional impression formation in terms of these three defining characteristics.

Evidence in Research Using Implicit and Explicit Measures

Preliminary evidence for unintentional influences in intentional impression formation comes from several studies using a combination of implicit and explicit measures to identify effects of mere co-occurrence and relational information. The central finding in this line of work is that implicit measures (e.g., implicit association test; evaluative priming task; for an overview, see Gawronski & De Houwer, 2014) sometimes reflect effects of mere co-occurrence even when explicit measures (e.g., evaluative rating scales) reflect effects of relational information.

In the first demonstration of such dissociative effects, Moran and Bar-Anan (2013) presented participants with sequences of images and sounds. Each sequence started with an image of one alien creature, followed by either a pleasant or an unpleasant sound (i.e., pleasant melody or unpleasant scream), followed by an image of a different alien creature. Participants were told that, depending on their position in the sequence, some aliens would start the following sound whereas other aliens would stop the preceding sound. Participants were asked to form an impression of the alien creatures based on the presented information. After the impression formation task, evaluative responses to the alien creatures were measured with an explicit and an implicit measure (i.e., implicit association test; see Greenwald et al., 1998). Whereas responses on the explicit measure reflected the particular relation of the aliens to the sounds, responses on the implicit measure reflected the mere co-occurrence of aliens and sounds regardless of their relation. Specifically, on the explicit measure, participants showed more favorable judgments of aliens that started pleasant sounds compared with aliens that stopped pleasant sounds. Conversely, participants showed less favorable judgments of aliens that started unpleasant sounds compared with aliens that stopped unpleasant sounds. In contrast, on the implicit measure, participants showed more favorable responses to aliens that co-occurred with pleasant sounds compared with aliens that co-occurred with unpleasant sounds, regardless of whether the aliens started or stopped the sounds.

Similar findings were obtained by Hu et al. (2017, Experiments 1 and 2). Participants were presented with image pairs involving pharmaceutical products and positive or negative health conditions (e.g., healthy hair, skin rash). Half of the participants were told that the pharmaceutical products cause the depicted health conditions; the other half was told that the pharmaceutical products prevent the depicted health conditions. Participants were asked to form an impression of the pharmaceutical products based on the presented information. After the impression formation task, evaluative responses to the pharmaceutical products were measured with an explicit and an implicit measure (i.e., evaluative priming task; see Fazio et al., 1995). Consistent with Moran and Bar-Anan's (2013) results, Hu et al. found that responses on the explicit measure reflected the relation between the pharmaceutical products and the depicted health conditions. In contrast, responses on the implicit measure reflected the mere co-occurrence of the products with the depicted health conditions regardless of their relation. Specifically, on the explicit measure, participants showed more favorable judgments of products that caused positive health conditions compared with products that prevented positive health conditions. Conversely, participants showed less favorable judgments of products that caused negative health conditions compared with products that prevented negative health conditions. In contrast, on the implicit measure, participants showed more favorable responses to products that co-occurred with positive health conditions than products that co-occurred with negative health conditions, regardless of whether the products caused or prevented the health conditions.

The findings by Moran and Bar-Anan (2013) and Hu et al. (2017) are consistent with the notion of unintentional influences in intentional impression formation. When the focal objects had a contrastive relation to a co-occurring stimulus, evaluative responses on implicit measures were influenced by mere co-occurrence, although responses on explicit measures reflected the intentional use of relational information in forming impressions of the focal objects. However, a more exhaustive review of the available evidence suggests that unqualified co-occurrence effects on implicit measures are not a ubiquitous outcome (see Kurdi & Dunham, 2020). Although some studies found mere co-occurrence effects on implicit measures that remained unqualified by relational information (e.g., Hu et al., 2017, Experiments 1 and 2; Moran & Bar-Anan, 2013), other studies found attenuated co-occurrence effects when the co-occurring stimuli had a contrastive relation (e.g., Zanon et al., 2012; Zanon et al., 2014). Yet, other studies found a full reversal of mere co-occurrence effects in cases involving contrastive relations (e.g., Gawronski et al., 2005; Hu et al., 2017, Experiment 3), suggesting that intentional processes completely overrode unintentional effects of mere co-occurrence. Together, these mixed findings suggest that the relative impact of mere co-occurrence and relational information on implicit measures may depend on specific conditions.

To date, there is empirical evidence for two moderators that seem to influence mere co-occurrence effects on implicit measures in the presence of contrastive relational information. First, Hu et al. (2017) found dissociative effects of mere co-occurrence and relational information on implicit and explicit measures only when the relational information was provided before the impression formation task and this information was consistent for all of the presented target stimuli (i.e., all of the pharmaceutical products either caused or prevented the depicted health conditions; see Experiments 1 and 2). However, when relational information was provided during the impression task and the specific relations varied on a trial-by-trial basis, both implicit and explicit measures were influenced by relational information without showing any effect of mere co-occurrence (Experiment 3). Second, Moran et al. (2015) found stronger mere co-occurrence effects on an implicit measure when participants were instructed to memorize the co-occurrence of the stimuli than when they were asked to form an impression of the target objects. However, memorization instructions also eliminated the effect of relational information on an explicit measure, which was influenced by mere co-occurrence instead of relational information under memorization conditions. Although these results suggest that effects of mere co-occurrence and relational information are goal-dependent, it is worth noting that the critical dissociation between implicit and explicit measures replicated under impression-formation instructions. In this case, the implicit measure was influenced by mere co-occurrence, while the explicit measure reflected the intentional use of relational information in forming impressions of the focal objects.

In sum, research using implicit and explicit measures provides mixed support for the idea that mere co-occurrence can have unintentional effects when people intentionally use contrastive relational information in forming impressions. When a CS has a contrastive relation to a co-occurring US, CS evaluations on explicit measures are typically opposite to the valence of the co-occurring US (e.g., more favorable evaluation of sunscreen in response to the message *sunscreen prevents skin cancer*), indicating that the contrastive relation influenced intentionally formed impressions. Yet, effects on implicit measures are inconsistent across studies, in that some studies found CS evaluations reflecting the valence of the US regardless of their relation (e.g., less favorable evaluation of sunscreen in response to the message *sunscreen prevents skin cancer*); some studies found CS evaluations that were opposite to the valence of the co-occurring US (e.g., more favorable evaluation of sunscreen in response to the message *sunscreen prevents skin cancer*); and some studies have found no effect at all (e.g., no change in the evaluation of sunscreen in response to the message *sunscreen prevents skin cancer*). Although a small number of studies has identified factors that make mere co-occurrence effects on implicit measures more or less likely to occur, the available evidence in research using implicit and explicit measures to study unintentional influences in intentional impression formation is mixed and

somewhat inconsistent. As we explain in the next section, at least some of these inconsistencies may be due to methodological limitations of using a task-dissociation approach to identify effects of mere co-occurrence and relational information.

Evidence in Research Using Multinomial Modeling

A major disadvantage of using a combination of implicit and explicit measures to identify effects of mere co-occurrence and relational information is that the two kinds of measures differ in numerous ways (for a discussion, see Payne et al., 2008). The large number of differences makes it impossible to identify which of these differences is responsible for the differential sensitivity to mere co-occurrence and relational information (see also Bading et al., 2020; Green et al., 2021). A superior approach that resolves this problem is the use of formal modeling procedures to estimate the impact of mere co-occurrence and relational information on responses within a single task. Indeed, research using multinomial modeling (Batchelder & Riefer, 1999; Erdfelder et al., 2009; Hütter & Klauer, 2016) to quantify effects of mere co-occurrence and relational information (e.g., Gawronski & Brannon, 2021; Heycke & Gawronski, 2020; Kukken et al., 2020) has obtained much more consistent evidence compared to studies that have used a task-dissociation approach.

The basic idea underlying the multinomial modeling approach can be illustrated by means of a processing tree that specifies potential patterns of responses to a target object as a function of whether the object has either an assimilative or a contrastive relation to either a positive or a negative stimulus (see Figure 11.1). The four paths on the left side of the figure depict the four potential cases that (1) responses to the object are driven by its relation to a co-occurring stimulus, (2) responses to the object are driven by its mere co-occurrence with the stimulus, (3) responses to the object are driven by a general positivity bias, and (4) response to the object are driven by a general negativity bias. The table on the right side of the figure depicts the response patterns for each of the four cases as a function of relational information and the valence of the co-occurring stimulus.

If responses to a given object are driven by relational information, participants should show a positive response when the object has an assimilative relation with a positive stimulus or a contrastive relation with a negative stimulus, and participants should show a negative response when the object has a contrastive relation with a positive stimulus or an assimilative relation with a negative stimulus (first path in Figure 11.1). If responses to a given object are driven by mere co-occurrence, participants should show a positive response when the object co-occurs with a positive stimulus and a negative response when the objects co-occurs with a negative stimulus (second path in Figure 11.1). If responses to a given object are driven by a general positivity bias, participants should show a positive response regardless of the valence of

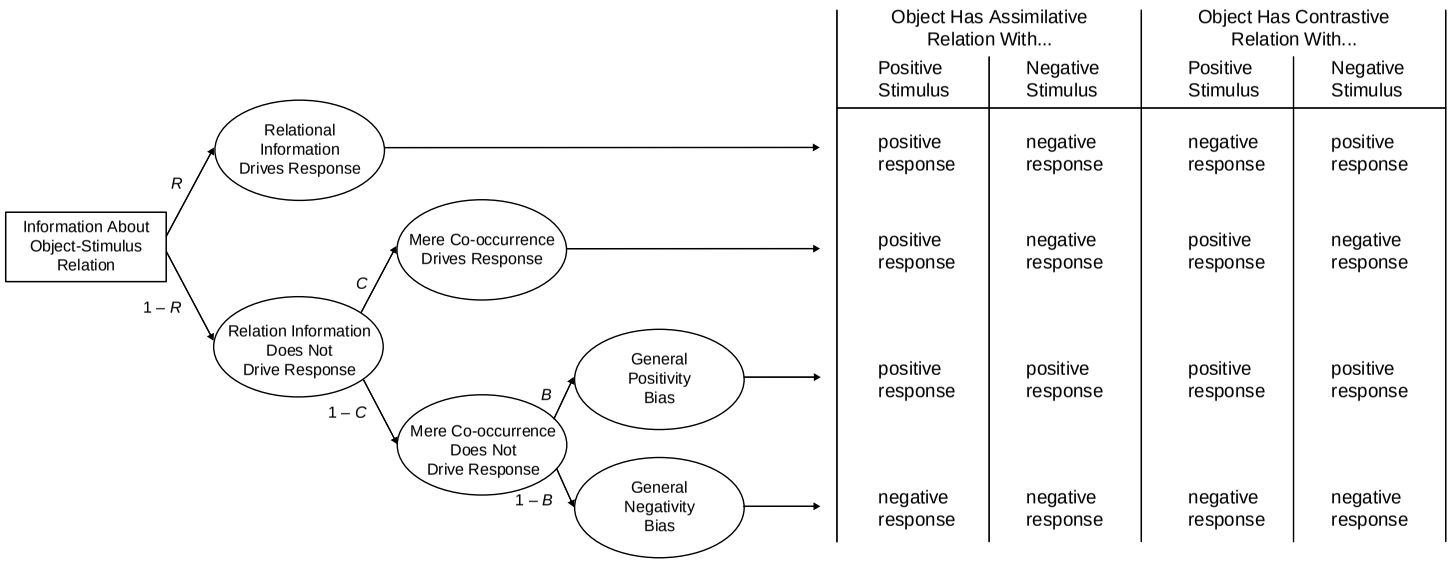

Figure 11.1 Multinomial processing tree depicting effects of relational information, mere co-occurrence, and general response biases on evaluative responses (positive response vs. negative response) as a function of relational information (assimilative relation vs. contrastive relation) and valence of co-occurring stimulus (positive stimulus vs. negative stimulus).

the co-occurring stimulus and the object's relation to that stimulus (third path in Figure 11.1). Conversely, if responses to a given object are driven by a general negativity bias, participants should show a negative response regardless of the valence of the co-occurring stimulus and the object's relation to that stimulus (fourth path in Figure 11.1).

Based on the processing tree depicted in Figure 11.1, multinomial modeling provides numerical estimates for (1) the probability that relational information drives responses (captured by the parameter *R* in Figure 11.1); (2) the probability that mere co-occurrence drives responses if relational information does not drive responses (captured by the parameter C in Figure 11.1); and (3) the probability that a general positivity or negativity bias drives responses if neither relational information nor mere co-occurrence drive responses (captured by the parameter *B* in Figure 11.1).[3] Numerical scores for the three probabilities are estimated by means of four non-redundant mathematical equations derived from the processing tree (see Appendix).[4] These equations include the three model parameters (*R*, C, *B*) as unknowns and the empirically observed probabilities of *positive* versus *negative* responses in the four object conditions (i.e., assimilative relation to positive stimulus; assimilative relation to negative stimulus; contrastive relation to positive stimulus; contrastive relation to negative stimulus) as known numerical values. Using maximum likelihood statistics, multinomial modeling generates numerical estimates for the three unknowns that minimize the discrepancy between the empirically observed probabilities of positive versus negative responses in the four object conditions and the probabilities of positive versus negative responses predicted by the model equations using the generated parameter estimates.

The adequacy of the model in describing the data can be evaluated by means of goodness-of-fit statistics, with poor model fit being reflected in a statistically significant discrepancy between the empirically observed probabilities in a given data set and the probabilities predicted by the model. The estimated scores for each parameter can vary between 0 and 1. For the *R* parameter, scores significantly greater than zero indicate that responses were affected by relational information. For the C parameter, scores significantly greater than zero indicate that responses were affected by mere co-occurrence. Finally, for the *B* parameter, scores significantly greater than 0.5 indicate a general positivity bias and scores significantly lower than 0.5 indicate a general negativity bias.

Differences from these reference points can be tested by enforcing a specific value for a given parameter and comparing the fit of the restricted model to the fit of the unrestricted model. If setting a given parameter equal to a specific reference point leads to a significant reduction in model fit, it can be inferred that the parameter estimate is significantly different from that reference point. For example, to test whether mere co-occurrence influenced responses, the C parameter is set equal to zero and the resulting model fit is compared to the fit of the model that does not include any restrictions for the C parameter. To the extent that enforcing a parameter estimate of zero leads

to a significant reduction in model fit, it can be inferred that mere co-occurrence significantly influenced participants' responses. The same approach can be used to test the influence of relational information captured by the *R* parameter. For the *B* parameter, comparisons to reference values are equivalent, except that the reference value reflecting the absence of a general response bias is 0.5. Similar tests can be conducted to investigate whether estimates for a given parameter significantly differ across groups, which can be tested by enforcing equal estimates for that parameter across groups. If setting a given parameter equal across groups leads to a significant reduction in model fit, it can be inferred that the parameter estimates for the two groups are significantly different.

A major advantage of the multinomial modeling approach is that it allows researchers to quantify effects of mere co-occurrence and relational information to overt responses on a single task, and this task can be rather simple (e.g., binary forced-choice judgments) without requiring a high level of procedural complexity (as it is the case for implicit measures). For example, combining Moran and Bar-Anan's (2013) impression-formation paradigm with a simple forced-choice task, Kukken et al. (2020) found that participants' responses to the alien creatures were influenced by both (1) their mere co-occurrence with a pleasant or unpleasant sound and (2) their particular relation to the co-occurring sound (i.e., whether they started or stopped the sound). Similarly, combining Hu et al.'s (2017) impression-formation paradigm with a simple forced-choice task, Heycke and Gawronski (2020) found that participants' responses to the pharmaceutical products were influenced by both (1) their mere co-occurrence with a pleasant or unpleasant health condition and (2) their particular relation to the co-occurring health condition (i.e., whether they caused or prevented the health condition). Interestingly, Heycke and Gawronski obtained reliable effects of mere co-occurrence with a procedural setup that failed to produce mere co-occurrence effects on implicit measures in Hu et al.'s research (Experiment 3). Although studies using a multinomial modeling approach have identified several contextual factors that moderate the relative impact of mere co-occurrence and relational information (see below), the obtained results provide strong support for the idea that mere co-occurrence can have unintentional effects when people intentionally use contrastive relational information in forming impressions.

Theoretical Explanations

A common explanation for joint effects of mere co-occurrence and relational information is that they are the products of two functionally distinct mechanisms operating during the learning of new information. For example, according to the associative-propositional evaluation (APE) model (Gawronski & Bodenhausen, 2006, 2011, 2014, 2018), mere co-occurrence effects are the product of an associative learning mechanism involving the

automatic formation of mental associations between co-occurring stimuli. In contrast, effects of relational information are claimed to be the product of a propositional learning mechanism involving the non-automatic generation and truth assessment of mental propositions about the relation between co-occurring stimuli. Based on the hypothesis that effects of mere co-occurrence and relational information are mediated by two distinct learning mechanisms, such accounts have been described as *dual-process learning accounts*.

An alternative explanation is offered by theories that interpret all learning effects as outcomes of a single propositional mechanism involving the non-automatic generation and truth assessment of mental propositions about stimulus relations (e.g., De Houwer, 2009, 2018; De Houwer et al., 2020). According to these theories, distinct effects of mere co-occurrence and relational information result from processes during the retrieval of stored propositional information rather than two functionally distinct learning mechanisms. For example, based on the assumptions of the integrated propositional model (IPM; De Houwer, 2018), mere co-occurrence effects can be expected to occur despite the successful learning of contrastive relational information when the retrieval of a stored proposition about a contrastive relation is incomplete (e.g., retrieval of *A is related to B* rather than *A stops B*; see Van Dessel et al., 2019). Based on the hypothesis that effects of mere co-occurrence and relational information can arise from a single propositional learning mechanism, such accounts have been described as *single-process learning accounts*.[5]

A major difference between the two accounts concerns the presumed (in)dependence of contextual effects on the impact of mere co-occurrence and relational information. Dual-process learning accounts such as the APE model suggest that contextual effects on the impact of mere co-occurrence and relational information are largely independent, in that a given factor may influence one without affecting the other. The critical question is whether a given contextual factor influences either (1) the automatic formation of mental associations between co-occurring stimuli or (2) the non-automatic generation and truth assessment of mental propositions about the relation between co-occurring stimuli (see Gawronski & Bodenhausen, 2006, 2007, 2011, 2018). In contrast, single-process learning accounts such as the IPM suggest that contextual factors should moderate the impact of mere co-occurrence and relational information in a complementary fashion. According to single-process learning theories, effects of mere co-occurrence in cases involving contrastive relations are due to incomplete retrieval of stored propositions about the relation between co-occurring stimuli. Thus, any factor that supports complete retrieval of stored propositions should increase the impact of relational information and reduce the impact of mere co-occurrence. Conversely, any factor that interferes with a complete retrieval of stored propositions should decrease the impact of relational information and increase the impact of mere co-occurrence (see De Houwer, 2018; De Houwer et al., 2020; Van Dessel et al., 2019).

The multinomial modeling approach is ideally suited for empirical tests of these competing predictions, because it permits experimental manipulations of contextual conditions during learning and retrieval while keeping everything else constant (Heycke & Gawronski, 2020). The latter is not feasible with the task-dissociation approach comparing responses on implicit and explicit measures, because it always includes multiple procedural differences between measurement instruments in addition to the focal difference of interest in the experimental manipulation (see Corneille & Mertens, 2020; Sherman et al., 2014). In the following sections, we review empirical evidence that speaks to competing predictions derived from dual-process and single-process accounts regarding the impact of various contextual conditions during learning and retrieval. In line with the proclaimed superiority of the multinomial modeling approach in testing these predictions, we focus specifically on studies that quantified effects of mere co-occurrence and relational information via multinomial modeling. Although some of the reviewed findings pose a challenge to both dual-process and single-process learning accounts, the available evidence provides valuable insights into unintentional influences in intentional impression formation by identifying factors that do or do not moderate such influences.

Time for Encoding

The amount of time devoted to the processing of new information during learning is an important determinant of memory strength (Craik & Lockhart, 1972). The more people elaborate on new information during encoding, the more likely it is that this information is successfully retrieved at a later time. These assumptions are shared by both dual-process and single-process accounts, which both suggest that more time for encoding should support the storage of relational information during learning, and thereby its subsequent retrieval. Hence, both dual-process and single-process accounts suggest that more time for encoding should increase effects of relational information. Yet, the two accounts have different implications for effects of mere co-occurrence. According to dual-process learning accounts, mere co-occurrence effects result from the automatic formation of mental associations between co-occurring stimuli, which should be independent of the available time to elaborate on new information. Thus, although more time for encoding should increase the impact of relational information, the impact of mere co-occurrence should be unaffected by time for encoding. In contrast, single-process learning accounts assume that mere co-occurrence effects result from incomplete retrieval of stored propositions about the relation between co-occurring stimuli. Thus, to the extent that more time for encoding supports the complete retrieval of stored information, it should increase the impact of relational information and reduce the impact of mere co-occurrence. Evidence addressing this question was presented by Heycke and Gawronski (2020, Experiments 2a and 2b) who found that more time for encoding significantly increased the impact of

relational information (consistent with both accounts) without affecting the impact of mere co-occurrence (consistent with dual-process learning accounts).

Repetition

Although dual-process learning accounts suggest that mere co-occurrence effects should be unaffected by how much people elaborate on new information, they predict that mere co-occurrence effects should increase as a function of repetition. This prediction is based on the assumption that mental associations between two stimuli should become stronger with increasing frequency of their co-occurrence (Smith & DeCoster, 2000). At the same time, repetition should support the storage of information about stimulus relations, and thereby the subsequent retrieval of this information. From this perspective, repetition should increase effects of both mere co-occurrence and relational information. In contrast, from a single-process learning view, repetition should support the storage of information about stimulus relations, and thereby a complete retrieval of this information. From this perspective, repetition should increase effects of relational information and decrease effects of mere co-occurrence. Interestingly, the available evidence regarding the impact of repetition on mere co-occurrence effects conflicts with both accounts. Specifically, Heycke and Gawronski (2020, Experiment 3) found that repetition significantly increased the impact of relational information (consistent with both accounts), but repetition had no significant effect on the impact of mere co-occurrence (inconsistent with both accounts).

Time during Judgment

Although dual-process and single-process learning accounts lead to different predictions regarding the impact of time for encoding, the two accounts have the same implications for the impact of time during judgment. According to dual-process accounts such as the APE model, effects of activated associations on judgments and behavior should be reduced when deliberate propositional reasoning leads to a rejection of the spontaneous evaluative response elicited by automatically activated associations (Gawronski & Bodenhausen, 2006, 2007, 2011, 2018). From this perspective, more time during judgment should have compensatory effects, in that it should increase effects of relational information and decrease effects of mere co-occurrence. Similarly, single-process accounts such as the IPM suggest that more time during judgment should support a complete retrieval of stored information about stimulus relations, which should increase effects of relational information and decrease effects of mere co-occurrence. Interestingly, the available evidence conflicts with the shared prediction regarding the impact of time during judgment on mere co-occurrence effects. Specifically,

Heycke and Gawronski (2020, Experiment 4) found that more time during judgment increased the impact of relational information (consistent with both accounts), but it also increased—rather than decreased—the impact of mere co-occurrence (inconsistent with both accounts).

Temporal Delay

Another factor for which the two accounts lead to different predictions is the temporal delay between encoding and judgment. Some dual-process learning accounts suggest that mental representations of relational information involve multiple layers within associative networks (Smith & DeCoster, 2000). According to such multi-layer network theories, activated concepts at higher levels specify the relation between activated concepts at lower levels (Gawronski & Bodenhausen, 2018; Gawronski et al., 2017). Thus, to the extent that hierarchical representations involving multiple layers of associative links are more likely affected by memory decay compared to direct associative links between two concepts, effects of mere co-occurrence should be more stable over time compared to effects of relational information. From this perspective, longer temporal delays between encoding and judgment should reduce the impact of relational information, with the impact of mere co-occurrence being less affected by temporal delays. In contrast, single-process learning accounts suggest that memory decay associated with temporal delays should increase the likelihood of incomplete retrieval of stored information about stimulus relations. From this perspective, a longer temporal delay between encoding and judgment should decrease effects of relational information and increase effects of mere co-occurrence. Evidence addressing this question was presented by Heycke and Gawronski (2020, Experiment 5) who found that a two-day delay between encoding and judgment decreased the impact of relational information (consistent with both accounts) without affecting the impact of mere co-occurrence (consistent with dual-process learning accounts).

Intentional Control

Another difference between the two accounts concerns the presumed impact of intentional control. According to dual-process learning accounts, enhanced attention to relational information during encoding should support the storage of this information, thereby increasing its effect on judgments. However, enhanced attention to relational information during encoding should have little impact on the effect of mere co-occurrence, which is assumed to result from the automatic formation of mental associations between co-occurring stimuli (see Gawronski & Bodenhausen, 2014). From this perspective, enhanced motivation to intentionally control the impact of mere co-occurrence by focusing on stimulus relations should increase the impact of relational information without affecting the impact of mere co-occurrence. In contrast,

single-process learning accounts suggest that enhanced attention of relational information during encoding should support the storage of information about stimulus relations, and thereby the complete retrieval of this information. From this perspective, enhanced motivation to intentionally control the impact of mere co-occurrence by focusing on stimulus relations should increase the impact of relational information and decrease the impact of mere co-occurrence. Evidence addressing this question was presented by Gawronski and Brannon (2021) who found that enhanced motivation to intentionally control the impact of mere co-occurrence by focusing on stimulus relations increased the impact of relational information (consistent with both accounts) without affecting the impact of mere co-occurrence (consistent with dual-process learning accounts). Similar findings were obtained by Kukken et al. (2020, Experiment 4).

Summary

Research testing competing predictions of dual-process and single-process learning accounts has provided valuable insights into unintentional influences in intentional impression formation by identifying factors that do moderate such influences and factors that do not. In line with the shared predictions of dual-process and single-process accounts, effects of relational information have been found to increase with more time for encoding, more frequent repetition, more time during judgment, shorter delays between encoding and judgment, and stronger motivation to process relational information. However, the two accounts fared less well in predicting the influence of these contextual factors on the effects of mere co-occurrence, which are the hallmark of unintentional influences in intentional impression formation. On the one hand, mere co-occurrence effects were unaffected by time for encoding, temporal delay, and intentional control. These results are consistent with the predictions of dual-process learning accounts and inconsistent with the predictions of single-process learning accounts. On the other hand, mere co-occurrence effects were unaffected by repetition and they increased with more time during judgment. These results are inconsistent with the predictions of both dual-process and single-process learning accounts. Although the latter findings raise important questions about the mental processes underlying mere co-occurrence effects, it is worth noting that they still provide valuable insights into the boundary conditions of unintentional influences in intentional impression formation, as reflected in dissociative effects of mere co-occurrence and relational information. Specifically, the available evidence suggests that unintentional influences in intentional impression formation are unaffected by time for encoding, repetition, temporal delay, and intentional control, but ironically increase with more time during judgment. An important task for future research is to investigate why these factors show the obtained effects, which could provide further insights into the processes underlying unintentional influences in intentional impression formation.

Implications for Social Impression Formation

Although extant theories are still facing empirical challenges in accounting for the moderators of unintentional influences in intentional impression formation, the phenomenon itself is supported by a solid body of evidence. While some of this research involves impressions of non-social objects (e.g., Gawronski & Brannon, 2021; Heycke & Gawronski, 2020; Hu et al., 2017), there is considerable evidence suggesting that unintentional influences can also occur for intentional impressions of social targets (e.g., Kukken et al., 2020; Moran & Bar-Anan, 2013; Moran et al., 2015). An interesting extension of the latter work is research on contrastive relations in social networks. Research on cognitive balance (Heider, 1958) suggests that interpersonal sentiments can influence social impressions in a manner similar to the relational information in the reviewed research. Whereas positive relations (e.g., liking someone, being liked by someone) have been found to influence social impressions in an assimilative manner, negative relations (e.g., disliking someone, being disliked by someone) tend to influence social impressions in a contrastive manner. For example, people tend to form positive impressions of individuals who are liked by a positively evaluated person and negative impressions of individuals who are liked by a negatively evaluated person. Conversely, people tend to form negative impressions of individuals who are disliked by a positively evaluated person and positive impressions of individuals who are disliked by a negatively evaluated person (e.g., Aronson & Cope, 1968; Gawronski et al., 2005; Langer et al., 2009). These findings raise the question of whether mere co-occurrence can influence social impressions when two individuals are known to have contrastive relations (e.g., they dislike each other).

Yet, counter to this idea, research using implicit and explicit measures suggests that relational information prevails over mere co-occurrence in impression formation based on social networks (e.g., Gawronski & Walther, 2008; Gawronski et al., 2005). Moreover, under conditions where mere co-occurrence has been found to influence responses on implicit measures, it also influenced responses on explicit measures with relational information being ineffective in influencing social impressions (e.g., Gawronski & Walther, 2008; Gawronski et al., 2005). These results suggest that unintentional influences of mere co-occurrence are unlikely to occur for intentional impressions of people based on their interpersonal relations in social networks.

That being said, all of this research has relied on a task-dissociation approach comparing responses on implicit and explicit measures. Considering that multinomial modeling has been found to be more sensitive in detecting mere co-occurrence effects that remain undetected by the task-dissociation approach, an interesting question for future research is whether multinomial modeling is also superior in detecting mere co-occurrence effects in impression formation based on social networks. We consider this question as an interesting direction for future research.[6]

An important theoretical insight of the reviewed research is the significance of distinguishing between (1) processes involved in the formation of mental representations and (2) processes involved in the behavioral expression of stored representations. Early domain-specific dual-process theories have been very precise about whether their assumptions refer to the formation of a mental representation or the effects of a stored representation on behavior. However, the distinction has become increasingly blurry in domain-independent dual-system theories (e.g., Epstein, 1994; Kahneman, 2003; Smith & DeCoster, 2000; Strack & Deutsch, 2004), which explain all social phenomena as the interactive product of two functionally distinct processing systems (for a discussion, see Gawronski, Luke, & Creighton, in press). In line with the rediscovered significance of distinguishing between the formation and behavioral expression of mental representations (e.g., Corneille & Stahl, 2019; De Houwer et al., 2020; Gawronski et al., 2017; Kurdi & Dunham, 2020; Mandelbaum, 2016; see also Ferguson et al., 2014), the reviewed debate on the processes underlying unintentional influences in intentional impression formation suggests that other research on social impressions might similarly benefit from drawing sharper distinctions between the two stages. An illustrative example is the modal approach in research on spontaneous social inferences, which is based on the assumption that spontaneous impressions can be identified by means of non-reactive measures that do not require intentional judgments of the focal targets. Examples of such non-reactive measures are cued recall tasks, recognition tasks, lexical-decision tasks, word-stem-completion tasks, and relearning tasks (see Uleman et al., 1996). However, in a strict sense, these non-reactive tasks ensure only the role of unintentional processes in the behavioral expression of stored impressions, but they do not ensure the role of unintentional processes in their formation. Thus, greater attention to the distinction between the formation and behavioral expression of mental representations may also provide more nuanced insights into the processes underlying spontaneous social impressions.

Conclusions

The current chapter reviewed evidence for unintentional influences in intentional impression formation, focusing particularly on the phenomenon that the mere co-occurrence of stimuli can influence evaluative responses in a manner that is diametrically opposite to intentionally formed impressions based on the relation between the co-occurring stimuli. This phenomenon is similar to spontaneous social inferences, in that it involves unintentional effects in impression formation. However, it is different from spontaneous social inferences, in that it arises in contexts where people do have the intention to form an impression. Moreover, while prior research on spontaneous inference has predominantly focused on impressions with specific semantic content, evidence for unintentional influences in intentional

impression formation primarily comes from studies on broad evaluative impressions. Although extant theories are facing some non-trivial challenges in accounting for the moderators of such unintentional influences, the phenomenon itself is supported by a considerable body of evidence in research using task-dissociation and formal modeling approaches. An important task for future research is to develop mental-process theories that explain not only the phenomenon itself, but also its (in)sensitivity to various contextual factors.

Notes

1 **Author's Note:** Preparation of this chapter was supported by National Science Foundation Grant #1649900. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

2 A notable exception to these modal trends is recent research on spontaneous evaluative inferences (e.g., Olcaysoy Okten et al., 2019; Schneid et al., 2015).

3 Following Heycke and Gawronski (2020), we use *R* for the parameter capturing effects of relational information, *C* for the parameter capturing effects of mere co-occurrence, and *B* for the parameter capturing general response biases. In a multinomial model that is structurally equivalent to the model in Figure 11.1, Kukken et al. (2020) used *m* instead of *R* (referring to meaning), *p* instead of *C* (referring to pairing), and *g* instead of *B* (referring to guessing).

4 Because multinomial modeling is based on binary responses with *p*(positive response) = 1 – *p*(negative response), there are only four non-redundant equations in the set of eight equations listed in the Appendix.

5 An alternative way to explain effects of mere co-occurrence and relational information from a single-process propositional view is to hypothesize that people generate and store two propositions for the same event, one capturing relational information (e.g., X prevents something negative) and one capturing co-occurrence information (e.g., X co-occurs with something negative). Expanding on this hypothesis, unintentional effects of mere co-occurrence despite intentional use of relational information can be explained with the additional assumption that mental propositions capturing co-occurrence information are generated and retrieved automatically. However, it is worth noting that such an explanation would make single-process propositional accounts empirically indistinguishable from accounts that propose two functionally distinct learning mechanisms, rendering the debate a matter of terminological preference rather than empirical evidence. While dual-process learning accounts explain mere co-occurrence effects in terms of automatic formation of associations between co-occurring stimuli, single-process propositional accounts endorsing the above assumptions would explain mere co-occurrence effects in terms of automatic processing of co-occurrence propositions.

6 An important caveat is that the standard model depicted in Figure 11.1 (see Heycke & Gawronski, 2020; Kukken et al., 2020) would have to be extended with an additional parameter capturing evaluative effects of interpersonal sentiments independent of the valence of the "co-occurring" person. Such an extension may be required, because being liked by someone has been found to lead to more favorable impressions than being disliked by someone, regardless of whether the (dis)liking person is evaluated positively or negatively (e.g., Gawronski et al., 2005). Similarly, liking someone has been found to lead to more favorable impressions than disliking someone regardless of whether the (dis)liked person is evaluated positively or negatively (e.g., Gawronski & Walther, 2008). These effects will have to be accounted for when applying a

multinomial modeling approach to studying effects of mere co-occurrence and relational information in impression formation based on social networks.

References

Aronson, E., & Cope, V. (1968). My enemy's enemy is my friend. *Journal of Personality and Social Psychology*, *8*, 8–12.

Bading, K., Stahl, C., & Rothermund, K. (2020). Why a standard IAT effect cannot provide evidence for association formation: The role of similarity construction. *Cognition and Emotion*, *34*, 128–143.

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, *6*, 57–86.

Corneille, O., & Mertens, G. (2020). Behavioral and physiological evidence challenges the automatic acquisition of evaluations. *Current Directions in Psychological Science*, *29*, 569–574.

Corneille, O., & Stahl, C. (2019). Associative attitude learning: A closer look at evidence and how it relates to attitude model. *Personality and Social Psychology Review*, *23*, 161–189.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671–684.

De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *The Spanish Journal of Psychology*, *10*, 230–241.

De Houwer, J. (2009). The propositional approach to associative learning as an alternative for association formation models. *Learning & Behavior*, *37*, 1–20.

De Houwer, J. (2018). Propositional models of evaluative conditioning. *Social Psychological Bulletin*, *13*(3), e28046.

De Houwer, J., Van Dessel, P., & Moran, T. (2020). Attitudes beyond associations: On the role of propositional representations in stimulus evaluation. *Advances in Experimental Social Psychology*, *61*, 127–183.

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist*, *49*, 709–724.

Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models: A review of the literature. *Zeitschrift für Psychologie/Journal of Psychology*, *217*, 108–124.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, *69*, 1013–1027.

Ferguson, M. J., Mann, T. C., & Wojnowicz, M. T. (2014). Rethinking duality: Criticisms and ways forward. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 578–594). New York: Guilford Press.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*, 692–731.

Gawronski, B., & Bodenhausen, G. V. (2007). Unraveling the processes underlying evaluation: Attitudes from the perspective of the APE Model. *Social Cognition*, *25*, 687–717.

Gawronski, B., & Bodenhausen, G. V. (2011). The associative-propositional evaluation model: Theory, evidence, and open questions. *Advances in Experimental Social Psychology*, *44*, 59–127.

Gawronski, B., & Bodenhausen, G. V. (2014). The associative-propositional evaluation model: Operating principles and operating conditions of evaluation. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 188–203). New York: Guilford Press.

Gawronski, B., & Bodenhausen, G. V. (2018). Evaluative conditioning from the perspective of the associative-propositional evaluation model. *Social Psychological Bulletin*, *13*(3), e28024.

Gawronski, B., & Brannon, S. M. (2021). Attitudinal effects of stimulus co-occurrence and stimulus relations: Range and limits of intentional control. *Personality and Social Psychology Bulletin*, *47*, 1654–1667.

Gawronski, B., Brannon, S. M., & Bodenhausen, G. V. (2017). The associative-propositional duality in the representation, formation, and expression of attitudes. In R. Deutsch, B. Gawronski, & W. Hofmann (Eds.), *Reflective and impulsive determinants of human behavior* (pp. 103–118). New York: Psychology Press.

Gawronski, B., & De Houwer, J. (2014). Implicit measures in social and personality psychology. In H. T. Reis, & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd edition, pp. 283–310). New York: Cambridge University Press.

Gawronski, B., Luke, D. M., & Creighton, L. A. (in press). Dual-process theories. In D. E. Carlston, K. Johnson, & K. Hugenberg (Eds.), *The Oxford handbook of social cognition* (2nd edition). New York: Oxford University Press.

Gawronski, B., & Walther, E. (2008). The TAR effect: When the ones who dislike become the ones who are disliked. *Personality and Social Psychology Bulletin*, *34*, 1276–1289.

Gawronski, B., Walther, E., & Blank, H. (2005). Cognitive consistency and the formation of interpersonal attitudes: Cognitive balance affects the encoding of social information. *Journal of Experimental Social Psychology*, *41*, 618–626.

Green, L. J. S., Luck, C. C., Gawronski, B., & Lipp, O. (2021). Contrast effects in backward evaluative conditioning: Exploring effects of affective relief/disappointment versus instructional information. *Emotion*, *21*, 350–359.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. K. L. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, *74*, 1464–1480.

Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.

Heycke, T., & Gawronski, B. (2020). Co-occurrence and relational information in evaluative learning: A multinomial modeling approach. *Journal of Experimental Psychology: General*, *149*, 104–124.

Hu, X., Gawronski, B., & Balas, R. (2017). Propositional versus dual-process accounts of evaluative conditioning: I. The effects of co-occurrence and relational information on implicit and explicit evaluations. *Personality and Social Psychology Bulletin*, *43*, 17–32.

Hütter, M., & Klauer, K. C. (2016). Applying processing trees in social psychology. *European Review of Social Psychology*, *27*, 116–159.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, *58*, 697–720.

Kukken, N., Hütter, M., & Holland, R. (2020). Are there two independent evaluative conditioning effects in relational paradigms? Dissociating effects of CS-US pairings and their meaning. *Cognition and Emotion*, *34*, 170–187.

Kurdi, B., & Dunham, Y. (2020). Propositional accounts of implicit evaluation: Taking stock and looking ahead. *Social Cognition*, *38*, s42–s67.

Langer, T., Walther, E., Gawronski, B., & Blank, H. (2009). When linking is stronger than thinking: Associative transfer of valence disrupts the emergence of cognitive balance after attitude change. *Journal of Experimental Social Psychology*, *45*, 1232–1237.

Mandelbaum, E. (2016). Attitude, inference, association: On the propositional structure of implicit bias. *Noûs*, *50*, 629–658.

Moran, T., & Bar-Anan, Y. (2013). The effect of object–valence relations on automatic evaluation. *Cognition and Emotion*, *27*, 743–752.

Moran, T., Bar-Anan, Y., & Nosek, B. (2015). Processing goals moderate the effect of co-occurrence on automatic evaluation. *Journal of Experimental Social Psychology*, *60*, 157–162.

Moskowitz, G. B., & Olcaysoy Okten, I. (2016). Spontaneous goal inference (SGI). *Social and Personality Psychology Compass*, *10*, 64–80.

Olcaysoy Okten, I., Schneid, E. D., & Moskowitz, G. B. (2019). On the updating of spontaneous impressions. *Journal of Personality and Social Psychology*, *117*, 1–25.

Payne, B. K., Burkley, M., & Stokes, M. B. (2008). Why do implicit and explicit attitude tests diverge? The role of structural fit. *Journal of Personality and Social Psychology*, *94*, 16–31.

Schneid, E. D., Carlston, D. E., & Skowronski, J. J. (2015). Spontaneous evaluative inferences and their relationship to spontaneous trait inferences. *Journal of Personality and Social Psychology*, *108*, 681–696.

Sherman, J. W., Krieglmeyer, R., & Calanchini, J. (2014). Process models require process measures. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual process theories of the social mind* (pp. 121–138). New York: Guilford Press.

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, *4*, 108–131.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, *8*, 220–247.

Uleman, J. S., Newman, L. S., & Moskowitz, G. B. (1996). People as flexible interpreters: Evidence and issues from spontaneous trait inference. *Advances in Experimental Social Psychology*, *28*, 211–279.

Van Dessel, P., Gawronski, B., & De Houwer, J. (2019). Does explaining social behavior require multiple memory systems? *Trends in Cognitive Sciences*, *23*, 368–369.

Zanon, R., De Houwer, J., & Gast, A. (2012). Context effects in evaluative conditioning of implicit evaluations. *Learning and Motivation*, *43*, 155–165.

Zanon, R., De Houwer, J., Gast, A., & Smith, C. T. (2014). When does relational information influence evaluative conditioning? *The Quarterly Journal of Experimental Psychology*, *67*, 2105–2122.

Appendix

Model equations for the estimation of effects of relational information (*R*), mere co-occurrence (C), and general response bias (*B*) on responses to objects that have an assimilative or a contrastive relation to a positive or a negative stimulus.

p(positive response | assimilative, positive) = $R + [(1 - R) \times C] + [(1 - R) \times (1 - C) \times B]$

p(positive response | assimilative, negative) = $(1 - R) \times (1 - C) \times B$

p(positive response | contrastive, positive) = $[(1 - R) \times C] + [(1 - R) \times (1 - C) \times B]$

p(positive response | contrastive, negative) = $R + [(1 - R) \times (1 - C) \times B]$

p(negative response | assimilative, positive) = $(1 - R) \times (1 - C) \times (1 - B)$

p(negative response | assimilative, negative) = $R + [(1 - R) \times C] + [(1 - R) \times (1 - C) \times (1 - B)]$

p(negative response | contrastive, positive) = $R + [(1 - R) \times (1 - C) \times (1 - B)]$

p(negative response | contrastive, negative) = $[(1 - R) \times C] + [(1 - R) \times (1 - C) \times (1 - B)]$

12 Stereotypes and Trait Inference

Jeffrey W. Sherman
University of California, Davis

For this chapter, I thought it would be fun to very briefly trace the influence of Jim Uleman's research on spontaneous trait inference to work in my own research career on stereotyping and social cognition more broadly. Although I have been influenced by much of Jim's work, there is one paper that stands out as particularly impactful in my own research life. I am referring to Winter and Uleman (1984), which demonstrated that people draw trait inferences from others' behavior spontaneously, without necessarily intending to or being aware of having done so. This, of course, is the spontaneous trait inference (STI) paper that launched a thousand research projects. When I began graduate school in 1989, this was one of the very first papers my advisor, Dave Hamilton, told me to read. Even five years after its publication, Dave considered this to be the absolute cutting edge of social cognition research, and he was right. It kind of blew my mind. Upon entering graduate school, I was not well versed in the burgeoning social cognition literature, and was just beginning to wrap my head around the methods that were being used to figure out what was going on in people's heads when they thought about other people. I found Winter and Uleman's (1984) adaptation of Tulving's encoding specificity approach especially clever. It almost seemed like a magic trick for reading people's minds. I became intensely interested in understanding what, when, how, and why we decide what other people (and, later, groups of people) are like. I pursued such questions in the context of deciphering the sources of self-knowledge, person perception, stereotyping, differences between individual and group perception, perceptions of group variability, stereotype formation, the processes surrounding the encoding and retrieval of expected and unexpected information, employee evaluations, and social role inferences. More broadly, Winter and Uleman (1984) was integral in kickstarting a career-long fascination with identifying the mechanisms of social cognition that reached a natural conclusion with an abiding interest in formal models designed to identify and measure the hidden processes that drive our judgments and evaluations of other people.

DOI: 10.4324/9781003045687-14

The Mental Representation of Social Knowledge

I can identify two broad research enterprises in my own work that owe a major debt of gratitude to Jim's work on trait inference. First, early in my career, I was engaged in a research program aimed at identifying whether people's judgments about the self and others are based on abstract mental representations that have been formed and stored in memory (e.g., trait inferences) versus specific pieces of information (e.g., episodic memory; category exemplars) that are retrieved at the time of judgment and summarized in order to make social judgments. Initially, this work was conducted with Stan Klein on the self-concept (Klein et al., 1993; Klein et al., 1996; Klein et al., 1997). Stan was interested in the fundamental nature of self-knowledge and whether judgments about the self require autobiographical memory. Could people know themselves without remembering their specific behaviors? Both philosophers and psychologists had long argued that such autobiographical memories were essential to the construction of self-knowledge.

The alternative is that people develop stable, semantic self-knowledge. That is, that people make inferences from their behavior about the traits that describe themselves and retain these inferences in memory. When judging themselves, rather than retrieving and summarizing autobiographical memories, they may simply access the stored trait inference. This work largely demonstrated that people need not access specific autobiographical memories in order to judge themselves. Moreover, the extent to which self-knowledge is independent of autobiographical memory is related to the amount of experience a person has with him or herself in a particular context. In novel contexts, in which people do not have much basis for self-knowledge, they rely on autobiographical memories. However, as they gain experience, they develop stable self-knowledge that is independent from autobiographical memory. In other words, over time and experience, people make inferences from their behavior about the stable traits that characterize them.

In subsequent work, we extended this analysis to knowledge of others (Sherman & Klein, 1994; see also Klein et al., 1992). In this case, the question was whether we can make judgments about other people without accessing specific memories of their behavior. As with self-knowledge, the answer is that it depends on the extent of experience one has with another person. Early on, as we are just getting to know others, our judgments about them involve the retrieval of specific biographical behaviors. However, as we become more familiar with them, we extract trait inferences that may be accessed independently of the specific behaviors upon which they were based. We also showed that, when exposed to relatively extreme behaviors that strongly exemplified a particular trait, this process occurred more rapidly. That is, when a person engages in highly diagnostic behavior, we make trait inferences very quickly.

Obviously, these ideas share much in common with Jim's work on STIs. Yet, they are distinct in important ways. First, whereas work on STIs tests

whether or not a trait is inferred, our work tested whether judgments about traits are based on the retrieval of specific behaviors. If judgments are not based on specific behaviors, we assume that they are based on already formed and stored trait inferences—they must be based on something. Note that the use of specific behaviors doesn't mean that a trait has not already been inferred and stored. It simply means that respondents are not content to rely solely on existing trait knowledge, perhaps due to a lack of confidence in the inference. Also note that, in both cases, judgments are based on trait inferences. In one case, the inferences have been made and stored in memory. In the other, the inference is based on the trait implications of the retrieved behaviors. Second, the extent to which the inferences in our work are made spontaneously or possess other features of automaticity is unclear. Subjects are asked to form impressions of the target, though they are not informed ahead of time that they will be asked about particular traits.

Finally, we expanded these ideas into the study of stereotype formation and group knowledge. With novel groups, for which perceivers do not possess pre-existing stereotypes, the results mirrored those for the self and for individual others (Sherman, 1996). Namely, at low levels of experience, judgments of the group involved the retrieval of specific behaviors performed by individual group members. However, as knowledge of the group increased, an abstract trait impression of the group (i.e., a stereotype) was created that formed the basis for group judgments, independent of memory for specific behaviors. In another study, I asked the same question about groups that were known to participants and for which they possessed pre-existing stereotypes (e.g., engineers). In this case, judgments about stereotype-relevant traits never involved the retrieval of group behaviors. Even when little was known about the specific group (of engineers), participants did not need to refer to specific group behaviors in order to judge the group. Rather, it seemed that the stereotype provided ready-made trait knowledge that permitted immediate inference, even in the absence of direct knowledge about the group in question. Thus, merely categorizing a person as a member of a stereotyped group invokes existing stereotypes about the group that are stored in memory and which provide ready-made inferences about stereotype-relevant traits. At the same time, judgments about non-stereotypic traits did invoke the retrieval of specific group behaviors. Thus, the stereotype permitted inferences only about stereotype-relevant traits.

In subsequent research, we examined how intergroup motivations influenced the development of group stereotypes (Sherman et al., 1998). In this case, via a minimal group manipulation, participants were assigned to an arbitrary group. Subsequently, they learned either positive or negative information about either their own group or an outgroup to which they did not belong. The results showed that the rate of trait inference (i.e., stereotype formation) varied as a function of trait valence and group membership. For positive attributes, participants retrieved specific behaviors to make judgments about the outgroup but not the ingroup. In contrast, for negative

behaviors, they retrieved behaviors to make judgments about the ingroup but not the outgroup. Thus, trait inferences were made in accordance with intergroup motives. Positive stereotypes of ingroups and negative stereotypes of outgroups developed quickly and judgments along these traits were made independent of specific group memories. In contrast, negative stereotypes of ingroups and positive stereotypes of outgroups developed slowly and judgments along these traits required the retrieval of specific behaviors.

Stereotype Efficiency and Encoding Flexibility

Our work on mental representation fed directly into the second line of research that builds on Jim's trait inference work. One of the conclusions from my studies on stereotype formation (Sherman, 1996) is that, once a group stereotype exists, it provides relevant trait inferences that no longer need be inferred from group members' behavior. This meaning supplying function of stereotypes is central to the view of stereotypes as judgmental heuristics that help to simplify the world and make social cognition more efficient (Hamilton & Sherman, 1994). Related research on stereotype efficiency focused not on the inference process but on how stereotypes direct our attention toward different kinds of information and how that affects our subsequent memory for that information. Though not directly focused on trait inference, per se, the inference process formed the theoretical basis and explanation of key results. In this work, stereotypes were seen as information filters that efficiently directed attention toward certain kinds of information and away from others, thus reducing overall cognitive load (for a review, see Sherman et al., 1998).

Specifically, according to this view, stereotypes are thought to direct attention toward others' stereotype-consistent behavior and away from stereotype-irrelevant and stereotype-inconsistent information. The logic is that, because behavior that fits stereotypic expectancies is easier to understand (i.e., it is easier to infer the trait meaning), stereotypes make social perception efficient by directing attention toward that information and away from information, such as stereotype-inconsistent behavior, that requires more cognitive resources to understand and integrate. This results in stereotype confirmation and subsequent superior memory for stereotypic behavior. Because the need for efficient processing is magnified under cognitive load, these processes were thought to be more prevalent in those circumstances. For example, subjects who were distracted by an irrelevant newscast when learning about a target person subsequently recalled more stereotypic than counter-stereotypic information about the person (Stangor & Duan, 1991).

My own reading of the literature led me to propose a different interpretation of the data and a new model for understanding how stereotypes affect the processing of stereotype-relevant information. At the heart of this analysis, again, is the trait inference process. As for the data, they were more

complex and nuanced than had generally been recognized. Although free recall favored stereotype-consistent over-inconsistent behaviors, particularly when encoded under cognitive load, recognition memory showed the opposite pattern—better memory for stereotype-inconsistent behavior, particularly under cognitive load (Stangor & McMillan, 1992). Free recall reflects not only attention and encoding, but retrieval advantages for expected (versus unexpected) information and response biases that lead people to set a lower threshold for reporting stereotype-consistent than-inconsistent behavior. Thus, greater recall of stereotype-consistent behavior is not clear evidence for an attentional filtering mechanism that favors that information. In contrast, recognition memory controls for retrieval and response biases by presenting the to-be-remembered behaviors to participants when memory is tested. As such, recognition performance is a much clearer index of attention and encoding effort than is free recall. Thus, the fact that recognition memory favors stereotype-inconsistent information, particularly when encoded under cognitive load, argues against the suggestion that stereotypes focus attention on consistent information and filter out inconsistent information.

Theoretical considerations further argue against a filter model. Given that stereotypes facilitate the processing of information that confirms the stereotype, it is not clear why extra attention would be devoted to that information. Because they confirm what is expected, the trait meaning of those behaviors may be easily inferred and, indeed, the trait impression of the actor may be inferred directly from the stereotype without attending to the behavior at all. This was one of the conclusions from my earlier work (Sherman, 1996). In my view, it made much more sense for attention to be directed toward information that cannot simply be inferred from a stereotype. In an efficient system, this should be particularly true when under cognitive load and processing resources are scarce. We called this model the Encoding Flexibility Model (Sherman et al., 1998) and supported its primary predications across many experiments. Specifically, we showed that people pay more attention to and better encode the perceptual and contextual details of stereotype-inconsistent than-consistent information, particularly under cognitive load. For example, using a dot probe technique, we showed that participants learning about a target person while under a cognitive load (rehearsing an eight-digit number) attended more carefully to stereotype-inconsistent than-consistent behaviors (Sherman et al., 1998). In particular, reactions to dot probes were faster when they appeared during the presentation of stereotype-inconsistent than-consistent behaviors, particularly when subjects were under cognitive load. This shows that those participants were attending more carefully to the stereotype-inconsistent than-consistent behaviors.

At the same time, people are better able to extract the conceptual (trait) meaning of consistent than inconsistent behavior (Allen et al., 2009; Sherman & Frost, 2000; Sherman et al., 1998; Sherman et al., 2004). For example, subjects who learned about a target person while under a cognitive

load were subsequently better able to accurately identify traits implied by stereotype-consistent than -inconsistent behaviors when those traits were flashed very quickly (33 ms; Sherman et al., 1998). This shows that, when under cognitive load, perceivers are more likely to infer the trait meanings of stereotype-consistent than-inconsistent behaviors, which are subsequently more accessible. Coming full circle, we (Wigboldus et al., 2004) demonstrated this latter effect most directly in a study on how stereotypes affect spontaneous trait inferences for stereotype-consistent and\-inconsistent behavior using a variant of the trait probe method pioneered by Winter and Uleman (1984; Uleman et al., 1996). Specifically, when under cognitive load, subjects required more time to accurately judge that traits only implied by stereotype-consistent behaviors that hadn't been explicitly presented than it took to make the same judgment about traits implied by stereotype-inconsistent behaviors. This demonstrates that subjects were more likely to spontaneously make trait inferences about stereotype-consistent than-inconsistent behavior, particularly when under cognitive load.

Summary

To summarize, questions about trait inference have been central to my research, and it was Uleman's work on spontaneous trait inference that ignited my interest in the topic. In one line of work, I studied when trait inferences occur and how they interact with and promote independence from (auto) biographical memory in social judgment. In another line of work, I examined how the trait inference process is informed by stereotypes and how that, in turn, influences the encoding of stereotype-relevant behavior, particularly in conditions that demand efficient social cognition.

I would like to conclude with a few personal observations about Jim and his influence on my professional life. Beyond the obvious influence of his research, Jim had a major impact on my socialization and sense of belonging in the guild of social psychology. As I'm sure is true for many social cognition researchers, Jim was the first big shot (other than my advisor) who seemed to take a genuine interest in me—not just as a researcher, but also as a person. In my case, this occurred at a Person Memory Interest Group conference, where Jim has long been a fixture. Unassuming, welcoming, and funny, I could not believe that this was Jim Uleman! The guy whose work had put a charge into my early life as a social cognition researcher? Whose work seemed impossibly sophisticated and precise, theoretically and methodologically? If not for Jim being Jim, I would have been intimidated, as I was in the presence of other big shots. Jim simply would not permit that. He approached *me*, asked about my research, and made me feel welcome and at ease. He remembered who I was and what I did. He helped me feel like maybe I belonged. I was but one of countless young social psychologists to whom Jim extended such kindness. For these reasons alone, I would be proud to contribute to this volume, and I am grateful to the editors for providing me with

an opportunity to express my admiration and appreciation of Jim as an exceptional scientist and human being.

References

Allen, T. J., Sherman, J. W., Conrey, F. R., & Stroessner, S. J. (2009). Stereotype strength and attentional bias: Preference for confirming versus disconfirming information depends on processing capacity. *Journal of Experimental Social Psychology*, *45*, 1081–1087.

Hamilton, D. L., & Sherman, J. W. (1994). Stereotypes. In R. S. Wyer, Jr. & T. K. Srull (Eds.), *Handbook of social cognition* (2nd Ed., Vol. 2, pp. 1–68). Hillsdale, NJ: Erlbaum.

Klein, S. B., Babey, S. H., & Sherman, J. W. (1997). The functional independence of trait and behavioral self-knowledge: Methodological considerations and new empirical findings. *Social Cognition*, *15*, 183–203.

Klein, S. B., Loftus, J., & Sherman, J. W. (1993). The role of summary and specific behavioral memories in trait judgments about the self. *Personality and Social Psychology Bulletin*, *19*, 305–311.

Klein, S. B., Loftus, J., Trafton, J. G., & Fuhrman, R. W. (1992). Use of exemplars and abstractions in trait judgments: A model of trait knowledge about the self and others. *Journal of Personality and Social Psychology*, *63*, 739–753.

Klein, S. B., Sherman, J. W., & Loftus, J. (1996). The role of episodic and semantic memory in the development of trait self-knowledge. *Social Cognition*, *14*, 277–291.

Sherman, J. W. (1996). Development and mental representation of stereotypes. *Journal of Personality and Social Psychology*, *70*, 1126–1141.

Sherman, J. W., Conrey, F. R., & Groom, C. J. (2004). Encoding flexibility revisited: Evidence for enhanced encoding of stereotype-inconsistent information under cognitive load. *Social Cognition*, *22*, 214–232.

Sherman, J. W., & Frost, L. A. (2000). On the encoding of stereotype-relevant information under cognitive load. *Personality and Social Psychology Bulletin*, *26*, 26–34.

Sherman, J. W., & Klein, S. B. (1994). The development and representation of personality impressions. *Journal of Personality and Social Psychology*, *67*, 972–983.

Sherman, J. W., Klein, S. B., Laskey, A., & Wyer, N. A. (1998). Intergroup bias in group judgment processes: The role of behavioral memories. *Journal of Experimental Social Psychology*, *34*, 51–65.

Sherman, J. W., Lee, A. Y., Bessenoff, G. R., & Frost, L. A. (1998). Stereotype efficiency reconsidered: Encoding flexibility under cognitive load. *Journal of Personality and Social Psychology*, *75*, 589–606.

Stangor, C., & Duan, C. (1991). Effects of multiple task demands upon memory for information about social groups. *Journal of Experimental Social Psychology*, *27*, 357–378.

Stangor, C., & McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: A review of the social and social developmentalliteratures. *Psychological Bulletin*, *111*, 42–61.

Uleman, J. S., Hon, A., Roman, R. J., & Moskowitz, G. B. (1996). On-line evidence for spontaneous trait inferences at encoding. *Personality and Social Psychology Bulletin*, *22*, 377–394.

Wigboldus, D. H. J., Sherman, J. W., Franzese, H. L., & van Knippenberg, A. (2004). Capacity and comprehension: Spontaneous stereotyping under cognitive load. *Social Cognition*, *22*, 292–309.

Winter, L., & Uleman, J. S. (1984). When are social judgments made? Evidence for the spontaneousness of trait inferences. *Journal of Personality and Social Psychology*, *47*, 237–252.

13 Perceiving Group Attributes Spontaneously: Broadening the Domain

David L. Hamilton[1] *and Joel A. Thurston*[2]
[1]*University of California, Santa Barbara*
[2]*University of Virginia, Biocomplexity Institute and Initiative*

In a chapter published in 1935 the famous psychologist Gordon Allport (1935) made the following observation:

> The concept of attitude is probably the most distinctive and indispensable concept in contemporary American social psychology. (p. 798)

> Without guiding attitudes the individual is confused and baffled… Attitudes determine for each individual what he will see and hear, what he will think and what he will do…they draw lines about and segregate an otherwise chaotic environment; they are our methods for finding our way about in an ambiguous universe. (p. 806)

As Allport states, attitudes are very influential in guiding and shaping people's thoughts, feelings and behaviors in many domains. They are pervasive in all aspects of life. The concept of attitude was a central concept in social psychology in 1935. It still is today.

Of course, attitude is not the only concept that is of great importance in social psychology's analysis of how people negotiate and adapt to the complexities of social life. In this chapter we argue that throughout the history of social psychology the concept of inference has proven to be incredibly important in people's cognitive functioning. Like attitudes, inferences are pervasive in people's everyday lives and, like attitude, the concept of inference is pervasive in the social psychological literature.

In their everyday lives people continually encounter events, other persons, and groups, and those experiences are the raw data that then can be processed in various ways by the individual. Inference involves going beyond those raw data, elaborating on them to broaden understanding of what has been encountered. It does so by drawing on knowledge and beliefs acquired from past experience to gain further understanding of the information being encoded, interpreted, and evaluated. An important benefit of expanding knowledge in this way is that it enhances one's ability to anticipate future occurrences. Inference, then, is a crucial process in comprehending the social world.

DOI: 10.4324/9781003045687-15

Important Role of Inferences in Social Psychology: Early Research

The study of inferences has been pervasive throughout the history of social psychology. In fact, some of the earliest empirical studies that laid the groundwork for important areas of research relied entirely on inference data to demonstrate their phenomena. Consider the following examples.

Stereotypes

Although not the first attempts to measure stereotypes (see Schneider, 2004) the studies published by Katz and Braly (1933, 1935) became the catalyst for future research, in part because they introduced a simple yet effective means of capturing and measuring stereotypes. They presented participants with a list of 84 traits and, for each of ten target groups (e.g., Germans, Italians, Negroes, Jews, Americans) participants were to indicate which traits were descriptive of each group. Those traits most commonly checked for a group were considered to define the stereotype of that group. Notice that the method directly asks participants to make trait inferences: "Given the named group (Italians), what traits do you infer to be characteristic of the group?" This became the standard way of measuring stereotypes for several decades (Brigham, 1971).

Impressions

Solomon Asch (1946) is generally acknowledged as having demonstrated that first impressions is a researchable topic. Obviously people form impressions all the time – of strangers, acquaintances, friends, family members – impressions that are perhaps tentative at first and they may be reinforced, strengthened, modified, or rejected based on the acquisition of new information. Nevertheless those first impressions are clearly established quickly and can influence our approach to or avoidance of persons from the very outset. We all routinely engage in impression formation, often without even thinking about it. Given the complexities of this life-long process, studying it empirically might seem a daunting task. Asch realized such research would require a means of measuring those impressions. Therefore, in his classic research he presented participants a half-dozen or so trait terms describing a person with the instruction to form an impression of that person. He then gave people a list of trait adjectives and asked them to indicate which terms also described the person. "If a person is intelligent, skillful, warm, determined, and practical, what other traits would describe him?" Thus was born the measure of impressions used in countless studies for several decades. Again, it is a trait inference task.

Implicit Personality Theories

The notion that people carry in their heads intuitive ideas about the nature of the personalities of the people they meet and interact with, and that those ideas can influence their perceptions of and interactions with those persons, may seem implausible. But that was exactly what Bruner and Taguiri (1954) proposed – a conception about the role of cognitive representations that was clearly out of step with the behaviorist tradition that then dominated academic theorizing in psychology. Bruner and Taguiri coined the term implicit personality theory to capture this notion, and it spawned considerable research (Schneider, 1973). The research focused on people's ideas about "what goes with what" (i.e., what attributes are correlated with what other attributes) in people's personalities. The means to measure these relations posed a challenge and several strategies were developed (see Schneider, 2004, Chapter 5). One of the earliest (and perhaps the simplest) method was the trait inference task in which participants made judgments of the co-occurrence of traits in general ("If a person is honest then how likely is she also intelligent?"). Many such questions can be answered in a short time and with ease, and they provide a direct measure of "what goes with what" in the person's belief system. Again, an intriguing new question was made manageable through the use of trait inferences.

Correspondent Inferences

Among the more important theoretical contributions in social psychology was Correspondent Inference Theory (Jones & Davis, 1965). The theory was particularly concerned with how perceivers move from observing behaviors to inferring psychological attributes or states – moving, as the title of their theory expressed, "from acts to dispositions." It was especially concerned with cases in which the inferred disposition corresponds directly to the manifest properties of the observed behavior, thereby called correspondent inferences. The theory was originally intended as a framework for understanding the nature and dynamics of attributions and causal judgments (Heider, 1958), but over time the research it inspired became increasingly focused on correspondent inferences and the conditions under which they are and are not made (as a function of the normativeness, desirability, and choice of the behavior in question). Therefore the theory became more a theory of the inference process (Hamilton, 1998). The research it generated, which was considerable (Jones, 1979, 1990), has been highly informative not only about the questions posed above but also about biases in this process (Gilbert & Malone, 1995; Jones & Harris, 1967; McArthur, 1972; Uleman et al., 1996), inferences about an individual's characteristics (such as goals) aside from traits (e.g., Aarts et al., 2004; Moskowitz & Olcaysoy Okten 2016; Read et al., 1990), and conditions fostering noncorrespondent dispositional inferences (Fein, 1996; Fein et al., 1990; Hilton et al., 1993).

In this brief historical overview we have argued that the inference process has been a focal point in social psychological research since the earliest days of the discipline. Foundational studies in several topic areas (stereotypes, impressions, implicit personality theories) studied the inferences made by participants to measure and document the parameters of their phenomena. Correspondent Inference Theory documented the pervasiveness of inferences in social perception, demonstrating a general tendency for social perceivers to infer dispositional qualities from the behaviors they observe, even when the information available would suggest that such inferences are not warranted.

In much of this work one gains the sense that perceivers attach great importance to the process by which they are making inferences. The checklist methods for measuring inferences about groups and persons (in assessing stereotypes and impressions) requires that participants consider the appropriateness of each trait for characterizing the target. Does this suggest that participants review and weigh the information they have (or believe) about the group or person in question before making their judgments? Correspondent Inference Theory stated that a correspondent inference will most likely be made when the behavior is not constrained by social norms, is undesirable, and when the actor has free choice. Does this imply that each time an observer notices someone's behavior she quickly assesses those three points before making an inference? From this perspective these seem to be thought-provoking and time-consuming tasks. But perceivers have busy lives and often don't have time and cognitive resources to consider the implications of each behavior they observe. There must be some alternative for understanding this process. Fortunately, things changed dramatically in 1984.

Spontaneous Trait Inferences (STIs)

As anyone reading the chapters in this volume is well aware, the now-classic article by Winter and Uleman (1984) introduced an entirely different perspective on the process of making trait inferences, specifically, that they are made spontaneously, unintentionally, and without conscious awareness of making them. Participants, who believed they were in a study on memory for verbal material, read a series of sentences, each one describing a behavior performed by a person. Later their memory for the information was assessed in a cued recall task. If the cue word was a trait implied by the behavior it increased the likelihood of recall. Winter and Uleman argued that, when reading the sentences, participants spontaneously inferred a trait implied by the behavior as characterizing the actor. Having made that inference during encoding, the trait word became associated with the sentence as it was stored in memory, and hence the trait could serve as a useful retrieval cue to aid recalling the sentence.

This radically different view ("What?? People make trait inferences about others without even knowing they're doing it?") was intriguing and was greeted with much interest among social psychologists. And for some readers,

it was greeted with skepticism. Some writers questioned the conclusions drawn from the cued recall method (Bassili, 1989; Bassili & Smith, 1986; D'Agostino, 1991; D'Agostino & Beegle, 1996). This led several investigators to create new methodologies for testing spontaneous inferences. These included the probe recognition method (Newman, 1991; Uleman et al., 1996), savings in relearning (Carlston & Skowronski, 1994; Carlston & Skowronski, 2005; Carlston et al., 1995), the false recognition paradigm (Todorov & Uleman, 2002, 2003, 2004), and the modified free association paradigm (Orghian et al., 2017). These new methodologies provided multiple opportunities to test for STIs. Other scholars were concerned whether a spontaneous process was the same or different from an automatic process (Bassili, 1989; Bassili & Smith, 1986; D'Agostino, 1991; D'Agostino & Beegle, 1996; Uleman et al., 1992; Winter et al., 1985). Winter and Uleman's (1984) article followed by only a few years the dramatic introduction into social psychology of the notion of automatic processes (Bargh, 1982, 1984; Bargh & Pietromonaco, 1982). Spontaneous trait inferences seemed to meet some of the important criteria ("the four horsemen;" Bargh, 1984) for automaticity (unintentional, outside of awareness) but less so for others (efficiency). In particular, there was debate as to whether STIs are influenced by cognitive load and by processing goals (e.g., impression vs. memory instructions) (Bassili & Smith, 1986; D'Agostino, 1991; Ferreira et al., 2012; Uleman, 1989, 1999; Uleman & Moskowitz, 1994; Wyer & Lambert, 1994).

All of these issues stimulated a considerable amount of research (see Uleman et al., 2008, for a review of STIs and related phenomena). Perhaps the simplest conclusions at this point are the following. First, STIs are highly robust. They occur spontaneously, without intention, and without awareness. They have been extensively documented and are manifested in studies using several different paradigms. Second, although occurring spontaneously and outside of awareness, they do not have all the properties of automaticity. In particular, their occurrence can sometimes be modified by both processing goals (comparing, for example, goal conditions that do (impression formation) and do not (memory) explicitly call for making inferences) and cognitive load (e.g., when simultaneously searching for the letter "t" in stimulus sentences). However, it is important to note that from the beginning (Winter & Uleman, 1984), in order to demonstrate STIs under conditions when participants are *not* consciously trying to form an impression, experiments have typically been introduced to participants as a study of memory. Therefore, although these processing goals can, under some conditions, influence the magnitude of STI effects, STIs typically persist despite these other influences (Uleman, 1989, 1999).

A related question concerns the downstream effects or consequences of STIs. These inferences are made without prior intention and without conscious awareness. Given these properties, of what importance are they? Interest in STIs was drawn by the fact that they represent an initial step in

forming first impressions but done so spontaneously. Moskowitz and Roman (1992) were the first to explore their downstream consequences and found that STIs functioned as a form of priming (e.g., Higgins et al., 1977). Inferences made spontaneously by the perceiver in response to one set of social stimuli made those traits accessible and therefore able to serve as implicit primes that then shaped conscious judgments toward a new social target. Of course, STIs should also influence judgments made about the same target since they represent a stable quality of the person that should persist over time. If so, then they may be a basis for anticipating a person's future behavior. McCarthy and Skowronski (2011) showed participants photos of different persons, each one paired with a behavior that implied a trait. Later they were shown the same photos, this time with a list of behaviors, one of which matched the trait implied by the person's own behavior. Participants' task was to indicate which behavior each stimulus person would perform. The behaviors selected corresponded to the trait implied by the actor's first behavior. Similarly, Olcaysoy Okten et al. (2019) showed such predictive utility of STIs. In their work initial STI formation was supplemented by learning a new behavior that was either congruent or incongruent with the STI. As expected, predictions of future behavior that corresponded with the STI were strongest when congruent information had fortified the initial inference. Thus, inferred traits can guide the anticipation of the actor's future behaviors.

Spontaneous Inferences: Expanding the Domain

Inferences go beyond the literal information available as a part of comprehending people's behaviors. They draw on the knowledge stored in memory, from past experience, to infer what else might be true (or assumed) about the target person. They therefore elaborate on what is known and flesh out the initial and developing impression. All of this occurs as part of comprehending the stimulus experience and provide a basis for anticipating future recurrences. In fact, because they happen immediately during encoding, STIs can be viewed as the initial elements in newly-forming impressions.

The literature cited in the previous section documents that inferences can occur spontaneously, unintentionally, and often without awareness of their occurrence. Interestingly, all of that research followed Winter and Uleman's lead in studying spontaneous *trait* inferences. Yet as we comprehend the information learned from observing others' behaviors, we seek to know more than simply the traits that characterize them. We seek to know why they are doing what they're doing, what goals they are pursuing, and what values they possess that guide their choices. We also have evaluative reactions to the person based on their behaviors, and we recognize role and situational constraints that may influence their behavior. The question then arises whether these additional aspects of comprehending a person can also occur spontaneously. The answer is Yes (see Schneid et al. chapter in this volume).

Knowing a person's *goals* is useful in adapting to the person's behavior. Research has investigated whether perceivers spontaneously infer an actor's goals as they process and comprehend information about him or her (Moskowitz & Olcaysoy Okten, 2016). Several studies have provided evidence of spontaneous goal inferences (SGIs) and that they are distinct from STIs (Hassin et al., 2005; Olcaysoy Okten & Moskowitz, 2018, 2020). Traits differ from goals in important respects. Traits are abstract concepts pertaining to general patterns of behavior, whereas goals are end states specific to certain tasks and situations. These differences influence the relative occurrence of STIs and SGIs (Olcaysoy Okten & Moskowitz, 2018, 2020; Skitka et al., 2002).

Observers typically have *evaluative reactions* in their perceptions of others' behaviors. Studies using paradigms for detecting spontaneous inferences have been adapted to detecting spontaneous evaluative inferences (SEIs) and have documented both their occurrence and their distinct properties from STIs (Olcaysoy Okten et al., 2019; Schneid et al., 2015). In addition, Ham and van den Bos (2008) reported evidence of spontaneous inferences about whether behavior is *just and fair* and that these spontaneous judgments differ from explicit judgments of the same situations. Moreover, these inferences are attuned to properties of the fairness context and to its personal relevance. Finally, behaviors always occur in a specific social context, and the context may moderate or alter the meaning of the behavior for comprehension. Research has shown that observers are sensitive to these *contextual constraints*, spontaneously make inferences based on the properties of the situation, and can make trait and situational inferences simultaneously (Ham & Vonk, 2003; Lupfer et al., 1990; Ramos et al., 2012; Todd et al., 2011).

People's behaviors are routinely both guided by and constrained by the *social roles* they are in. For example, occupations prescribe appropriate behaviors for job fulfillment. Policemen, pediatricians, and professors have roles that require different kinds of behaviors and people know the behaviors prescribed by those roles. Will an observer of those behaviors spontaneously infer that these behaviors are constrained by role fulfillment needs? Will they make spontaneous role inferences (SRIs)? Or will those role constraints not be recognized and instead observers make STIs from the person's behavior? Chen et al. (2014) presented role-implying sentences in which persons performed behaviors consistent with a role. Using two different STI paradigms (probe recognition, savings in relearning), participants were then tested for whether SRIs were made. Results showed that role-consistent behaviors did generate SRIs and, on subsequent judgments, actors were rated higher on traits consistent with those social roles.

Spontaneous trait inferences even extend beyond the person whose behavior inspires the trait inference. Suppose Evan mentions to you that Ken moved some heavy boxes for an elderly neighbor. The STI would be that you infer that Ken is helpful. Beyond that, however, studies have shown that you will also infer that Evan is helpful, even though you've learned nothing about his behavior. This *spontaneous trait transference* (STT) happens when the

inferred trait based on the behavior of the actor is transferred to the communicator (Crawford, Skowronski, & Stiff, 2007; Mae et al., 1999; Skowronski et al, 1998). Whereas STIs are made from first-hand observation of behavior, STTs are based on second-hand information communicated by another person. Both STIs and STTs are highly replicable (see Skowronski and McCarty chapter in this volume), though STTs typically are not as strong as STIs. Authors disagree on whether they are based on the same or different processes. Some (Carlston & Skowronski, 2005) have argued that they reflect different processes (STIs involve an attributional process, STTs are based on simple associations) whereas others (Brown & Bassili, 2002; Orghian et al., 2015) view both phenomena as based on associative processes. Nevertheless, both effects are robust, reflect spontaneous and involuntary processes, occur outside of awareness, and contribute to the formation of new impressions.

The extent and breadth of spontaneous inference processes demonstrated in this body of research is impressive. It is interesting to note, however, that all of this work has focused on inferences based on the behavior of individual persons. As observers, we often see behaviors performed by groups as well and certainly we make inferences, form impressions, and develop stereotypes about those groups. The question then becomes, do observers make spontaneous inferences about groups as they engage in these processes?

Spontaneous Inferences about Groups

Given the extensive body of research on spontaneous inferences about individuals, it is surprising to realize that there has been relatively little research investigating spontaneous processes in perceptions of groups. Groups are comprised of individual group members, and the evidence we have reviewed shows that observers spontaneously infer attributes from a person's behaviors. Once made, do those inferences about individual group members have implications for the perceiver's impressions of a group? If so, does the nature of the group influence that process? And would this process influence stereotypes of the group? Do people spontaneously make inferences that serve to differentiate groups? These are a few of the many important questions one could ask regarding inferences about groups, yet it is only recently that investigators have begun to pursue those questions. The remainder of this chapter focuses on those issues.

STIs and Group Impressions

Observers spontaneously make inferences about the people they see from the behaviors they observe. Crawford et al. (2002) explored the extent to which the attributes inferred, when made about group members, would generalize to other members of the group. If that happens, then members of the group would be perceived as having the same or similar attributes, even if those attributes were directly inferred only about one of the group's members. In

this way members of the group would come to be perceived as being similar to each other to a greater extent than warranted by the information acquired about them. Exaggerated perceptions of similarity can be a precursor to the development of group stereotypes.

Groups differ in many ways. One dimension underlying those differences is the tightness or coherence of the groups. Some are tight-knit groups in which members share many attributes and goals, interact a lot, and have shared goals and outcomes. In other groups the members are more loosely connected, spend less time together, and have differing objectives and outcomes. These are differences in the perceived *entitativity* or "groupness" of groups, a property that has been extensively researched (Brewer, 2015; Brewer & Harasty, 1996; Hamilton et al., 2002; Hamilton et al., 2004; Hamilton et al., 2011). Crawford et al. (2002) predicted that the type of generalization described above would most prominently occur in high entitativity groups.

In their study, using the savings in relearning paradigm, participants read about members of two different groups whose behaviors implied different traits. In a later phase the same stimulus persons were shown again, paired with a trait word. Sometimes those traits were implied by the person's behavior, other times the trait had been implied by the behavior of a different member of the same group. In a third phase the stimulus faces were shown again and the participants' task was to recall which trait had been paired with that person in the previous phase. Results showed that participants made STIs about the group members. However, transferring an attribute to another group member (recalling incorrectly that a trait had been paired with a different group member) occurred only in high entitativity groups. In these cases, then, traits inferred from the behaviors of some group members were transferred or generalized to other members of the same group. In this way, members of high entitativity groups became similar to, and interchangeable with, other members of the same (but not a different) group. This perceived interchangeability can be a foundation for stereotyping. A follow-up study showed that this generalization occurred to other members of the same group but not to members of the other group. Thus it produced within-group homogeneity while maintaining intergroup differences.

Crawford et al.'s (2002) finding of the spontaneous transfer of attributes from one group member to other group members is important not only because it fosters preconditions for stereotype formation but also because it occurs without the participants' intention or awareness. Other research using other (non-spontaneous inference) paradigms has shown similar generalization across group members (see Bray and Zarate chapter in this volume). This research suggests that such generalization can increase in magnitude with the passage of time and is particularly likely to occur for negatively-valued attributes (Enge et al., 2015; Lupo & Zarate, 2019). We will return to the issue of generalization in a different context in the next section.

Spontaneous Trait Inferences about Groups (STIGs)

Crawford et al.'s (2002) research documented that group impressions can develop on the basis of spontaneous inferences made about individual group members – without intention, without conscious awareness of this process. The stimulus materials presented in that research were trait-implying behaviors performed by individual members of groups, and those traits were spontaneously inferred about those individuals. Under certain conditions (when the group was known to be high in entitativity) the traits inferred about some group members were transferred to other members of that group, members whose behavior had not implied those traits. Thus a group impression emerged from the generalization of the inferred attributes of individual group members.

More recent research (Hamilton et al., 2015) has extended this analysis by investigating whether perceivers make spontaneous inferences from observing a group's behavior. Like individuals, groups enact behaviors and in this case the object of perception is the group as an entity, not as a collection of individual persons. A men's club may spend a Saturday working to restore the equipment at a children's park. A company's employees may go on strike and demonstrate against their company. In these cases the unit whose behavior is being observed is not an individual person or even a collection of persons. It is a group as an entity. Do observers spontaneously infer that the men's club is generous? Or that the employees are hostile or aggressive? If they did, their spontaneous inferences would be about the group, based on the group's behavior. Hamilton et al.'s (2015) research has demonstrated that people do in fact make such *spontaneous trait inferences about groups (STIGs)* and that they have important influences on perceptions of those groups.

The research used the false recognition paradigm (Todorov & Uleman, 2002, 2003, 2004) in which faces are presented, each one accompanied by a behavior description (e.g., "This individual participated in a protest."). In contrast to past studies focused on STIs about individuals, in Hamilton et al.'s (2015) research four faces were shown and the sentence described a group action (e.g., "This group participated in a protest."). After reading a series of such stimuli, the groups (sets of four faces) were shown again, this time with a trait word instead of behavior description. The trait word was either one that was implied by the behavior that had described the group (Match trial) or was a trait implied by the behavior of a different group (Mismatch trial). Participants were told the study concerned memory for verbal information and their task was to indicate whether or not the trait word was in the sentence that had described the group. The rationale for this method is that if the participant had made a spontaneous trait inference while encoding the original behavior information they would be more likely to say Yes to the probe question when in fact the word had not been in the group description. Thus, saying Yes was a false recognition and more of them were predicted to occur on match than on mismatch trials. If that happens it

would provide evidence that a spontaneous inference about the group had been made in encoding the group's behavior.

In a series of experiments Hamilton et al. (2015) provided compelling evidence that STIGs were made while processing group behavior information. The first study compared the occurrence of STIs and STIGs using the same materials. That is, in one condition participants saw one target person on each trial whereas in the other condition they saw groups (represented by four faces, as described above). The same behavior descriptions and probe traits were used in both conditions. The frequency of false recognitions (more Yes responses on Match than on Mismatch trials) was compared for the individual and group target conditions. Results showed that participants made STIs in the individual condition and STIGs in the group condition, and STIs and STIGs occurred with the same frequency; there were no target differences in making spontaneous trait inferences. Thus the common finding of STIs for individuals was replicated and, for the first time, evidence of STIGs for group targets was produced.

This evidence that STIGs occur as group-relevant behavioral information is processed raises several important questions. Under what conditions do STIGs occur and when do they not occur? Are STIGs made in processing information about all groups or are there certain types of groups for which they do not occur? If so, do STIGs become the basis for group impressions? What are the implications of STIGs for the formation of group stereotypes? Each of these questions poses multiple issues that will require more research to fully answer them. Subsequent studies reported by Hamilton et al. (2015) began to explore these questions empirically.

One issue concerns the efficiency of the process, such that it is initiated without intention or awareness and is not disrupted by other simultaneous tasks. In STI studies care is taken to assure that instructions focus participants on a presumed purpose of the study (memory) that would neither call for inferences nor draw attention to the underlying process of interest. In addition, the argument that the process is spontaneous is enhanced by demonstrating that imposing a second task simultaneously does not interfere with the initial process of interest. To test these ideas, Hamilton et al.'s (2015) next study tested whether a cognitive load would prevent STIGs from occurring. This study used the same procedure as the group condition of the first study, with the addition of a cognitive load manipulation in which participants had to keep a seven-digit (high load) or a two-digit (low load) number in mind as they read the stimulus information. This cognitive load would disrupt processing, and influence false recognitions, unless the trait inference process is truly spontaneous. In fact, participants made significantly more false recognitions in the match (compared to mismatch) condition, and this difference (i.e., evidence of STIGs) did not differ for the high vs. low cognitive load conditions. Thus STIGs are spontaneous and do not require extensive cognitive resources.

The first two studies establish that people do make spontaneous inferences about groups. But do they do so for all groups or only certain types of groups? If so, what types? There are, of course, many varieties of groups (Lickel et al., 2000) and we know that perceivers sometimes (but not always) process information differently about persons and groups (Hamilton & Sherman, 1996) so we need evidence for the generality of the findings produced thus far. Given their great variety, it's not clear, however, which types of groups to compare. As mentioned earlier, groups vary along a continuum of entitativity, or the perceived groupness of groups. Research has shown that people assume greater consistency and common underlying attributes in high compared to low entitativity groups (see Hamilton et al., 2014). If so, one might expect that STIGs are more likely to be made about high than about low entitativity groups. On the other hand, if STIGs are made spontaneously during the comprehension of behavioral information, they may occur routinely in processing group behavior, regardless of group properties (such as differences in entitativity). To test these ideas, two group conditions were created, manipulating the perceived entitativity of the groups through instructions describing the groups in high or low entitativity terms. Participants in both conditions learned about the groups following the same procedure used in previous studies. They then were tested for memory of the trait probes from which false recognitions were coded. Analyses showed no difference between the two conditions in the number of false recognitions made. Thus STIGs were made for both high and low entitativity groups, suggesting that they occur with generality across types of groups. Of course, there may be qualifications on such a broad conclusion, which can be explored in future research.

Based on the findings thus far, STIGs are robust and reflect a highly efficient process that appears to occur quite generally in comprehending group-descriptive information. One might wonder, though, what difference it makes. Once these spontaneous inferences are made, what role do they play in subsequent processing and with what consequences? Presumably spontaneous inferences become the basis for the first impressions participants form of the target groups. To test this idea, the next study (Hamilton et al., 2015) used the same paradigm with one change. After the initial phase in which participants learned the behaviors performed by the stimulus groups, the recognition task was not included. Instead, participants were shown the groups (sets of four faces) again and were asked to rate their impressions of each group on rating scales. Specifically, they rated each target group on three attributes: a trait implied by the group's behavior (a match trait), a trait implied by a different group's behavior (mismatch trait), and a control trait that was equated with the match trait on overall likeability (to test for halo effects). Analyses of these ratings revealed that participants made higher ratings of the groups on the match traits than on either mismatch or control traits. These comparisons support two important points. First, note that the traits implied by both match and mismatch trials would have been activated

by those behaviors presented during the first (learning) phase. Therefore the difference in ratings on match and mismatch traits reveals the specificity of the effect of STIGs on judgments of the groups. Second, the higher ratings on match than control traits (that were equated on overall likeability) shows that the higher ratings on match traits were not due to halo or valence effects, again revealing the specificity of the effects of STIGs on judgments. Thus the effects of STIGs on ratings are content based, not simply evaluative inferences. In sum, STIGs – unintended, nonconsciously-produced inferences – can be the basis of group impressions.

Stereotypes are, in some sense, group impressions. Do these results imply that STIGs can generate stereotypes? Perhaps, but more evidence of stereotyping would be useful. One of the hallmarks of stereotypes is that they generalize to group members about whom little is known, other than their membership in a stereotyped group. Would these STIG-based group impressions generalize to a new group member about whom no information has been provided? To explore this possibility, another experiment (Hamilton et al., 2015) used a modification of the paradigm from previous studies. Participants again were presented a series of slides, each one showing four faces and a one-sentence description of a behavior performed by the group. As in previous studies, the same groups of four faces were shown again, accompanied by a trait word, and participants' task was to indicate if the trait word had occurred in the sentence describing that group (assessing whether false recognitions were made). After completing this recognition task, participants were shown the same groups again, without either behavior description or probe trait. Instead, they were shown a picture of a new member of the group, not previously seen, with no information about him. Their task was to rate this new group member on several attributes, specifically, the same traits used in the previous study (match, mismatch, and control traits, unique to each group). Analyses of these ratings showed that the new member was rated higher on match traits than on either mismatch or control traits. Thus the STIG-based inferences generalized to a new group member about whom no information had been provided.

A further analysis substantiated the generalization of STIG-based effects on these ratings. The generalization of a spontaneously inferred trait to a new group member should only occur when a STIG had been made during the initial processing of the group's behavior. Therefore participants' responses on each trial were coded for whether a STIG had or had not been made for each stimulus group. Ratings of the new member were then compared for these two sets of trials. For those cases in which the participant had made a STIG about the group (as reflected in a false recognition), the new group member was rated significantly higher on the match than on the mismatch or control traits. In contrast, for those cases in which STIGs had not been made there was no difference in ratings on match and mismatch traits. This finding provides useful evidence for the process underlying the generalization effect. Specifically, this generalization effect was conditional on the participant

having made a STIG about the target group during behavior encoding and comprehension. Thus, STIG-based beliefs have generalized and have been transferred to a new group member.

In sum, the five experiments reported by Hamilton et al. (2015) have extended the evidence for spontaneous inferences in important ways. First, they provide the first evidence that spontaneous inferences occur as perceivers process behaviors enacted by groups. Second, these STIGs occur for different types of groups and even when processing under cognitive load. Third, once made, STIGs provide the basis for first impressions formed of these groups. And fourth, STIGs generalize to a new group member about whom no information has been learned. Together, this package of studies is rich in implications for further explorations of spontaneous inferences about groups.

New Processes for Stereotype Formation?

The vast literature on stereotypes has delineated numerous enabling preconditions and processes by which stereotypic beliefs may form. Such beliefs may arise through first-hand intergroup experiences or they may be acquired through social learning and socialization by significant others. These beliefs may originate in and be reinforced when accompanied by intergroup conflict, experiences of relative deprivation, or competition for scarce resources. They may be sustained and perpetuated by both cognitive and motivational biases and by system-justifying beliefs. Thus many forces can lead to stereotype formation.

The findings we have just reviewed are intriguing because they suggest new and previously unexplored processes by which stereotypic beliefs may form. It is important to note that none of the enabling preconditions enumerated earlier exist in the research on STIs and STIGs we have just described. In Crawford et al.'s (2002) work participants learned about members of two groups, and each person performed a trait-implying behavior. Their results showed that not only did participants make STIs about these individual persons but also that traits inferred about members of a group were transferred to other members of the same group (but not to members of the other group). These results demonstrated the generalization of spontaneously inferred traits of individual group members (STIs) to other members of the same group. Such generalization generates perceptions of group homogeneity and the interchangeability of group members (because they are perceived as sharing the same traits). Perceiving homogeneity and interchangeability are two ingredients that lay the groundwork for stereotype formation. Hamilton et al.'s (2015) research on STIGs demonstrated that such generalized characterizations of groups can be spontaneously inferred directly from the group's behavior. Merely observing a group's behavior can lead to inferred attributes describing the group as a whole.

In both cases spontaneous inferences (STIs and STIGs) spawn perceptions of groups that, although based on minimal information, can foster the

beginnings of group impressions. Moreover, these group impressions occurred in the absence of any of the preconditions that have been commonly cited as the bases for stereotype formation. Although these group impressions do not qualify as full-blown stereotypes, this research suggests some new and different processes by which the groundwork for stereotype formation may be laid. Could these spontaneously formed initial impressions develop and evolve into group stereotypes? An enormous amount of social psychological research has documented the mechanisms by which it could happen. First impressions, even based on minimal information, generate expectancies; those expectancies in turn lead to confirmatory biases; and those biases then serve to preserve and perpetuate existing beliefs about groups. Could spontaneous inferences be the catalyst for a parallel process in group perception? The findings we have reviewed suggest that STIGs could plant seeds that grow and blossom into more consequential stereotypes.

Stereotypes and Spontaneous Inferences

In the preceding section we have seen that spontaneous inferences (STIGs) could lay groundwork on which stereotypes may develop. In this section we explore the interplay between existing stereotypes and spontaneous inferences.

In the years since Winter and Uleman's (1984) original research introducing the notion of spontaneous trait inferences there have been hundreds of studies on the topic. In almost all of these experiments the participant knows virtually nothing about the person or group being described – except for one trait-implying behavior the person or group has performed. The paucity of information is of course intentional, as an important means of controlling (or preventing) the influence of other variables in order to study the inference process in pure form. Yet in everyday life this doesn't happen; we virtually always know something about the target person or group in advance. One thing we almost always know is the person's membership in certain groups and social categories (or, in the case of group stimuli, the nature of the group). Many of these categories – gender, race, age, nationality, occupation – have well-developed stereotypes that generate inferences that are used in understanding the target person or group. Consequently, when observing a person's behavior we would have two kinds of inferences available – stereotype-based inferences and behavior-based inferences (STIs). What is the interplay between these two kinds of inferences?

In a series of studies (Wigboldus et al., 2003; see also Wigboldus et al., 2004) participants were shown a series of behaviors on a computer screen, each one enacted by a person identified by a group membership (e.g., professor, garbage man). The behavior (e.g., "won the science quiz") was either consistent (professor) or inconsistent (garbage man) with the activated stereotype. Following the sentence a trait probe word (e.g., smart) appeared and the participant's task was to indicate whether the word was in the stimulus

sentence. The participant's response time in responding to the probe was recorded. In none of the theoretically relevant trials did the probe word appear in the sentence, so the correct answer was always No. The behavior (won the science quiz) implies the probe word (smart), so there may be uncertainty in making the response (Did I see the word or did I infer it?). The question of interest in the study was whether the consistency or inconsistency of the behavior with the activated stereotype would influence the response times. If it was the professor who won the science quiz, the behavior is consistent with the activated stereotype, which should add further uncertainty. The probe trait (smart) might have been in the sentence describing the professor or it might have been inferred from the behavior. On the other hand, if the garbage man won the science quiz the implications of the behavior (smart) would be inconsistent with the stereotype of garbage men so the inferred attribute might be dismissed and the inference from the behavior would not occur. In the latter case the STI is not made so it should take less time to respond than in the former case. In other words, a stereotype can inhibit an STI from occurring when the behavior is inconsistent with the activated stereotype. Wigboldus et al.'s (2003) results showed exactly that outcome. More recent research has documented other variables that can inhibit STIs (Crawford, Skowronski, Stiff, & Scherer, 2007; Rim et al., 2009; Ramos et al., 2012; Wigboldus et al., 2004; Yan et al., 2012).

The findings reported by Wigboldus et al. (2003) are noteworthy. As the literature reviewed in this chapter indicates, research on STIs has repeatedly shown that spontaneous inferences are routinely made in almost every case when they have been studied. What was unusual in the Wigboldus et al. (2003) findings is that they revealed a case in which the STI was not made; the prior activation of a stereotype inhibited the STI from occurring.

This finding raises another intriguing possibility. Essentially the Wigboldus et al. (2003) study pitted two spontaneous inferences against each other: one activated by the group stereotype, the other activated by the behavior. When there was an inconsistency between the two, the stereotype-based inference blocked the behavior-based inference. In Wigboldus et al.'s study the actor's group membership was always presented in advance of the behavior (The garbage man won the science quiz.) so the stereotype was activated *before* the behavior that would promote a trait inference. But what if things occurred in the reverse order? What if the participant learned about the actor's behavior (The man won the science quiz.) before learning the actor's group membership (The man is a garbage man.). In that case the STI should occur when the behavior is encoded, prior to activation of the stereotype. Can the same inhibitory effect occur in reverse? Would the STI made in comprehending the behavior (Won science quiz → smart) inhibit activation of the stereotype-based inference (Garbage man → stupid)? If so, it would mean that an STI – an inference that is unintended and occurs outside of conscious awareness – had prevented the use of a pre-existing group stereotype. We know of no study that has tested this hypothesis, but we consider it a

possibility worth exploring because of its interesting, and potentially important, implications.

Making Sense of Inconsistencies

Wigboldus et al. (2003) had shown that stereotypes can inhibit STIs when the behavior is inconsistent with stereotype-based expectancies. In that case, what will happen instead? For example, what will happen when relevant situational information is included? Will stereotypes diminish the likelihood of any spontaneous inferences? This might happen if strong *a priori* expectancies can inhibit all forms of spontaneous inferencing. Alternatively, when faced with stereotype disconfirming information, will other types of spontaneous inferences become more likely? This might occur if the inconsistency triggers processes aimed at finding some interpretive meaning in the stereotype-inconsistent behavior.

Ramos et al. (2012) studied the interplay between STIs and *spontaneous situation inferences (SSIs)*. Their stimulus behaviors implied traits and also referred to situational factors that could provide an alternative (non-trait) interpretation of the behavior. That is, the stimulus information implied both STIs and SSIs. How would stereotype-disconfirmation affect making STIs and/or SSIs? Ramos et al. (2012) postulated that making SSIs would be one way of understanding the meaning of the stereotype-inconsistent behaviors.

In their research the recognition probe paradigm was used to test these possibilities. Each trait-implying behavior was performed by a person for whom the behavior was consistent or inconsistent with the persons' group membership. For example, "the old man (dancer) stepped on his partner's feet while dancing." The sentence was extended with a continuation that permitted a situational inference, for example, "after a long day of work." Each sentence was then followed by either a trait probe (fumbling) or a situational gist probe (tired). Participants' response times to indicate that the probe word was not in the behavior sentence were recorded. The results of two experiments showed that, when the behavior was inconsistent with the actor's group membership, group stereotypes inhibited STIs but they facilitated SSIs. These results are interesting in that they demonstrate two effects of stereotypes on spontaneous processing. On the one hand, STIs are made following stereotype consistent behaviors, thereby facilitating the anticipation and prediction of the actor's future behavior. On the other hand, SSIs are made following stereotype inconsistent behaviors as a means of interpreting the meaning of those expectancy-inconsistent behaviors.

Spontaneous Inferences and Intergroup Differentiation

Thus far we have reviewed research on STIs made as perceivers comprehend the behaviors of individual target persons, and we have reviewed a smaller but more recent literature on spontaneous inferences about groups, based on

group behaviors. In both cases it is clear that spontaneous inferences are robust, they occur without intention, and they are often made without the perceiver's awareness of these inferences being made. An important question that has not been investigated is whether perceivers spontaneously make inferences that differentiate between targets. When processing behavioral information about, and forming impressions of, two or more individuals, do people simultaneously make different STIs about those persons, unique to each one? That is, can people spontaneously form different impressions simultaneously? We know of no research that has explored this possibility. In the group domain the issue seems even more relevant. Throughout the history of social psychology immense effort has been devoted to understanding the formation, perpetuation, and change of stereotypes of and prejudicial attitudes toward different groups, often comparing one versus another group. When processing information about more than one group, do perceivers spontaneously make different STIGs about the groups? That is, do people spontaneously make intergroup differentiations in comprehending group behaviors? These important questions have received little research attention. We briefly summarize examples of how they have been explored.

STIs and Evaluative Ingroup Bias

One of the most consistent and pervasive findings in research on intergroup perception is the ingroup bias, in which the ingroup and its members are evaluated more favorably than the outgroup and its members (Brewer, 1979). These effects occur even in the minimal group paradigm in which no actual differences between the groups is known (Tajfel, 1970; Tajfel et al., 1971). Otten and Moskowitz (2000) showed that differentiation between these arbitrary groups can foster evaluatively biased STIs that are consistent with and reinforce the ingroup bias.

Participants were randomly assigned to arbitrary groups that participants were told reflected different perceptual styles. They then read trait-implying behaviors, some positive and some negative, performed by ingroup and outgroup members. Each sentence was followed by a positively or negatively valenced trait word that was or was not implied by the behavior. Participants' task was to indicate whether that probe word had been in the preceding sentence. Their response times in making these responses were recorded. Making an STI would slow these judgments. Analyses showed that STIs were made when sentences described ingroup members performing positively-valenced behaviors implied by the probe trait. Thus ingroup favoritism was evident in the STIs made when behaviors of ingroup members were processed. There was no evidence of outgroup derogation, which would be manifested in longer response times for outgroup members performing negatively valenced behaviors, consistent with other research on ingroup bias (Brewer, 1979).

STIGs and Intergroup Differentiation

As we have noted, there is little research focused on spontaneous inferences about group targets, despite the fact that we encounter, perceive, and interact with groups every day. The research on STIGs (Hamilton et al., 2015) showed that people do make such inferences, but that research investigated inferences made about one group. Other than the Otten and Moskowitz (2000) experiment on evaluative ingroup bias, we know of no research on spontaneous inferences about two or more groups. We often perceive members of two (or more) groups and clearly we form impressions (and sometimes stereotypes) of those groups. We are also aware that there are times when we devote considerable attention and thought to those impressions, often thinking and discussing with others the similarities and differences between groups. In these cases, as in the STIGs research, the concern is with the groups as entities and their properties. If people spontaneously make inferences about groups (as the STIGs work shows), can they spontaneously (without intention and awareness) form different impressions of two different groups? Can they form different group impressions simultaneously? If so, can we provide evidence of *spontaneous intergroup differentiation*?

Recently we have begun to explore these questions (Thurston & Hamilton, unpublished data). We extended and adapted the STIG paradigm used by Hamilton et al. (2015). At the outset the instructions stated that the study was concerned with people's "memory for visual and verbal information" and at several points throughout the procedure instructions reiterated that participants' memory for stimulus information would be tested. Because participants would be learning about several groups and their behaviors, these memory instructions were designed to diminish the likelihood of suggesting forming impressions.

Participants learned about unidentified groups from classes at two universities, labelled University A and University B (instructions indicated that neither university was UCSB, the participants' own school). Fourteen groups from each university were shown. As in Hamilton et al.'s studies, each group was presented by faces of four people along with a trait-implying behavior description. For the groups from University A, nine of the 14 behaviors implied competence (e.g., solved a computer problem for the company; class project received highest honors). For University B nine of the 14 groups performed behaviors that implied warmth (volunteered all weekend at a homeless shelter; collected toys for a children's hospital). For both universities the remaining five groups (filler trials) performed neutral behaviors that included a relevant trait, but in every case it was a trait different from competence or warmth (e.g., This group was quiet while they took the exam). The 28 groups were presented in random order.

After presentation of the 28 groups participants were given a 3-min filler task followed by the recognition task. Each group (consisting of four faces) was presented again without the behavior but with a probe trait word, with the

instruction to indicate if that word was in the sentence describing that group. Except for the filler trials, the correct answer in all cases was No; the probe traits were implied by the group's behavior so responding Yes was a false recognition and an indication that an inference had been made. For each university there were nine groups of interest, and for each one the recognition responses were compared for three types of cases: (a) match trials in which the probe word was a trait implied by that group's behavior; (b) mismatch-within trials in which the probe word was implied by the behavior of a different group from the *same* university; and (c) mismatch-between trials, in which the probe trait was implied by the behavior of a group from the *other* university. Again, a higher number of false recognitions on match than on mismatch trials indicates that more STIGs were made. The filler trials (five for each university) were included to provide opportunities to correctly say Yes to the probe question. Responses on these trials were not included in data analyses.

Analyses of the number of false recognitions for the three types of trials, separately for University A and University B, showed results strongly indicating that STIGs were made for both universities. The largest number of STIGs occurred on match trials for both universities, the smallest number of STIGs occurred for mismatch-between trials, and mismatch-within trials were intermediate. These differences were strongly significant and indicate that, at the level of spontaneous inferences, the two universities were differentially represented in memory.

After completing the recognition task participants rated University A and University B on several trait dimensions. Some of the traits assessed participants' impressions of the universities on the attributes predominantly represented in information about the two universities (warmth, competence). These scales were designed to assess whether participants had formed different impressions of the two schools. Four of the scales were traits reflecting competence and four represented warmth. The final three scales were general evaluation measures (good-bad, favorable-unfavorable, and positive- negative). Ratings on these scales were collapsed to form measures of Competence, Warmth, and General Evaluation. Analyses of these data confirmed that participants formed different impressions of the two universities, in accordance with the manipulation of competence and warmth information.

The false recognition results clearly confirm that participants made STIGs from the behavioral information they processed. Thus people can spontaneously differentiate between two groups, they differentially make inferences about the two groups, and they spontaneously form different impressions – without intention to do so and without awareness they're doing it – on a task that they think is intended to assess their memory.

In all of these studies on STIGs, people were learning about anonymous groups about whom they know very little. For each stimulus group participants are shown four faces, Caucasian males in all cases, along with one trait-implying behavior, and they see multiple groups. They have no other knowledge or expectancies about these groups, yet they spontaneously make

inferences about them. In real life, however, people usually have some information about the groups they observe, and they may have stereotypic or other beliefs about them. One wonders, then, how other group information would influence the processing of information. We know that group characteristics typically influence explicit judgments of groups. Would knowing groups' properties also influence the likelihood of making STIGs, and if so, how? As a first step toward exploring these questions, we conducted a study much like our previous studies but in which we varied the racial composition of the stimulus groups.

Participants were shown a series of 51 stimulus groups, each one comprised of photos of four men with a one-sentence description of a group behavior. There were three different conditions, varying in the racial composition of the groups. Seventeen of the groups were comprised of four black males, 17 had four white males, and 17 had two black and two white males. In each condition 12 of the 17 groups performed trait-implying behaviors, but they were not stereotypic behaviors. Also, all behaviors were moderately positive in valence, with no undesirable behaviors about any of the groups. The remaining five groups in each race condition were fillers.

After viewing the presentation of the groups, participants were shown the groups again, accompanied by a trait word, and their task was to indicate whether that word had been in the sentence describing that group. The words presented created either match trials or two kinds of mismatch trials. For example, if the stimulus group consisted of four black men, then the probe trait word could be a trait implied by the behavior performed by that group (match trial), it could be a trait implied by the behavior of one of the white groups, or it could be a trait implied by the behavior of one of the mixed race groups. Similar options were created for responding when white and mixed race groups were shown. The number of Yes responses (false recognitions, indicative of a STIG having been made during encoding) for each group type and probe words was determined.

Analyses of these false recognitions revealed evidence of spontaneous differential impression formation. Overall, for reasons that are not entirely clear, participants made more false recognitions about black groups than about either white or mixed race groups. They were quite willing to say that Yes, that trait word was in the sentence describing that group – regardless of what type of group's behavior implied that trait. Perhaps the salience of four black men for our non-black participants evoked a greater number of false recognitions, or perhaps participants had reduced attention to outgroup members, processing information less thoroughly, leading to more false recognition errors. Of greater theoretical interest are the comparisons among trait types for each group type. Specifically, for each racial group type, the number of false recognitions (STIGs) was (with one exception) significantly higher on match trials than on either type of mismatch trial. The greater number of STIGs on match than on mismatch trials documents not only that participants made spontaneous inferences about these groups but also that

these implied traits were uniquely associated with the groups that performed the trait-implying behavior. That is, they not only made STIGs but also significantly differentiated among the groups at an implicit level.

After completing the false recognition task participants rated each group on a series of trait scales, including both positive and negative attributes. Overall, the ratings of the groups were quite favorable, a result that is not surprising given that all behavior sentences describing the groups were positively valenced. In contrast to the STIG results based on false recognition data, these explicit ratings showed that on positively valued traits black groups were rated highest and white groups lowest, with mixed race groups receiving intermediate ratings. On negatively valued traits white groups were rated higher than black or mixed race groups, with the latter two group types not differing in the valence of their mean ratings. These results are the opposite of what one might expect if traditional stereotypes were driving people's ratings. Instead, they clearly reveal that participants' ratings were strongly influenced by social desirability and/or self-presentational biases. More importantly, the differing patterns of findings for explicit (ratings) and implicit (STIGs) measures further substantiate that these spontaneous processes are less affected by biases that influence conscious judgments. Thus STIGs independently contribute to the initial formation of group impressions.

Concluding Remarks

There are several ways that researchers can become well known, respected and admired for their contributions to the field. One way is to introduce a new theory that either redefines how people think about an existing domain or provides a conceptual framework that introduces a new theory that opens up an entirely new line of inquiry. Heider on attribution, Festinger on dissonance, Bem on self-perception, Tajfel on social identity are examples that come to mind. Each of them broke new ground and generated new growth in the discipline. A second route to prominence would be to develop a new paradigm or methodology that demonstrates that a phenomenon can be studied empirically or that reveals a new outcome. Such landmark contributions include Milgram's studies of obedience to authority, Tajfel's minimal intergroup paradigm, and two groundbreaking instances by Asch, who introduced seemingly simple means of studying the complex topics of impression formation and conformity. These are but a few of the milestone contributions that have shaped and (re)directed the course of conceptual and empirical development of social psychology. Merely mentioning one of these names immediately brings to mind the singular work the person produced. In addition, the discipline is well populated with figures whose work, over time, has contributed in important ways to one topic, then moved on to another topic to which the person has made new advances, then has taken up a third area of work where new contributions are made, and so on. Such people are rightly admired and recognized for the impressive breadth of their work and

the meaningful contributions made in each of several domains. These are some of the ways people develop importance and a central place on the landscape of social psychology through their work.

A major goal of this volume is to recognize and honor the work of Jim Uleman and his contributions to the study of impression formation. Now that Jim has retired, we can use the above observations to consider how his research has established his long-lasting prominence in social psychology. People who are even casually familiar with the world of social psychology, when hearing the term *spontaneous trait inference*, or even hearing the initials *STI*, will instantly realize it refers to an entire area of research initiated by a researcher named James Uleman. If they have had some exposure to the literature in social psychology they may even know there was an article by Winter and Uleman (1984) that started it all, and that a considerable body of research has emanated from that initial paper – a line of research that has now continued for almost 40 years and shows no signs of diminishing. These people may further know that several different paradigms have been developed or adapted, mostly from the memory literature, in order to study these STIs, and that three of those paradigms were developed by Jim Uleman and his colleagues. And if they wondered why one would adapt paradigms from the memory tradition to study processes in social psychology, they might then learn that this intriguing enterprise is focused on mental processes that happen involuntarily and without the person's awareness that it is happening, and to pursue that work requires clever methods by clever investigators. Finally, they may know that this entire 40+ year history was not only initiated by Jim Uleman but it has developed and transpired under his watchful eye and guidance.

The contributors to this volume are scientists who work long hours, with little hope of getting rich or of having fame and glory (outside of their small corner of the Ivory Tower), yet they love the work they do and they take their work very seriously. Why? Because they are investigating one part of the answer to the broad and fundamental question of how the mind works. Jim Uleman has contributed enormously to solving this piece of that puzzle. Through his own and his students' work in his lab, and through the vast, extensive research he has stimulated in others in their labs, Jim Uleman has advanced our understanding of processes that are fascinating, elusive and difficult to study, but that underlie the process of impression formation. For all that he has accomplished, we say to him, "Well done, Jim!"

References

Aarts, H., Gollwitzer, P. M., & Hassin, R. R. (2004). Goal contagion: Perceiving is for pursuing. *Journal of Personality and Social Psychology*, *87*(1), 23–37.

Allport. G. W. (1935). Attitudes. In C. Murchison (Ed.), *Handbook of social psychology* (pp. 798–844). Worcester, MA: Clark University Press.

Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal and Social Pychology*, *41*, 258–290.

Bargh, J. A. (1982). Attention and automaticity in the processing of self-relevant information. *Journal of Personality and Social Psychology*, *43*, 425–436.

Bargh, J. A. (1984). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. Wyer, Jr. & Thomas K. Srull (Eds.), *Handbook of social cognition* (2nd ed., Vol. 1, Basic processes, pp. 1–40). Hillsdale, NJ: Lawrence Erlbaum.

Bargh, J. A., & Pietromonaco, P. (1982). Automatic information processing and social perception: The influence of trait information presented outside of conscious awareness on impression formation. *Journal of Personality and Social Psychology*, *43*, 437–449.

Bassili, J. N. (1989). Trait encoding in behavior identification and dispositional inference. *Personality and Social Psychology Bulletin*, *15*, 285–296.

Bassili, J. N., & Smith, M. C. (1986). On the spontaneity of trait attribution: Converging evidence for the role of cognitive strategy. *Journal of Personality and Social Psychology*, *50*, 239–245.

Brewer, M. B. (1979). Ingroup bias in the minimal intergroup situation: A cognitive-motivational analysis. *Psychological Bulletin*, *86*, 307–324.

Brewer, M. B. (2015). Motivated entitativity: When we'd rather see the forest than the trees. In S. J. Stroessner & J. W. Sheman (Eds.), *Social perception from individuals to groups* (pp. 161–176). New York: Psychology Press.

Brewer, M. B., & Harasty, A. S. (1996). Seeing group as entities: The role of perceiver motivation. In R. M. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition* (Vol. 3, pp. 347–370). New York: Guilford Press.

Brigham, J. C. (1971). Ethnic stereotypes. *Psychological Bulletin*, *76*, 15–38.

Brown, R. D., & Bassili, J. N. (2002). Spontaneous trait associations and the case of the superstitious banana. *Journal of Experimental Social Psychology*, *38*, 87–92.

Bruner, J. S., & Taguiri, R. (1954). The perception of people. In G. Lindzey (Ed.), *Handbook of social psychology* (Vol. 2, pp. 634–654). Cambridge, MA: Addison Wesley.

Carlston, D. E., & Skowronski, J. J. (1994). Savings in the relearning of trait information as evidence of spontaneous inference generalization. *Journal of Personality and Social Psychology*, *66*, 840–856.

Carlston, D. E., & Skowronski, J. J. (2005). Linking versus thinking: Evidence for the different associative and attributional biases of spontaneous trait inference and spontaneous trait transference. *Journal of Personality and Social Psychology*, *89*, 884–898.

Carlston, D. E., Skowronski, J. J., & Sparks, C. (1995). Savings in relearning: II. On the formation of behavior based trait associations and transferences. *Journal of Personality and Social Psychology*, *69*, 420–435.

Chen, J. M., Banerji, I., Moons, W. G., & Sherman, J. W. (2014). Spontaneous social role inferences. *Journal of Experimental Social Psychology*, *55*, 146–153.

Crawford, M. T., Sherman, S. J., & Hamilton, D. L. (2002). Perceived entitativity, stereotype formation, and the interchangeability of group members. *Journal of Personality and Social Psychology*, *83*, 1076–1094.

Crawford, M. T., Skowronski, J. J., & Stiff, C. (2007). Limiting the spread of spontaneous trait transference. *Journal of Experimental Social Psychology*, *43*, 466–472.

Crawford, M. T., Skowronski, J. J., Stiff, C., & Scherer, C. R. (2007). Interfering with inferential, but not associative, processes underlying spontaneous trait inference. *Personality and Social Psychology Bulletin*, *33*, 677–690.

D'Agostino, P. R. (1991). Spontaneous trait inferences: Effects of recognition instructions and subliminal priming on recognition performance. *Personality and Social Psychology Bulletin*, *17*, 70–77.

D'Agostino, P. R., & Beegle, W. (1996). Reevaluation of the evidence for spontaneous trait inferences. *Journal of Experimental Social Psychology*, *32*, 153–164.

Enge, L. R., Lupo, A. K., & Zarate, M. A. (2015). Neurocognitive mechanisms of prejudice formation: The role of time-dependent memory consolidation. *Psychological Science*, *26*, 964–971.

Fein, S. (1996). Effects of suspicion on attributional thinking and the correspondence bias. *Journal of Personality and Social Psychology*, *70*, 1164–1184.

Fein, S., Hilton, J. L., & Miller, D. T. (1990). Suspicion of ulterior motivation and correspondence bias. *Journal of Personality and Social Psychology*, *58*, 753–764.

Ferreira, M. B., Garcia-Marques, L., Hamilton, D. L., Ramos, T., Uleman, J. S., & Jeronimo, R. (2012). On the relation between spontaneous trait inferences and intentional inferences: An inference monitoring hypothesis. *Journal of Experimental Social Psychology*, *48*, 1–12.

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, *117*, 21–38.

Ham, J., & van den Bos, K. (2008). Not fair for me! The influence of personal relevance on social justice inferences. *Journal of Experimental Social Psychology*, *44*, 699–705.

Ham, J. & Vonk, R. (2003). Smart and easy: Co-occurring activation of spontaneous trait inferences and spontaneous situational inferences. *Journal of Experimental Social Psychology*, *39*, 434–447.

Hamilton, D. L. (1998). Dispositional and attributional inferences in person perception. In J. M. Darley & J. Cooper (Eds.), *Attribution and social interaction: The legacy of Edward E. Jones* (pp. 99–114). Washington, DC: American Psychological Association.

Hamilton, D. L., & Sherman, S. J. (1996). Perceiving persons and groups. *Psychological Review*, *103*, 336–355.

Hamilton, D. L., Chen, J. M., Ko, D., Winczewski, L., Banerji, I., & Thurston, J. A. (2015). Sowing the seeds of stereotypes: Spontaneous inferences about groups. *Journal of Personality and Social Psychology*, *109*, 569–588.

Hamilton, D. L., Chen, J. M., & Way, N. (2011). Dynamic aspects of entitativity: From group perception to social interaction. In R. M. Kramer, G. J. Leonardelli, & R. W. Livingston (Eds.), *Social cognition, social identity, and intergroup relations: A festschrift in honor of Marilynn Brewer* (pp. 27–52). New York: Psychology Press.

Hamilton, D. L., Sherman, S. J., & Castelli, L. (2002). A group by any other name – The role of entitativity in group perception. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 12, pp.139–166). Chichester, England: Wiley.

Hamilton, D. L., Sherman, S. J., & Rodgers, J. (2004). Perceiving the groupness of groups: Entitativity, homogeneity, essentialism, and stereotypes. In V. Yzerbyt, C. M. Judd, & O. Corneille (Eds.), *The psychology of group perception: Perceived variability, entitativity and essentialism* (pp. 39–60). Philadelphia, PA: Psychology Press.

Hamilton, D. L., Sherman, S. J., Way, N., & Percy, E. (2014). Convergence and divergence in perceptions of persons and groups. In M. Mikulincer, P. R. Shaver (Eds.), J. F. Dovidio, & Simpson, J. A. (Assoc. Eds.), *APA handbook of personality*

and social psychology: Vol. 2. Group processes (pp. 229–261). Washington, DC: American Psychological Association.

Hassin, R. R., Aarts, H., & Ferguson, M. J. (2005). Automatic goal inferences. *Journal of Experimental Social Psychology*, *41*, 129–140.

Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.

Higgins, E. T., Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology*, *13*, 141–154.

Hilton, J. L., Fein, S., & Miller, D. T. (1993). Suspicion and dispositional inference. *Personality and Social Psychology Bulletin*, *19*, 501–512.

Jones, E. E. (1979). The rocky road from acts to dispositions. *American Psychologist*, *34*, 107–117.

Jones, E. E. (1990). *Interpersonal perception*. New York: W.H. Freeman.

Jones, E. E., & Davis, K. E. (1965). From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 219–266). New York: Academic Press.

Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology*, *3*, 1–24.

Katz, D., & Braly, K. (1933). Racial stereotypes in one hundred college students. *Journal of Abnormal and Social Psychology*, *28*, 280–290.

Katz, D., & Braly, K. (1935). Racial prejudice and facial stereotypes. *Journal of Abnormal and Social Psychology*, *30*, 175–193.

Lickel, B., Hamilton, D. L., Wieczorkowska, G., Lewis, A., Sherman, S. J., & Uhles, A. N. (2000). Varieties of groups and the perception of group entitativity. *Journal of Personality and Social Psychology*, *78*, 223–246.

Lupfer, M. B., Clark, L. F., & Hutcherson, H. W. (1990). Impact of context on spontaneous trait and situational attributions. *Journal of Personality and Social Psychology*, *58*, 239–249.

Lupo, A. K., & Zarate, M. A. (2019). Guilty by association: Time-dependent memory consolidation facilitates the generalization of negative – but not positive – person memories to group and self-judgments. *Journal of Experimental Social Psychology*, *83*, 78–87.

Mae, L., Carlston, D. E., & Skowronski, J. J. (1999). Spontaneous trait transfer to familiar communicators: Is a little knowledge a dangerous thing? *Journal of Personality and Social Psychology*, *77*, 233–246.

McArthur, L. Z. (1972). The how and what of why: Some determinants and consequences of causal attribution. *Journal of Personality and Social Psychology*, *22*, 171–193.

McCarthy, R. J., & Skowronski, J. J. (2011). What will Phil do next? Spontaneously inferred traits influence predictions of behavior. *Journal of Experimental Social Psychology*, *47*, 321–332.

Moskowitz, G. B., & Olcaysoy Okten, I. (2016). Spontaneous goal inferences (SGI). *Social and Personality Compass*, *10*, 64–80.

Moskowitz, G. B., & Roman, R. J. (1992). Spontaneous trait inferences as self-generated primes: Implications for conscious social judgment. *Journal of Personality and Social Psychology*, *62*, 728–738.

Newman, L. S. (1991). Why are traits inferred spontaneously? A developmental approach. *Social Cognition*, *9*, 221–253.

Olcaysoy Okten, I., & Moskowitz, G. B. (2018). Goal versus trait explanations: Causal attributions beyond the trait-situation dichotomy. *Journal of Personality and Social Psychology*, *114*, 211–229.

Olcaysoy Okten, I., & Moskowitz, G. B. (2020). Spontaneous goal versus spontaneous trait inferences: How ideology shapes attributions and explanations. *European Journal of Social Psychology*, *50*, 177–188.

Olcaysoy Okten, I., Schneid, E. D., & Moskowitz, G. B. (2019). On the updating of spontaneous impressions. *Journal of Personality and Social Psychology*, *117*, 1–25.

Orghian, D., Smith, A., Garcia-Marques, L., & Heinke, D. (2017). Capturing spontaneous trait inferences with the modified free association paradigm. *Journal of Experimental Social Psychology*, *73*, 243–257.

Orghian, D., Garcia-Marques, L., Uleman, J. , & Heinke, D. (2015). a connectionist model of spontaneous trait inference and spontaneous trait transference: Do they have the same underlying processes? *Social Cognition*, *33*, 20–66.

Otten, S., & Moskowitz, G. B. (2000). Evidence for implicit evaluative in-group bias: Affect biased spontaneous trait inference in a minimal group paradigm. *Journal of Experimental Social Psychology*, *36*, 77–89.

Ramos, T., Garcia-Marques, L., Hamilton, D. L., Ferreira, M. B., & van Acker, K. (2012). What I infer depends on who you are: The influence of stereotypes on trait and situational spontaneous inferences. *Journal of Experimental Social Psychology*, *48*, 1247–1256.

Read, S. J., Jones, D. K., & Miller, L. C. (1990). Traits as goal-based categories: The importance of goals in the coherence of dispositional categories. *Journal of Personality and Social Psychology*, *58*(6), 1048–1061.

Rim, S., Uleman, J., & Trope, Y. (2009). Spontaneous trait inference and construal level theory: Psychological distance increases nonconscious trait thinking. *Journal of Experimental Social Psychology*, *45*, 1088–1097.

Schneid, E. D., Carlston, D. E., & Skowronski, J. J. (2015). Spontaneous evaluative inferences and their relationship to spontaneous trait inferences. *Journal of Personality and Social Psychology*, *108*, 681–696.

Schneider, D. J. (1973). Implicit personality theory: A review. *Psychological Bulletin*, *79*, 294–309.

Schneider, D. J. (2004). *The psychology of stereotyping*. New York: Guilford Press.

Skitka, L. J., Mullen, E., Griffin, T., Hutchinson, S., & Chamberlin, B. (2002). Dispositions, scripts, or motivated correction? Understanding ideological differences in explanation for social problems. *Journal of Personality and Social Psychology*, *83*, 470–487.

Skowronski, J. J., Carlston, D. E., Mae, L., & Crawford, M. T. (1998). Spontaneous trait transference: Communicators take on the qualities they describe in others. *Journal of Personality and Social Psychology*, *74*, 837–848.

Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American*, *223*, 96–102.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behavior. *European Journal of Social Psychology*, *1*, 149–177.

Todd, A. R., Molden, D. C., Ham, Y., & Vonk, R. (2011). The automatic and co-occurring activation of multiple social inferences. *Journal of Experimental Social Psychology*, *47*, 37–49.

Todorov, A., & Uleman, J. S. (2002). Spontaneous trait inferences are bound to actors' faces: Evidence from a false recognition paradigm. *Journal of Personality and Social Psychology*, *83*, 1051–1065.

Todorov, A., & Uleman, J. S. (2003). The efficiency of binding spontaneous trait inferences to actors' faces. *Journal of Experimental Social Psychology*, *39*, 549–562.

Todorov, A., & Uleman, J. S. (2004). The person reference process in spontaneous trait inferences. *Journal of Personality and Social Psychology*, *87*, 482–493.

Uleman, J. S. (1989). A framework for thinking intentionally about unintended thought. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 425–449). New York: Guilford Press.

Uleman, J. S. (1999). Spontaneous versus intentional inferences in impression formation. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 141–160). New York: Guilford Press.

Uleman, J. S., Hon, A., Roman, R., Moskowitz, G. B. (1996). On-line evidence for spontaneous trait inferences at encoding. *Personality and Social Psychology Bulletin*, *22*, 377–394.

Uleman, J. S., & Moskowitz, G. B. (1994). Unintended effects of goals on unintended inferences. *Journal of Personality and Social Psychology*, *66*, 490–501.

Uleman, J. S., Newman, L. S., & Moskowitz, G. B. (1996). People as flexible interpreters: Evidence and issues from spontaneous trait inference. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 28, pp. 211–279). San Diego, CA: Academic Press.

Uleman, J. S., Newman, L. S., & Winter, L. (1992). Can personality traits be inferred automatically?: Spontaneous inferences require cognitive capacity at encoding. *Consciousness and Cognition*, *1*, 77–90.

Uleman, J. S., Saribay, S. D., & Gonzalez, C. M. (2008). Spontaneous inferences, implicit impressions, and implicit theories. *Annual Review of Psychology*, *59*, 329–360.

Wigboldus, D. H. J., Dijksterhuis, A., & van Knippenberg, A. (2003). When stereotypes get in the say: Stereotypes obstruct stereotype-consistent trait inferences. *Journal of Personality and Social Psychology*, *84*, 470–484.

Wigboldus, D. H. J., Sherman, J. W., Franzese, H. L., & van Knippenberg, A. (2004). Capacity and comprehension: Spontaneous stereotyping under cognitive load. *Social Cognition*, *22*, 292–309.

Winter, L., & Uleman, J. S. (1984). When are social judgments made? Evidence for the spontaneousness of trait inferences. *Journal of Personality and Social Psychology*, *47*, 237–252.

Winter, L., Uleman, J. S., & Cuniff, C. (1985). How automatic are social judgments? *Journal of Personality and Social Psychology*, *49*, 904–917.

Wyer, R. S., Jr., & Lambert, A. J. (1994). The role of trait constructs in person perception: A historical perspective. In P. G. Devine, D. L. Hamilton, & T. M. Ostrom (Eds.), *Social cognition: Impact on social psychology* (pp. 109–142). San Diego, CA: Academic Press.

Yan, X., Wang, M., & Zhang, Q. (2012). Effects of gender stereotypes on spontaneous trait inferences and the moderating role of gender schematicity: Evidence from Chinese undergraduates. *Social Cognition*, *30*, 220–231.

14 Forming and Managing Impressions Across Racial Divides

Cydney H. Dupree
School of Management, University College London

Decades of social cognitive research examines *impression formation*—how people (perceivers) form impressions of others and the implications for their attitudes, thoughts, and behaviors. Related research examines *impression management*—how people (actors) attempt to manage others' impressions and the implications for others' attitudes, thoughts, and behaviors. Like all areas of psychological research (Roberts et al., 2020), this work has primarily featured White American participants and tested perceptions of others assumed to be White—largely ignoring the roles of perceiver or actor racial group membership. Thus, impression formation and management research evolved separately from group dynamics research, including work on stereotyping, prejudice, and intergroup relations. However, in an increasingly diverse society, interacting with members of other racial groups—both online and in person—is a requisite part of social and professional life, and science is hardly generalizable if phenomena are only tested among White Americans (see Dupree & Kraus, 2021, for comment). Moreover, impression formation and impression management research have evolved largely separately from each other, but these phenomena go hand-in-hand. People engage in impression management—often countering negative stereotypes in the process; this directly impacts how they are seen by others. We thus cannot fully understand impression formation or impression management without studying these phenomena across group contexts, examining how the racial group membership of actor and perceiver impact people's impressions of others and how they try to manage others' impressions. This chapter reviews the state of research on impression formation and impression management across racial divides, briefly describing foundational research before exploring new and upcoming research that examines these phenomena in diverse group contexts.

Impression Formation

I will first briefly describe the foundational research on impression formation—featuring largely White perceivers and actors—before exploring the classic and modern research on impression formation across racial group divides.

DOI: 10.4324/9781003045687-16

Foundational Research

Asch's seminal research on "Forming Impressions of Personality" (Asch, 1946) was among the first to study impression formation in a controlled laboratory setting. In the original publication, participants from 10 studies read different lists of traits (e.g., "industrious", "practical", "intelligent") pertaining to a target person. For example, one list of traits included "warmth", while another list included "coldness". Across various studies, participants wrote down their open-ended impression of the target person, picked a trait that they deemed most likely to represent the target person, or ranked the traits in order of importance for their impression. Based on the results, Asch concluded that people form unitary, coherent impressions of others, particularly surrounding interpersonal warmth. For example, "a warm intelligent person is wise, while a cold intelligent person is sly" (Fiske et al., 2007, p. 78).

In the decades that followed, several scholars built upon Asch's work to examine how people explicitly form impressions of others. Heider (1958) suggested that sharing a broad commonality with the target person matters, creating a grouping with the target person or party that is characterized by increased liking and desire to affiliate. Rosenberg et al. (1968) examined how certain traits sort into clusters by asking participants to sort 64 traits into categories thought to be associated with a target person. Through multidimensional scaling, he identified two primary dimensions of person perception: social good-bad and intellectual good-bad. Other scholars explored how contextual factors such as lighting or clothing (Zebrowitz-McArthur, 1981) and physical appearance cues (Eagly et al., 1991; Zebrowitz, 1996) impact judgements of others. Kelley's attribution model (1967, 1973) explored the perceiver's use of both schemas and the rational consideration of evidence. There is a long line of research that examines the process of explicitly evaluating others.

Forming impressions of others often occurs implicitly—somewhat automatically, and often beyond conscious awareness (Uleman et al., 2005). Much foundational work explores spontaneous trait inference, defined as "unintended, unconscious, and relatively effortless inferences of traits" (Uleman et al., 2008, p. 331). These inferences are based on numerous cues, including faces and behaviors. Faces are of particular import. From faces, we distinguish others, inferring their social group membership (most efficiently, age, gender, and race; Uleman et al., 2008) and ascribing presumed-relevant traits. People infer personality traits from others' faces in as quickly as 100 milliseconds (Willis & Todorov, 2006). People also infer traits from others' behaviors, attending to trait-implying behavior across contexts (Uleman & Moskowitz, 1994), quickly labeling people as competent (incompetent) or warm (cold), with little consideration of their situations or personal histories.

Spontaneous trait inferences have numerous implications. They can provide input for slower, more intentional trait inferences (Gilbert, 1998), which are "shaped by current motives and the demands of communication"

(Zárate et al., 2001). Spontaneous trait inferences can also influence rapid unrelated judgments (see Moskowitz & Roman, 1992; see Uleman, 1987 and Uleman et al. 1996, for more details). Importantly, spontaneous trait inferences can predict real-world outcomes. For example, inferring competence from political candidates' faces predicts real-world election outcomes and margins of victory (Todorov et al., 2005). Finally, people not only infer traits; they also infer others' goals, values, and beliefs (Leslie, 1987; Premack, 1990), potentially impacting affiliation or cooperation with others.

Impression Formation Across Groups

I will next discuss impression formation across racial group divides, exploring the history of stereotyping research, modern examinations of stereotype content, and the implications of stereotyping for interracial attitudes and behaviors.

Stereotypes

Group dynamics researchers have long studied inferred attributes of social groups, or stereotypes. Stereotypes have been defined as associations between social groups and the content of specific attributes (Dupree et al., 2021). Social scientists have documented stereotypes for nearly one hundred years. The earliest of this work emphasized the descriptive content of stereotypes (e.g., Devine & Elliot, 1995; Gilbert, 1951; Karlins et al., 1969; Katz & Braly, 1933). Katz and Braly (1933) asked (White) Princeton undergraduates to check which of 84 different adjectives described racial, ethnic, and national groups. Researchers have since repeated this experimental procedure twice with Princeton undergraduates, finding that many of their stereotypes persisted (e.g., Italians were consistently viewed as passionate and Chinese as intelligent) (Bergsieker et al., 2012). Later, the rise of feminist movements in the 1970s saw the documentation of gender stereotypes in addition to racial/ethnic stereotypes, including the stereotype that women are more emotional than men—now thought to be one of the strongest gender stereotypes in Western culture (Shields, 2002). The late 20th century saw a growing interest in the process of stereotyping—how stereotypes form, why they persist, and how they influence attitudes and behavior. I review those trends in more detail below.

Spontaneity of Stereotypes

Stereotyping has clear connections to inferences of individuals. Much like the cognitive revolution in social psychology of the late 20th century gave birth to research on spontaneous trait inference, it also revitalized stereotyping research, which began to focus on the cognitive processes through which stereotyping occurs. Until then, the focus had primarily been on describing various labels applied to social groups (descriptive stereotypes) and exploring

whether these stereotypes were seen positively or negatively (prescriptive stereotypes). However, dovetailing with work on spontaneous trait inferences, social cognitive researchers began to study the spontaneity of stereotyping. A spontaneous categorization approach (Macrae & Bodenhausen, 2000) emerged, with consensus that prototypes represent social category members as average, ideal, or extreme. These social categories are activated spontaneously and rapidly, but not necessarily automatically—these processes can be interrupted, as I'll discuss below. People quickly categorize individuals as members of social groups (social categorization), infer associated traits (stereotype activation), and apply them to social group members (stereotype application).

Similar to spontaneous trait inference, stereotyping is typically activated by exposure to social group members' faces and behaviors. People exposed to targets with more phenotypically-Black facial features experience stronger activation of Black stereotype concepts (Eberhardt et al., 2004). Individuals with more phenotypically-Black features are seen as having more stereotypically-Black attributes, even if those individuals are categorized as White (Blair et al., 2002; Maddox, 2004). Outgroup members' behaviors also activate stereotypes, including perceived traits or values. For example, Black Americans infer White Americans' egalitarianism after reading their written statements or seeing how they describe themselves (Dupree & Foster-Gimbel, 2022; Jacoby-Senghor et al., 2021).

The rapid processes of stereotype activation and stereotype application can be interrupted by motivation and information (see Fiske & Taylor, 2013, for a review). People with the explicit motivation to avoid stereotyping group members can have some success (e.g., Kunda & Spencer, 2003), as can those who are trained to override stereotypes (Burns et al., 2017; Gawronski et al., 2008; Kawakami et al., 2000; Kawakami et al., 2005; Woodcock & Monteith, 2013). Additional research suggests that taking group members' perspectives can reduce stereotyping (Galinsky & Moskowitz, 2000; Vescio et al., 2003), though evidence on this is rather mixed (see Skorinko & Sinclair, 2013). Finally, according to Self-Regulation of Prejudice models, stereotyping can be overridden by being reminded of a better self and incentivized to judge others' accurately (Monteith, 1993; Monteith et al., 2002; Monteith et al., 2009). Triggering of egalitarian goals to which a person is committed can do more than override stereotypes, but inhibit them (e.g., Moskowitz et al., 1999; Moskowitz & Li, 2011).

Stereotype Content

In the early 21st century, stereotyping researchers refocused on stereotype content, converging upon the notion that two-dimensions dominate person and group perception. According to the Stereotype Content Model, upon encountering others, people must first determine their intent (warmth) and their ability to carry out their intent (competence). These two dimensions of warmth and competence broadly capture the prevalent stereotypes toward

various social groups. In the original study, Fiske and colleagues (2002) asked participants to rate 23 groups in society based on how society views each group on warmth and competence. Here, the findings for person- and group-based impressions differ. On an individual level, these two dimensions tend to correlate positively: people anticipate that others will be either high or low in both dimensions (Fiske & Taylor, 2013; Rosenberg et al., 1968). However, when judging social groups, researchers find a negative relationship between perceived warmth and competence: People tend to judge social groups as high in either warmth or competence, but not both (Fiske et al., 2002). Even when perceivers describe a group as high in one dimension, people infer that this means they are low in the other (Bergsieker et al., 2012).

Different names denote these two dimensions of group impressions, including warmth versus competence (Cuddy et al., 2007; Fiske et al., 2002) and agency versus communality (Abele, 2003; Bakan, 1966), but the core components of this two-dimensional framework are largely agreed upon (Abele & Wojciszke, 2007; Abele et al., 2008; Fiske, 2018). Various scholars have since refined and clarified theory, merging, in the process, research on person and group perception. Some scholars defined different facets of warmth and competence, such as dividing warmth into sociability and morality (Ellemers, 2017; Goodwin, 2015) and dividing competence into ability and assertiveness (Abele et al., 2016). Others added new stereotyping dimensions, such as beliefs (e.g., political orientation) (Koch et al., 2016). Scholars have also determined that perceptions of warmth are primary in person and group perception, reinforcing the notion that individuals prioritize determining others' intent before concerning themselves with competence (Wojciszke et al., 1998; for a review see Wojciszke, 2005). The rapid testing and refinement of this two-dimensional theory demonstrated how much can be learned and accomplished by incorporating both person and group-based impression formation research. Unfortunately, much of this work still concerned itself primarily with assessing perceptions of groups and individuals by White participants; recent work has only just begun to remedy this issue. (I will discuss this further in the upcoming *Future Directions*section.)

Intersectional Stereotypes

More recently, scholars have begun to move beyond only examining single stereotypes, which has traditionally focused on White-Black dynamics. Scholars have also begun to take an intersectional perspective, describing the specific stereotypes of group members at the intersection of multiple identities (e.g., race and gender; Purdie-Vaughns & Eibach, 2008; Sesko & Biernat, 2010). For example, my own work (Dupree et al., 2021) has examined stereotyping at the intersection of race and status. Recruiting thousands of White and Black participants, I found that White Americans are associated with high status while Black Americans are associated with low status. I measured these race-status associations in multiple ways. More direct

measures included a rank-based measure (how high perceivers rank White Americans and Black Americans on a social status ladder) and an attribute-based measure (how high perceivers rate White Americans and Black Americans on status-relevant attributes like "wealthy", "powerful", or "high-status"). The more indirect, job-based measure tested how likely people are to guess that White Americans hold high-status jobs (e.g., doctor, lawyer) and Black Americans hold low-status jobs (e.g., cashier, janitor). Both White and Black Americans held race-status associations across all three measures; they were more likely to associate White Americans with high-status and Black Americans with low-status. For White Americans, the more indirect, jobs-based race-status associations predicted more anti-Black prejudice, support for inequality (social dominance orientation; Ho et al., 2015), belief in meritocracy, less support for equalizing policies, and rejection of Black applicants seeking a high-status job. For White and Black Americans, the more direct rank- or attribute-based measures predicted less anti-Black prejudice, less support for inequality, and more support for equalizing policies. Examining intersectional stereotypes among diverse participants can illuminate how specific stereotypes maintain—or mitigate—inequality.

Implications for Attitudes and Behaviors

Stereotypes have crucial implications for individuals' attitudes and behaviors, along with groups' status and power in society. Throughout the 20th century, the dominant view of stereotyping was its uniform negativity. However, the late 20th century saw researchers begin to systematically examine how attitudes toward different groups vary based on the traits associated with these groups. Groups seen as higher in competence are also seen as higher in status (Fiske et al., 2002). This relationship between groups' perceived competence and their perceived status has been replicated worldwide, across 37 different countries (Durante et al., 2013). Stereotypical perceptions of groups' competence directly correspond to group status—particularly in unequal societies. Thus, stereotypes can maintain and justify inequality.

Stereotypes predict emotional prejudices, which in turn predict behavior (Cuddy et al., 2007). In fact, the proximate cause of much intergroup behavior is thought to be emotions, not cognition (Dovidio et al., 1996; Tropp & Pettigrew, 2005; Talaska et al., 2008). Consistent with this idea, researchers found that perceived warmth and competence associated with each group significantly predicted emotions and behavioral intent toward these groups (Cuddy et al., 2007). The BIAS map of intergroup behavior examines how intergroup stereotypes directly shape emotions toward different social groups, eliciting certain behaviors. For example, groups seen as warm and competent are more likely to be admired, prompting helping behaviors. In contrast, groups seen as lacking in warmth and competence are more likely to be viewed with disgust, prompting harmful behaviors. My own work further supports the notion that stereotypes correspond to behaviors; race-status stereotypes can

directly predict preferences for policies that reduce inequality and support (or rejection) of Black or White job applicants (Dupree et al., 2021).

Impression Management

People not only form impressions of others, but they also concern themselves with the impressions others form about them, and this can impact how they present themselves to others (Leary, 1995; Goffman, 1959). Thus, impression management and impression formation go hand-in-hand. A full understanding of impression formation requires an understanding of how people behave in order to elicit certain impressions from others. Unfortunately, impression formation research has traditionally neglected to consider how it intersects with impression management, largely ignoring the perspective of the actor, who is actively involved in shaping perceivers' impressions. Impression formation research has also typically ignored the roles of perceivers' and actors' social identities—including the attitudes, values, and ideologies that often come with these identities (see Dupree & Kraus, 2021)—in shaping how people perceive others. In the following section, I bring these bodies of work together, exploring the intersection of impression management and impression formation across racial divides. I first briefly describe foundational research on impression management before examining how impression management is impacted by stereotypes, thus influencing interaction goals, behaviors, and, ultimately, perceivers' impressions.

Foundational Research

When people interact with others, they adopt a wide variety of goals (Jones & Thibaut, 1958), prompting a need to manage others' impressions across dynamic social contexts. As fundamentally social animals, people are especially motivated to belong (Baumeister & Leary, 1995). People generally wish to be liked and respected by others (Baumeister, 1982), and they behave accordingly to meet these goals. They may smile and laugh at dinner parties to be seen as warm, and they may discreetly mention accomplishments to be seen as competent. Impression management (or self-presentation; Goffman, 1959) can be strategic. Thus, impression management represents a form of social influence, whereby one person (the actor) attempts to influence, or gain power over, another (the perceiver) (Jones, 1990; see also Tedeschi & Norman, 1985). As with many goals, impression management goals are most salient when under threat. When external rewards are at stake—be they personal or professional—impression management goals become highly salient (Buss & Briggs, 1984; Leary & Kowalski, 1990; Schlenker, 1980). For example, people are especially motivated to manage others' impression of them when at a job interview or on a first date. People are also especially likely to engage in impression management when they are being monitored, such as during videotaped interactions (Carver & Scheier, 1985; Scheier & Carver, 1982). Though

classic (and much modern) impression management research has paid little attention to the racial groups that actors belong to, racial identities—and the stereotypes that are associated with them—play an important role in impression management. Below, I explain how.

Impression Management Across Groups

Meta-stereotypes

Impression management in diverse group contexts are strongly influenced by meta-stereotypes. Meta-stereotypes activate during interactions with racial outgroup members, prompting specific interaction goals and behaviors (impression management) that can change how people are perceived by others (impression formation). Meta-stereotypes are defined as "a person's beliefs regarding the stereotypes that outgroup members hold about them" (Vorauer et al., 1998), and they vary dependent on the outgroup in question. They are distinct from self-stereotypes (Hogg & Turner, 1987)—the stereotypes a person holds about other groups—in that they have a distinctly relational component. Self-stereotypes can be somewhat relational—recall intergroup image theory, which suggests that stereotypes about outgroups predict prejudice based on the relationship between the target outgroup and the perceiver's ingroup (Alexander et al., 1999). However, meta-stereotypes are built upon one's sense of the outgroup's impressions of their ingroup—and these meta-stereotypes tend to be rather negative (Vorauer et al., 1998). Moreover, much like stereotypes (Fiske et al., 2002), meta-stereotypes are often ambivalent. For example, White Canadians and White Americans have the meta-stereotype that racial minorities view them as high status, but bigoted: competent, but cold (Vorauer et al., 1998; Vorauer et al., 2000). In contrast, Black Americans have the meta-stereotype that White Americans view them as low in competence (e.g., Krueger, 1996).

Interpersonal Goals

Meta-stereotypes directly relate to interpersonal goals in diverse settings; these goals go on to dictate how people behave with racial outgroup members. Of note, meta-stereotypes are not always accurate—people are not always right about how outgroup members view their ingroup. Though, in the prior example, Black Americans are indeed stereotyped as middling in warmth —"fun-loving" (Allport, 1954) or "happy-go-lucky" (Katz & Braly, 1933)—but low in competence— "lazy", "ignorant", and "low in intelligence" (Dupree et al., 2020; Devine & Elliot, 1995; Weaver, 2007)—by White Americans. Regardless of their accuracy, meta-stereotypes are theorized to activate impression management goals that directly impact intergroup behaviors. For example, in interracial settings, Bergsieker et al. (2010) found that White Americans had the goal to be seen as warm or likeable by racial minorities—thus, disconfirming White Americans' meta-stereotype

that racial minorities see them as competent but cold. In contrast, Latinx and Black Americans had the goal to be seen as competent or worthy of respect, thus countering racial minorities' meta-stereotype that White Americans see them as incompetent (Bergsieker et al., 2010).

Interpersonal Behavior

These interracial impression management goals—and the meta-stereotypes theorized to drive them—are thought to differentially affect interracial behavior. Recent work finds that White liberals, unlike conservatives, are more interested in racial equality and affiliating with racial minorities (Eastwick et al., 2009; Ho et al., 2015; Graham et al., 2009; Jost et al., 2004; Kteily et al., 2019); White liberals' affiliation motivation produces a competence downshift when they interact with racial minorities. White liberals describe themselves as less competent with a Black (versus White) interaction partner (Dupree & Fiske, 2019). Across six studies featuring thousands of participants, I found that White Americans who are lowest in self-reported conservatism and support for inequality (social dominance orientation; Ho et al., 2015) use fewer words indicating competence in a work task with a Black partner, describe themselves as less competent when completing a personality questionnaire to be shown to a Black partner, and use fewer words related to competence in written introductions to a Black partner. I also found this effect outside of the lab, in a real-world setting. Collecting campaign speeches delivered by White Democratic and Republican presidential candidates over 25 years revealed that White Democratic candidates used fewer words related to competence (e.g., "assertive", "competitive") when speaking to a mostly-minority audience (e.g., NAACP, National Council of La Raza) than a mostly-White audience (e.g., Americans for Prosperity).

This subtle but reliable effect suggests that White liberals distance themselves from stereotypes that depict them as competent, but cold—ironically, enough, approaching stereotypes that depict Black Americans as incompetent. This phenomenon was unique to White liberals—who are more affiliative toward racial minorities and more likely to be concerned about negative meta-stereotypes—consistent with theorizing that meta-stereotypical concerns activate affiliative impression management goals (Vorauer et al., 1998) and that interracial impression management goals can drive interracial behavior (Bergsieker et al., 2010). Thus, this behavioral phenomenon may be rooted in both distancing from White Americans' meta-stereotypes depicting the ingroup as bigoted (Vorauer et al., 1998) and approaching their stereotypes depicting Black Americans as less competent than the ingroup (Dupree et al., 2021). The competence downshift thus provides an example of how White Americans (specifically, White liberals) alter their behavior to meet an impression management goal—that of appearing more affiliative and less dominant.

Another line of work found that White Americans also downshifted conservatism toward racial minorities, describing themselves as less

supportive of conservative policies and candidates when discussing politics with a Black (versus White) interaction partner. Mediation analyses revealed this effect was driven by stereotypical inferences of racial groups' values—specifically, the stereotype that a Black (versus White) interaction partner is more egalitarian (Dupree & Foster-Gimbel, 2021). White participants described themselves as less conservative with a Black (versus White) interaction partner via perceptions that a Black (versus White) partner is more egalitarian. Stereotypes clearly play a role in White Americans' impression management across racial divides.

Low-status group members also engage in impression management by countering meta-stereotypes. Black and Latinx conservatives tend to be more interested in affiliating with the high-status outgroup and less interested in affiliating with the low-status ingroup than liberals (Bejarano, 2013; Eastwick et al., 2009; Ho et al., 2015; Jost & Thompson, 2000; Stern & Axt, 2018); as a result, Black and Latinx conservatives upshift competence relative to liberals in mostly-White settings. For example, my own work recently found that conservative Black and Latinx politicians used more words related to high power and ability than liberals when speaking in Congress and when posting on social media; in addition, conservative Black Americans used more words related to high status than liberals did when introducing themselves to an interaction partner (Dupree, 2021a). Women in leadership positions also upshift competence in mostly-male professional settings. Female politicians—specifically, White women—used more words related to high power than same-race men when speaking in Congress and when posting on social media (Dupree, 2021b). Thus, White women in politics reversed stereotypes that depict White women as more submissive than men (Purdie-Vaughns & Eibach, 2008; Sesko & Biernat, 2010; Livingston et al., 2012) and therefore unsuitable to leadership (Rosette et al., 2016). This effect was specific to White women, for Black and Latinx women are more less likely to be stereotyped as submissive; rather, they are stereotyped as loud, angry, and aggressive (Purdie-Vaughns & Eibach, 2008; Rosette et al., 2016). Scholars must also consider intersectionality when examining impression management in diverse settings.

Implications for Impression Formation

As noted, impression management and impression formation go hand-in-hand. People engage in impression management—often countering negative stereotypes in the process—which directly impacts how they are seen. My own and others' research empirically supports this link between strategic, counter-stereotypical self-presentation and impression formation. Sure enough, when actors reverse stereotypes by engaging in counterstereotypical behavior, perceivers' impressions change. The more Black or Latinx conservatives (versus liberals) upshift competence in Congress, the more journalists represent them as high in power (Dupree, 2021b). In addition, the more

female politicians—specifically, Black and Latina women—reference power in Congress, the more journalists use powerful words in editorials about them (Dupree, 2021b). People notice others' verbal behavior—especially when others are in positions of power and when they are members of traditionally disadvantaged social groups. This is not always to the actor's advantage. For women in leadership, behaving powerfully can prompt well-studied backlash effects (Bowles et al., 2007; Heilman & Okimoto, 2007; LaFrance, 1992; Livingston et al., 2012; Okimoto & Brescoll, 2010; Rudman, 1998; Rudman & Glick, 1999; Rudman et al., 2012; Williams & Tiedens, 2016). Indeed, the more female politicians reference high power in a given Congressional term, the lower vote share they receive in the subsequent election (Dupree, 2021b).

Future Research

The future of impression formation and impression management research is bright, particularly at the intersection of impression formation, impression management, and stereotyping. Much remains to be learned, For example, do White liberals, women, or racial minorities reverse meta-stereotypes by altering their tone of voice or shifting their physical behavior (e.g., eye gaze, posture)? Such non-verbal impression management behaviors can directly impact outgroup members' impression formation. For example, Dovidio and colleagues (2002) found that, in interracial interactions, White Americans' nonverbal friendliness (as rated by coders) predicted Black Americans' ratings of White Americans' bias. Verbal friendliness, in contrast, more strongly predicted how biased White Americans thought they were (rather than how biased Black Americans thought they were). Researchers can also further explore the role of threat in impression management goals theorized to drive interracial behaviors. For example, are White liberals who downshift competence toward racial minorities—or Black conservatives who upshift competence toward White Americans—more anxious about these interactions, or are they less anxious and more confident due to their counter-stereotypical behavior? Anxiety predicts how interested people are in intergroup contact, and, when they do engage in these interactions, how they are perceived by outgroup members (e.g., West et al., 2014). Thus, further examining the role of anxiety in impression management and impression formation across racial divides is a promising area of study.

Stereotyping research is also starting to home in on the specific stereotypes associated with people who hold multiple social identities, and this can inform investigations of impression management. Everyone is a member of not one, but multiple social groups, including gender, race, age, social class, sexual orientation, and others. Exploring stereotypes at the intersection of social identities thus constitutes a ripe area for future study. The past decade has seen scholars begin to explore stereotype content at the intersection of race and social class (Brown-Iannuzzi et al., 2019; Dupree et al., 2020; Freeman et al., 2011; Kahn et al., 2009; Kunstman et al., 2016; Lei &

Bodenhausen, 2017; Moore-Berg & Karpinski, 2019) and race and gender (Hall et al., 2019; Purdie-Vaughns & Eibach, 2008; Sesko & Biernat, 2010; Livingston et al., 2012; Rosette et al., 2016). For example, White women upshift competence relative to men in professional settings, reversing stereotypes that depict them as submissive (Dupree, 2021b), but do Black and Latina women upshift warmth, thus reversing stereotypes that depict them as loud or angry? Moreover, Black and Latinx conservatives upshift competence relative to liberals in mostly-White settings, reversing stereotypes that depict them as low-status or incompetent (Dupree, 2021a), but do working class Latinx or Black Americans also upshift competence upshift relative to their upper-class counterparts, reversing stereotypes that depict them as uneducated? Determining the unique stereotypes associated with multiple social identities and how these stereotypes can elicit specific impression management and behaviors will give us a broader sense of how impression formation and impression management operates among a wider swath of individuals.

Relatedly, stereotyping research is also beginning to branch out from Black-White and female-male dynamics, for these are not the only social identities with important consequences for perceivers' and actors' attitudes, cognition, and behaviors. Researchers have begun to examine the stereotypes associated with other racial groups, including Asian Americans and Latino/a/x (e.g., Zou & Cheryan, 2017) and other stigmatized groups, including LGBTQ individuals (Madon, 1997), obese individuals (Hunger et al., 2015), the poor (Lott & Bullock, 2001), the elderly (Abrams et al., 2016; North & Fiske, 2013), and those with physical or mental disability (Dunn, 2010; Rohmer & Louvet, 2018). Stereotyping and impression management models are ultimately incomplete without considering additional and intersecting social identities, including those that aren't immediately visible, such as social class, sexual orientation, and disability.

Finally, much of the scholarship described in this chapter has been decidedly U.S.-focused—specifically, White American focused—emphasizing the perceptions and behaviors of White Americans (typically highly educated, liberal-leaning students). Scholarship that relies on WEIRD (Western, Educated, Industrialized, Rich, and Democratic) populations is hardly representative of the global population (Henrich et al., 2010; Henry, 2008; Sears, 1988). As social scientists begin to reckon with a lack of diversity in its samples, methods, and journals (Dupree & Kraus, 2021; Roberts et al., 2020), future work should draw upon more representative populations both within and outside of the United States, incorporating participants and perspectives from more than one swath of humanity.

Conclusion

Social cognitive researchers have long studied impression formation, the dimensions that perceivers use when forming first impressions, the spontaneity and universality of these impressions, and their implications for

emotions and behavioral tendencies. However, this work has evolved largely separately from research on group dynamics, despite the obvious areas of overlap. Examining how the social group membership of perceiver and actor affect impression formation and impression management is a crucial area of study that provides a more holistic and generalizable understanding of these phenomena. Research on stereotyping and impression management across group divides dovetails with classic research on impression formation and impression management, making for scholarship that is more cognizant of and applicable to the diverse world in which we all live.

References

Abele, A. E. (2003). The dynamics of masculine-agentic and feminine-communal traits: Findings from a prospective study. *Journal of Personality and Social Psychology*, *85*, 768–776.

Abele, A. E., Cuddy, A. J. C., Judd, C. M., & Yzerbyt, V. (2008). Fundamental dimensions of social judgment. *European Journal of Social Psychology*, *38*, 1063–1065.

Abele, A. E., & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *Journal of Personality and Social Psychology*, *93*, 751–763. doi: 10.1037/0022-3514.93.5.751

Abele, A. E., Hauke, N., Peters, K., Louvet, E., Szymkow, A., & Duan, Y. (2016). Facets of the fundamental content dimensions: Agency with competence and assertiveness—Communion with warmth and morality. *Frontiers in Psychology*, *7*, 1810.

Abrams, D., Swift, H. J., & Drury, L. (2016). Old and unemployable? How age-based stereotypes affect willingness to hire job candidates. *Journal of Social Issues*, *72*(1), 105–121. doi: 10.1111/josi.12158

Alexander, M. G., Brewer, M. B., & Herrmann, R. K. (1999). Images and affect: A functional analysis of out-group stereotypes. *Journal of Personality and Social Psychology*, *77*, 78–93.

Allport, G. W. (1954). *The nature of prejudice*. Reading, MA: Addison-Wesley.

Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology*, *41*, 258–290.

Bakan, D. (1966). *The duality of human existence. An essay on psychology and religion*. Chicago: Rand Mcnally.

Baumeister, R. F. (1982). A self-presentational view of social phenomena. *Psychological Bulletin*, *91*, 3–26.

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, *117*, 497–529.

Bejarano, C. E. (2013). *The Latino Gender Gap in U.S. Politics*. Routledge.

Bergsieker, H. B., Leslie, L. M., Constantine, V. S., & Fiske, S. T. (2012). Stereotyping by omission: Eliminate the negative, accentuate the positive. *Journal of Personality and Social Psychology*, *102*, 1214–1238.

Bergsieker, H. B., Shelton, J. N., & Richeson, J. A. (2010). To be liked versus respected: Divergent goals in interracial interactions. *Journal of Personality and Social Psychology*, *99*, 248–264.

Blair, I. V., Judd, C. M., Sadler, M. S., & Jenkins, C. (2002). The role of Afrocentric features in person perception: Judging by features and categories. *Journal of Personality and Social Psychology*, *83*, 5–25.

Bowles, H. R., Babcock, L., & Lai, L. (2007). Social incentives for gender differences in the propensity to initiate negotiations: Sometimes it does hurt to ask. *Organizational Behavior and Human Decision Processes*, *103*, 84–103. doi: 10.1016/j.obhdp.2006.09.001.

Brown-Iannuzzi, J. L., Cooley, E., McKee, S. E., & Hyden, C. (2019). Wealthy Whites and poor Blacks: Implicit associations between racial groups and wealth predict explicit opposition toward helping the poor. *Journal of Experimental Social Psychology*, *82*, 26–34.

Burns, M., Monteith, M., & Parker, L. (2017). Training away bias: The differential effects of counterstereotype training and self-regulation on stereotype activation and application. *Journal of Experimental Social Psychology*, *73*, 97–110.

Buss, A. H., & Briggs, S. R. (1984). Drama and the self in social interaction. *Journal of Personality and Social Psychology*, *47*, 1310–1324.

Carver, C. S., & Scheier, M. F. (1985). Aspects of self, and the control of behavior. In B. R. Schlenker (Ed.), *The self and social life* (pp. 146–174). McGraw-Hill.

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2007). The BIAS map: Behaviors from intergroup affect and stereotypes. *Journal of Personality and Social Psychology*, *92*, 631–648.

Devine, P. G., & Elliot, A. J. (1995). Are racial stereotypes really fading? The Princeton trilogy revisited. *Personality and Social Psychology Bulletin*, *11*, 1139–1150.

Dovidio, J. F., Brigham, J. C., Johnson, B. T., & Gaertner, S. L. (1996). Stereotyping, prejudice, and discrimination: Another look. In C. N. Macrae, C. Stangor & M. Hewstone (Eds.), *Stereotypes and stereotyping* (pp. 276–319). Guilford Press.

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology*, *82*, 62–68. doi: 10.1037/0022-3514.82.1.62.

Dupree, C. H. (2021a). Black and Latinx conservatives upshift competence relative to liberals in mostly white settings. *Nature Human Behavior*, *5*, 1652–1662. doi: 10.1038/s41562-021-01167-9

Dupree, C. H., (2021b). *Words of a leader: An intersectional analysis of gender differences in leaders' speech*. [Manuscript submitted for publication]. School of Management, Yale University.

Dupree, C. H., & Fiske, S. T. (2019) Self-presentation in interracial settings: The competence downshift by White liberals. *Journal of Personality and Social Psychology*, *117*(3), 579–604.

Dupree, C. H. & Foster-Gimbel, O. (2022). *Going for woke: White Americans downshift conservatism in interracial settings*. [Manuscript submitted for publication]. School of Management, Yale University.

Dupree, C. H., Torrez, B., Obioha, O., & Fiske, S. T. (2021). Race–status associations: Distinct effects of three novel measures among White and Black perceivers. *Journal of Personality and Social Psychology*, *120*(3), 601–625. doi: 10.1037/pspa0000257.

Dupree, C. H. & Kraus, M. K. (2021). Psychological science is not race neutral. *Perspectives in Psychological Science*, *17*(1), 270–275. doi: 10.1177/1745691620979820.

Dupree, C. H., Torrez, B., Obianuju, O., & Fiske, S. T. (2020). Race-status associations: Distinct effects of three novel measures among White and Black perceivers. *Journal of Personality and Social Psychology*, *120*(3), 601–625. doi: 10.1037/pspa0000257.

Dunn, D. S. (2010). The social psychology of disability. In R. G. Frank, M. Rosenthal, & B. Caplan, *Handbook of rehabilitation psychology* (pp. 379–390). American Psychological Association.

Durante, F., Fiske, S. T., Kervyn, N., Cuddy, A. J. C., Akande, A., Adetoun, B. E., Adewuyi, M. F., Tserere, M. M., Al Ramiah, A., Mastor, K. A., Barlow, F. K., Bonn, G., Tafarodi, R. W., Bosak, J., Cairns, E., Doherty, S., Capozza, D., Chandran, A., Chryssochoou1, X., Iatridis, T., Contreras, J. M., Costa-Lopes, R., González, R., Lewis, J. I., Tushabe, G., Leyens, J-Ph., Mayorga, R., Rouhana, N. N., Smith Castro, V., Perez, R., Rodríguez-Bailón, R., Moya, M., Morales Marente, E., Palacios Gálvez, M., Sibley, C. G., Asbrock, F., & Storari, C. C. (2013). Nations' income inequality predicts ambivalence in stereotype content: How societies mind the gap. *British Journal of Social Psychology*, *52*, 726–746.

Eberhardt, J. L., Goff, P. A., Purdie, V. J., & Davies, P. G. (2004). Seeing Black: Race, crime, and visual processing. *Journal of Personality and Social Psychology*, *87*, 876–893.

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good, but…: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin*, *110*, 109–128.

Eastwick, P. W., Richeson, J. A., Son, D., & Finkel, E. J. (2009). Is love colorblind? Political orientation and interracial romantic desire. *Personality & Social Psychological Bulletin*, *35*, 1258–1268.

Ellemers, N. (2017). *Morality and the regulation of social behavior: Groups as moral anchors*. Routledge.

Fiske, S. T. (2018). Stereotype content: Warmth and competence endure. *Current Directions in Psychological Science*, *27*, 67–73.

Fiske, S. T., & Taylor, S. E. (2013). *Social cognition: From brains to culture*. Sage.

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*, 878–902.

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: warmth and competence. *Trends in Cognitive Sciences*, *11*(2), 77–83.

Freeman, J. B., Penner, A. M., Saperstein, A., Scheutz, M., & Ambady, N. (2011). Looking the part: Social status cues shape race perception. *PLoS ONE*, 6(9), e25107. doi: 10.1371/journal.pone.0025107.

Galinsky, A. D., & Moskowitz, G. B. (2000). Perspective taking: Decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of Personality and Social Psychology*, *78*, 708–724.

Gawronski, B., Deutsch, R., Mbirkou, S., Seibt, B., & Strack, F. (2008). When "just say no" is not enough: Affirmation versus negation training and the reduction of automatic stereotype activation. *Journal of Experimental Social Psychology*, *44*, 370–377.

Goffman, E. (1959). *The presentation of self in everyday life*. Penguin.

Goodwin, G. P. (2015). Moral character in person perception. *Current Directions in Psychological Science*, *24*, 38–44.

Gilbert, G. M. (1951). Stereotype persistence and change among college students. *Journal of Abnormal and Social Psychology*, *46*, 245–254.

Gilbert, D. T. (1998). Ordinary personology. In D. T. Gilbert, S. T. Fiske & G. Lindzey (Eds.). *The Handbook of Social Psychology* (4th edn., Vol. II, pp. 89–150). McGraw-Hill.

Graham, J., Haidt, J., & Nosek, B. A. (2009) Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*, 1029–1046.

Hall, E. V., Hall, A. V., Galinsky, A. D., Phillips, K. W. (2019). MOSAIC: a model of stereotyping through associated and intersectional categories. *Academy of Management Review*, *44*, 643–672.

Heider, F. (1958). *The psychology of interpersonal relations*. Wiley.

Heilman, M. E., & Okimoto, T. G. (2007). Why are women penalized for success at male tasks?: The implied communality deficit. *Journal of Applied Psychology*, *92*, 81–92. doi: 10.1037/0021-9010.92.1.81

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*, 61–83.

Henry, P. J. (2008). College sophomores in the laboratory redux: Influ-ences of a narrow data base on social psychology's view of the nature of prejudice. *Psychological Inquiry*, *19*, 49–71.

Ho, A. K., Sidanius, J., Kteily, N., Sheehy-Skeffington, J., Pratto, F., Henkel, K. E., Foels, R. , & Stewart, A. L. (2015). The nature of social dominance orientation: Theorizing and measuring preferences for intergroup in- equality using the new SDO7 scale. *Journal of Personality and Social Psychology*, *109*(6), 1003–1028. doi: 10.1037/pspi0000033.

Hogg, M. A., & Turner, J. C. (1987). Intergroup behaviour, self-stereotyping and the salience of social categories. *British Journal of Social Psychology*, *26*, 325–340.

Hunger, J. M., Major, B., Blodorn, A., & Miller, C. T. (2015). Weighed down by stigma: How weight-based social identity threat contributes to weight gain and poor health. *Social and Personality Psychology Compass*, *9*(6), 255–268. doi: 10.1111/spc3.12172.

Jacoby-Senghor, D., Rosenblum, M., & Brown, N. D. (2021). Not all egalitarianism is created equal: Claims of nonprejudice inadvertently communicate prejudice between ingroup members. *Journal of Experimental Social Psychology*, *94*, 104104.

Jones, E. E. (1990). *Interpersonal perception*. W.H. Freeman and Co.

Jost, J. T., Banaji, M. R., & Nosek, B. A. (2004). A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. *Political Psychology*, *25*, 881–919.

Jones, E. E., & Thibaut, J. W. (1958). Interaction goals as bases of inference in interpersonal perception. In Taguiri, R. , & Petrullo, L. (Eds.), *Person perception and interpersonal behavior*, (pp. 151–178). Stanford: Stanford University Press.

Jost, J. T., & Thompson, E. P. (2000). Group-based dominance and opposition to equality as independent predictors of self-esteem, ethnocentrism, and social policy attitudes among African Americans and European Americans. *Journal of Experimental Social Psychology*, *36*, 209–232.

Kahn, K., Ho, A. K., Sidanius, J., & Pratto, F. (2009). The space between us and them: Perceptions of status differences. *Group Processes & Intergroup Relations*, *12*, 591–604.

Katz, D., & Braly, K. (1933). Racial stereotypes of one hundred college students. *Journal of Abnormal and Social Psychology*, *28*, 280–290.

Karlins, M., Coffman, T. L., & Walters, G. (1969). On the fading of social stereotypes: Studies in three generations of college students. *Journal of Personality and Social Psychology*, *13*, 1–16.

Kawakami, K., Dovidio, J. F., Moll, J., Hermsen, S., & Russin, A. (2000). Just say no (to stereotyping): Effects of training in the negation of stereotypic associations on stereotype activation. *Journal of Personality and Social Psychology*, *78*, 871–888.

Kawakami, K., Dovidio, J. F., & van Kamp, S. (2005). Kicking the habit: Effects of non- stereotypic association training and correction processes on hiring decisions. *Journal of Experimental Social Psychology*, *41*, 68–75.

Kelley, H. H. (1967). Attribution in social psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivation* (Vol. 15 pp. 192–238). University of Nebraska Press.

Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist*, *28*, 107–128.

Koch, A., Imhoff, R., Dotsch, R., Unkelbach, C., & Alves, H. (2016). The ABC of stereotypes about groups: Agency/socioeconomic success, conservative–progressive beliefs, and communion. *Journal of Personality and Social Psychology*, *110*(5), 675.

Kunda, Z., & Spencer, S. J. (2003). When do stereotypes come to mind and when do they color judgment? A goal-based theoretical framework for stereotype activation and application. *Psychological Bulletin*, *129*, 522–544.

Kunstman, J. W., Plant, E. A., & Deska, J. C. (2016). White? poor: Whites distance, derogate, and deny low-status ingroup members. *Personality and Social Psychology Bulletin*, *42*, 230–243.

Krueger, J. (1996). Personal beliefs and cultural stereotypes about racial characteristics. *Journal of Personality and Social Psychology*, *71*, 536–548.

Kteily, N. S., Rocklage, M. D., McClanahan, K., & Ho, A. K. (2019). Political ideology shapes the amplification of the accomplishments of disadvantaged vs. advantaged group members. *Proceedings of the National Academy of Sciences*, 201818545.

LaFrance, M. (1992). Gender and interruptions. *Psychology of Women Quarterly*, *16*, 497–512. doi: 10.1111/j.1471-6402.1992.tb00271.x.

Leary, M. R. (1995). *Self-presentation: Impression management and interpersonal behavior*. Westview Press.

Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin*, *107*, 34–47.

Lei, R. F., & Bodenhausen, G. V. (2017). Racial assumptions color the mental representation of social class. *Frontiers in Psychology*, 8.

Leslie, A. M. (1987). Pretense and representation: The origins of "theory of mind." *Psychological Review*, *94*, 412–426.

Livingston, R. W., Rosette, A. S., & Washington, E. F. (2012). Can an agentic Black woman get ahead? The impact of race and interpersonal dominance on perceptions of female leaders. *Psychological Science*, *23*, 354–358.

Lott, B., & Bullock, H. E. (2001). Who are the poor? *Journal of Social Issues*, *57*(2), 189– 206.

Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology*, *51*, 93–120.

Maddox, K. B. (2004). Perspectives on racial phenotypicality bias. *Personality and Social Psychology Review*, *8*, 383–401.

Madon, S. (1997). What do people believe about gay males? A study of stereotype content and strength. *Sex Roles*, *37*, 663–685.

Monteith, M. J. (1993). Self-regulation of prejudiced responses: Implications for progress in prejudice-reduction efforts. *Journal of Personality and Social Psychology*, *65*, 469–485.

Monteith, M. J., Ashburn-Nardo, L., Voils, C. I., & Czopp, A. M. (2002). Putting the brakes on prejudice: On the development and operation of cues for control. *Journal of Personality and Social Psychology*, *83*, 1029–1050.

Monteith, M. J., Lybarger, J. E., & Woodcock, A. (2009). Schooling the cognitive monster: The role of motivation in the regulation and control of prejudice. *Social and Personality Psychology Compass*, *3*, 211–226.

Moore-Berg, S. L., & Karpinski, A. (2019). An intersectional approach to understanding how race and social class affect intergroup processes. *Social and Personality Psychology Compass*, *13*, e12426.

Moskowitz, G. B., & Roman, R. J. (1992). Spontaneous trait inferences as self-generated primes: Implications for conscious social judgment. *Journal of Personality and Social Psychology*, *62*, 728–738.

Moskowitz, G. B., & Li, P. (2011). Egalitarian goals trigger stereotype inhibition: A proactive form of stereotype control. *Journal of Experimental Social Psychology*, *47*(1), 103–116. doi: 10.1016/j.jesp.2010.08.014.

Moskowitz, G. B., Gollwitzer, P. M., Wasel, W., & Schaal, B. (1999). Preconscious control of stereotype activation through chronic egalitarian goals. *Journal of Personality and Social Psychology*, *77*(1), 167–184. doi: 10.1037/0022-3514.77.1.167.

North, M. S., & Fiske, S. T. (2013). Act your (old) age: Prescriptive, ageist biases over succession, consumption, and identity. *Personality and Social Psychology Bulletin*, *39*, 720–734.

Okimoto, T. G., & Brescoll, V. L. (2010). The price of power: Power seeking and backlash against female politicians. *Personality and Social Psychology Bulletin*, *36*, 923–936. doi: 10.1177/0146167210371949.

Purdie-Vaughns V., & Eibach R. P. (2008). Intersectional invisibility: The distinctive advantages and disadvantages of multiple subordinate-group identities. *Sex Roles*, *59*, 377–391.

Premack, D. (1990). The infant's theory of self-propelled objects. *Cognition*, *36*, 1–16.

Roberts, S. O., Bareket-Shavit, C., Dollins, F. A., Goldie, P. D., & Mortenson, E. (2020). Racial inequality in psychological research: Trends of the past and recommendations for the future. *Perspectives on Psychological Science*, 15, 1295–1309.

Rohmer, O., & Louvet, E. (2018). Implicit stereotyping against people with disability. *Group Processes & Intergroup Relations*, *21*(1), 127–140.

Rosenberg, S., Nelson, C., & Vivekananthan, P. S. (1968). A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology*, *9*, 283–294.

Rosette, A. S., Koval, C. Z., Ma, A., & Livingston, R. (2016). Race matters for women leaders: Intersectional effects on agentic deficiencies and penalties. *Leadership Quarterly*, *26*, 429–445.

Rudman, L. A. (1998). Self-promotion as a risk factor for women: The costs and benefits of counterstereotypical impression management. *Journal of Personality and Social Psychology*, *74*, 629–645. doi: 10.1037/0022-3514.74.3.629.

Rudman, L. A., & Glick, P. (1999). Feminized management and backlash toward agentic women: The hidden costs to women of a kinder, gentler image of middle managers. *Journal of Personality and Social Psychology*, *77*, 1004–1010. doi: 10.1037/0022-3514.77.5.1004.

Rudman, L. A., Moss-Racusin, C. A., Glick, P., & Phelan, J. E. (2012). Reactions to vanguards: Advances in backlash theory. In Devine, P. G. , & Plant, E. A. (Eds.), *Advances in experimental social psychology*, (Vol. 45, pp. 167–227) .San Diego, CA: Elsevier. doi: 10.1016/B978-0-12-394286-9.00004-4.

Scheier, M. F. & Carver, C. S. (1982). Two sides of the self: One for you and one for me. In J. Suls & A. G. Greenwald (Eds.), *Psychological perspectives on the self* (Vol. 2, pp. 123–157). Lawrence Erlbaum Associates.

Shields, S. A. (2002). *Speaking from the heart: Gender and the social meaning of emotion*. Cambridge, U.K.: Cambridge University Press.

Sears, D. O. (1988). Symbolic racism. In P. A. Katz & D. A. Taylor (Eds.), *Eliminating racism: Profiles in controversy* (pp. 53–84). Plenum Press. 10.1007/978-1-4899-0818-6_4

Sesko, A. K., & Biernat, M. (2010). Prototypes of race and gender: The invisibility of Black women. *Journal of Experimental Social Psychology*, *46*, 356–360.

Schlenker, B. R. (1980). *Impression management: The self-concept, social identity, and interpersonal relationships*. Brooks/Cole.

Skorinko, J. L., & Sinclair, S. A. (2013). Perspective taking can increase stereotyping: The role of apparent stereotype confirmation. *Journal of Experimental Social Psychology*, *49*, 10–18.

Stern, C., & Axt, J. R. (2018). Group status modulates the associative strength between status quo supporting beliefs and anti-Black attitudes. *Social Psychology and Personality Science*, *10*, 946–956.

Talaska, C. A., Fiske, S. T., & Chaiken, S. (2008). Legitimating racial discrimination: A meta-analysis of the racial attitude-behavior literature shows that emotions, not beliefs, best predict discrimination. *Social Justice Research: Social Power in Action*, *21*, 263–296.

Tedeschi, J. T. & Norman, N. (1985). Social power, self-presentation, and the self. In B. R. Schlenker (Ed.), *The self and social life* (pp. 293–322). McGraw-Hill.

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science*, *308*, 1623–1626.

Tropp, L. R., & Pettigrew, T. F. (2005). Relationships between inter-group contact and prejudice among minority and majority status groups. *Psychological Science*, *16*, 951–957.

Uleman, J. S. (1999). Spontaneous versus intentional inferences in impression formation. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 141–160). Guilford.

Uleman, J. S. (1987). Consciousness and control. *Personality and Social Psychology Bulletin*, *13*(3), 337–354. doi: 10.1177/0146167287133004.

Uleman, J. S., Blader, S. L., & Todorov, A. (2005). Implicit impressions. In R. R. Hassin, J. S. Uleman & J. A. Bargh (Eds.), *The new unconscious* (pp. 362–392). Oxford University Press.

Uleman, J. S., & Moskowitz, G. B. (1994). Unintended effects of goals on unintended inferences. *Journal of Personality and Social Psychology*, 66, 490–501.

Uleman, J. S., Newman, L. S., & Moskowitz, G. B. (1996). People as flexible interpreters: evidence and issues from spontaneous trait inference. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 28, pp. 211–279). Academic Press.

Uleman, J., Saribay, S., Gonzalez, C. (2008). Spontaneous inferences, implicit impressions, and implicit theories. *Annual Review of Psychology*, *59*, 329–360.

Vescio, T. K., Sechrist, G. B., & Paolucci, M. P. (2003). Perspective taking and prejudice reduction: The mediational role of empathy arousal and situational attributions. *European Journal of Social Psychology*, *33*, 455–472.

Vorauer, J. D., Hunter, A., Main, K., & Roy, S. (2000). Concerns with evaluation and meta-stereotype activation. *Journal of Personality and Social Psychology*, *78*, 690–707.

Vorauer, J. D., Main, K. J., & O'Connell, G. B. (1998). How do individuals expect to be viewed by members of lower status groups? Content and implications of meta-stereotypes. *Journal of Personality and Social Psychology*, *75*, 917–937.

Weaver, C. N. (2007). The effects of contact on the prejudice between Hispanics and non-Hispanic Whites in the United States. *Hispanic Journal of Behavioral Sciences*, *29*, 254–274.

West, T. V., Pearson, A. R., & Stern, C. (2014). Anxiety perseverance in intergroup interaction: When incidental explanations backfire. *Journal of Personality and Social Psychology*, *107*(5), 825–843. doi: 10.1037/a0037941.

Williams, M. J., & Tiedens, L. Z. (2016). The subtle suspension of backlash: A meta-analysis of penalties for women's implicit and explicit dominance behavior. *Psychological Bulletin*, 142, 165–197.

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, *17*, 592–598.

Woodcock, A., & Monteith, M. J. (2013). Forging links with the self to combat implicit bias. *Group Processes & Intergroup Relations*, *16*, 44–461.

Wojciszke, B. (2005). Morality and competence in person- and self-perception. *European Review of Social Psychology*, *16*, 155–188.

Wojciszke, B., Bazinska, R., & Jaworski, M. (1998). On the dominance of moral categories in impression formation. *Personality and Social Psychology Bulletin*, *24*, 1251–1263.

Zárate, M. A., Uleman, J. S., & Voils, C. I. (2001). Effects of culture and processing goals on the activation and binding of trait concepts. *Social Cognition*, *19*, 295–323.

Zebrowitz, L. A. (1996). Physical appearance as a basis of stereotyping. In C. N. Macrae, C. Stangor & M. Hewstone (Eds.), *Stereotypes and stereotyping* (pp. 79–120). Guilford.

Zebrowitz-McArthur, L. A. (1981). What grabs you? The role of attention in impression formation and causal attribution. In E. T. Higgins, C. P. Herrman & M. P. Zanna (Eds.), *Social cognition* (pp. 201–246). Erlbaum.

Zou, L. X., & Cheryan, S. (2017). Two axes of subordination: A new model of racial position. *Journal of Personality and Social Psychology*, *112*, 696–717. doi: 10.1037/pspa0000080.

15 Understanding Guilt-by-Association: A Review of the Psychological Literature on Attitude Transfer and Generalization

Kate A. Ratliff
University of Florida

In his 1573 book, *The Garden of Pleasure*, the British writer James Sandford wrote: "He that goeth to bedde wyth Dogges, aryseth with fleas." This is the first-known English-language record of a saying my grandmother repeated often: "If you lie down with dogs, you'll get up with fleas." This refers, of course, to the idea that we should be careful of the company we keep, either because we may be led astray by our disreputable associates, or, more relevant to this chapter, because other people might *think* we have been led astray, and our reputation will suffer as a result. Some of us might decide our love for dogs is worth the risk of fleas, but my grandmother's advice is not entirely incorrect—stimuli often do take on meaning in the absence of direct experience through their relationships with other stimuli (Shanks, 1995). A growing body of research on impression formation speaks to *attitude transfer*, an effect by which evaluations of one individual transfer to another individual who is associated[1] in some way (Ratliff [Ranganath] & Nosek, 2008).

Attitude transfer is a specific instantiation of *generalization*, a principle in classical learning theory by which a response conditioned to one stimulus generalizes to other stimuli that are similar (see Till & Priluck, 2000, for a review). In early studies of classical conditioning, Pavlov (1927) found that dogs would salivate (a conditioned response) at the sound of a bell (a conditioned stimulus) after repeated pairings of the bell with food (an unconditioned stimulus). Subsequent studies showed that a bell with a different tone than the original would also produce the salivation response through generalization. According to these theories, the strength of generalization is directly proportional to how similar a new stimulus is to the original conditioned stimulus (i.e., the *generalization gradient*); a new bell is more likely to elicit salivation to the extent that it is similar in tone to the original bell (Klein, 2019).

In studies of learning, the experimental situation is tightly controlled and the observed behavior is relatively unambiguous (e.g., milliliters of saliva produced, number of times a lever is pushed). But understanding the role of generalization in people's evaluations of other people is not so straightforward.

DOI: 10.4324/9781003045687-17

First, understanding evaluations of people is complicated by the fact that target's behavior and traits are inherently ambiguous and context-dependent (Uleman, 2005). Further, whereas a sound can be made more or less similar to another sound by manipulating frequency in hertz (i.e., pitch), judgments of the similarity between two people can be influenced by the perceiver's motivations, expectations, prejudices, personality, and situational constraints. And, of course, those of us studying person perceptions are continually attuned to the discrepancies between what people think, say, and do. Thus, documenting attitude generalizations in person perception has unique challenges in comparison to our behaviorist colleagues' observations of whether pigeons learn to peck disks of differing color wavelengths (Blough, 1967) or whether the fear that a rodent learns in a box with a square roof transfers to a box with a triangular roof (Huckleberry et al., 2016).

The goal of this chapter is to provide an overview of attitude generalization and transfer in person perception to better understand when, why, and how people use information about one person or group to judge related others, and how these processes can form and maintain group-based prejudices and stereotypes by "spreading" evaluative information—particularly negative information—across group members. The chapter will focus on what we have learned about attitude generalization and transfer, primarily in person perception, under the following general themes: lay beliefs about attitude generalization and transfer; evidence for attitude transfer on direct and indirect measures; associative and propositional explanations for attitude transfer effects; and similarity, categorization, and valence effects in attitude transfer. This chapter concludes with a discussion of related phenomena that may have implications for understanding attitude transfer and generalization, such as transference (Andersen et al., 1995), cognitive balance (Heider, 1958), spontaneous trait transfer (Skowronski et al., 1998), and stereotyping (Hamilton & Trolier, 1986).

Lay Beliefs about the Acceptability of Attitude Transfer

Over the years, my lab has explored people's lay beliefs about the acceptability of attitude transfer among people with various kinds of relationships.[2] In one study, we collected demographic information from U.S. American participants along with several individual difference measures. Participants also responded to the stem: "To what extent is it acceptable to use information about one person to form an impression of [X]", where X was 24 possible relationships, individually presented, including those of a coincidental nature (e.g., person standing at the same bus stop), those that include a choice ("their best friend") and those that differ in family closeness (e.g., "sibling" or "cousin"). Table 15.1 shows the average response, on a five-point scale ranging from 1 = Very Unacceptable to 5 = Very Acceptable.

There are a few observations about these lay beliefs that I would like to highlight, with a clear caveat that these are largely post-hoc explanations.

Table 15.1 Average perceived acceptability of attitude transfer by relationship type

Relationship Type	*Mean Transfer Acceptability (SD)*
Someone standing at the same bus stop	1.72 (0.86)
Someone of the same race	1.72 (0.89)
Someone who is the same gender	1.82 (0.91)
Someone who lives in the same large city	1.85 (0.90)
Someone shopping at the same grocery store	1.88 (0.90)
Someone in the same large lecture class	1.88 (0.88)
Their co-worker	1.91 (0.88)
Their employee	2.01 (0.90)
Their cousin	2.02 (0.90)
Someone in the same small lecture class	2.05 (0.94)
Someone who lives in the same neighborhood	2.07 (0.95)
Their employer	2.08 (0.96)
Someone who lives in the same small town	2.10 (0.96)
Someone shopping at the same clothing store	2.11 (0.97)
Someone on the same sports team	2.22 (1.00)
Someone who likes the same music	2.31 (1.06)
Someone who attends the same church	2.44 (1.05)
Their sibling	2.44 (1.06)
Their child	2.54 (1.12)
Someone who belongs to the same political party	2.60 (1.08)
Someone who belongs to the same sorority	2.63 (1.11)
Their parent	2.65 (1.12)
Their romantic partner	2.78 (1.14)
Their best friend	2.80 (1.15)

Participants (N = 1,848) responded to the stem "To what extent is it acceptable to use information about one person to form an impression of [X]" on a five-point scale ranging from 1 = Very Unacceptable to 5 = Very Acceptable.

Overall, people see it as relatively unacceptable to use information about one person to inform their opinion of another. This finding is consistent with the idea that most people find it unreasonable and immoral to use information about one person to judge another (Banaji & Bhaskar, 2000), and with the observation from my lab that people report their judgments about the acceptability of attitude transfer being more influenced by logic (M = 3.40) and fairness (M = 3.39) than by social norms or conventions (M = 2.58), $d_{diff} > 0.20$ (scale = 1 to 5).

In addition, people who agree that attitude transfer is acceptable for one relationship are likely to agree for other relationships. Treating the ratings as a scale, the reliability is $\alpha = .96$. The intraclass correlation (ICC) between scale items is .49 ($p < .0001$). That said, there is variability in the extent to which ratings correlate with one another; the range of correlations between any two of the 24 items ranges from $r = .25$ (best friends and people at the same bus stop) at the lowest to $r = .68$ (parent and child; $p < .0001$) at the highest.

Individual Differences in Lay Beliefs about the Acceptability of Attitude Transfer

In the study for which data are presented in Table 15.1, neither overall acceptability ratings nor discrimination between the relationships with the highest (best friends) and lowest (people standing at the same bus stop) ratings are related to participant age, gender, education, racial/ethnic group, political orientation, or religious identity.

Although demographics do not seem to moderate belief about the acceptability of attitude transfer, ideological and motivational orientations do. For example, there is a small, positive correlation (r ranging from .18 to .24) between political conservativism and the extent to which attitude transfer is perceived to be acceptable and the extent to which information about one group member is used to inform evaluations of another. Other factors that are associated with beliefs about the acceptability of attitude transfer in our unpublished data include personal attitude stability (the extent to which people believe their own attitudes are stable; Xu et al., 2020), personal need for structure (the desire to structure the world in simple, more manageable form; Thompson et al., 2001), right-wing authoritarianism (a set of attitudes including dogmatism, a preference for conformity, willingness to coercively enforce behavioral standards, punitiveness toward enemies, and strong concern with hierarchy; Altemeyer, 1988; Costello et al., 2021), and essentialist beliefs (the belief that group membership is immutable, described in more detail below; Haslam et al., 2006). We have focused largely on U.S. Americans in our work, though it is likely that beliefs about the acceptability of attitude transfer—like other naïve theories of impression formation (Shimizu et al., 2017) differs across cultures and subcultures.

Entitativity and Lay Beliefs about the Acceptability of Attitude Transfer

The pattern of acceptability ratings presented in Table 15.1 are consistent with theories about group *entitativity*—the extent to which a collection of individuals is perceived as being a coherent, unified entity (i.e., "groupiness"; Campbell, 1958). Judgments of entitativity reflect that some assemblages of people, like those standing at the same bus stop, are less likely to be seen as a single, meaningful unit than others, such as people playing together on the same sports team (Hamilton, 2022, this volume; Lickel et al., 2000). While judgments of group entitativity do have some overlap with judgments of similarity between group members (Dasgupta et al., 1999), they are distinct (Crump et al., 2010). I will return to this point later in the chapter. There are two dominant approaches to understanding entitativity: the "essence-based" and "agency-based" approaches (Brewer et al., 2004; see Agadullina & Lovakov, 2018, for a meta-analysis and review).

Essence-Based Entitativity

According to the essence-based approach to entitativity (Crump et al., 2010), groups are high in entitativity to the extent that they are perceived as homogenous, particularly in physical or mental/trait characteristics. Haslam et al. (2002) conception of essentialism is an example of the essence-based approach, focusing on the underlying, inherent nature of social categories. Consistent with this conceptualization, participants generally see it as more acceptable to use information about one person to judge their family members than non-family, and the distance among family members matters; for example, transfer between siblings is seen as more acceptable than transfer between cousins (Cohen's $d = 0.43$). This is consistent with the finding that mock jurors see criminal defendants as being more likely to be guilty to the extent that they are seen as similar—without providing a definition of similarity—to family members who have been convicted of a crime (Rerick et al., 2021).

In line with the essence-based approach to entitativity, we also measured endorsement of *psychological essentialism*—the belief that differences between (racial) groups are immutable and naturally occurring (Haslam et al., 2006). Participants completed an eight-item racial essentialism scale (Pauker, unpublished), which included items such as "knowing what race someone is tells you a lot about their abilities and traits" and "Race is determined by biological factors such as genes and hormones." There was a significant, positive correlation between essentialist beliefs and ratings of the acceptability of using information about one person to form an impression of another person of the same race ($r = .28$, $p < .0001$). A comparable correlation was observed between essentialism and the average of *all* acceptability ratings ($r = .25$, $p < .0001$), suggesting that essentialist views of groups in general—and racial groups in particular—are related to judgments about the acceptability of attitude transfer.

Agency-Based Entitativity

The agency-based approach to entitativity considers the group's heterogeneity, motivations, intentions, and level of interaction among group members, with the latter factor argued to be the most important (Brewer et al., 2004). Group size factors into these judgments. For example, transfer acceptability ratings are higher when considering people living in a small town compared to a big city ($d = 0.27$) and students in a small lecture class compared to a large one ($d = 0.19$). Narrower categories also seem to promote transfer; for example, people believe it is more acceptable to judge people interchangeably who are shopping in the same clothing store compared to the same grocery store ($d = 0.25$).

Entitativity and Attitude Transfer Effects

Crawford et al. (2002) more directly demonstrated that the transfer of traits from one individual group member to another is dependent on perceived

group entitativity. They presented participants with valenced information about one group member and then gave people the opportunity to evaluate a second group member. The group was manipulated between-subjects to be high in entitativity (i.e., made up of similar people with shared background, attitudes, and personalities) or low in entitativity (diverse people with different backgrounds, attitudes, and personalities). Participants treated the group members as interchangeable only when the group was perceived as highly entitative. Presumably, when a social group was thought to be large and diverse, participants recognized that one individual is not representative of all group members and resisted the transference of traits—and presumably the valenced evaluation implied by those traits—from one person to another person in the same group.

It is clear from this exploration of lay beliefs and acceptability judgments that there is variation in perceptions of how acceptable it is to use information about one person as the basis of evaluating another who is related in some way. However, a deliberate judgment that a particular association between two people is not a sufficient basis for judgments ignores that those two people *are* in fact associated, whether by group membership, identity, family relationship, shared interests, or proximity. Thus, even when attitude transfer is deemed unacceptable and deliberately resisted, we might expect generalization to occur anyway through processes that are spontaneous (i.e., without instruction or intention to make them; Uleman, 1987) or difficult to control (see Moors & De Houwer, 2006, for an overview of the features of automaticity).

Evidence for Attitude Transfer on Direct and Indirect Measures of Evaluation

Participants in my earliest studies of attitude transfer (Ratliff [Ranganath] & Nosek, 2008) were first exposed to an attitude formation paradigm in which they read behaviors performed by Reemolap, a member of the group *Laapians*, and Vabbenif, a member of the group *Niffians*; there is considerable evidence that people spontaneously form impressions of people based on their behavior (i.e, spontaneous trait inference; Uleman et al., 1996). One of these original group members performed predominantly positive behaviors and the other predominantly negative behaviors, manipulated between-subjects. Participants were then given minimal information about two new individuals, Bosaalap and Ibbonif, belonging to the same groups, for example:

> *Ibbonif is a sculptor and very much enjoys gardening, biking, and playing card games. Ibbonif is kind and thoughtful, but tends to be slightly greedy at times.*
>
> *Bosaalap is a painter and very much enjoys cooking, hiking, and listening to music. Bosaalap is warm and considerate, but tends to be slightly dishonest at times.*

Pretesting showed these descriptions to be evaluated as similarly valenced; the descriptions were also randomized across participants, ensuring that any differences in attitudes toward the new people could be a function only of attitudes formed toward the original people from the induction phase.

To evoke low entitativity, we described the groups as being large, diverse, and made up of many kinds of people who do many kinds of activities; it was plainly stated that Bossalaap and Reemolap, and Ibbonif and Vabbenif, had never met one another. At the same time, the group members in this study did share some physical features (see image below). Participants were then tasked with evaluating either Reemolap and Vabbenif (the original people) or Bosaalap and Ibbonif (the new people). Self-reported evaluations of the new people were not influenced by the behavioral descriptions of the original people; however, evaluations of the new people measured with the Implicit Association Test (IAT; Greenwald & Lai, 2020; Greenwald et al., 1998) were equal in strength and direction to evaluations of the original people, a phenomenon that, for now, we will refer to as *implicit attitude transfer*. This effect was replicated as part of the Reproducibility Project: Psychology (Open Science Collaboration, 2015). Together, we interpreted these findings as suggesting that, even in the absence of self-reported attitude transfer, we might expect generalization on indirect measures of attitudes that capture evaluations that are (relatively) more spontaneously generated or difficult to control.[3]

An Associative Learning Explanation for Attitude Transfer and Generalization Effects

There were two assumptions that guided our original interpretation of these findings (an interpretation that has shifted over time, as described below). The first assumption is that attitudes may be acquired through simple associative learning mechanisms (Cacioppo et al., 1992; Eagly & Chaiken, 1993; Olson & Fazio, 2002; Walther et al., 2005; Uleman et al., 2008). Associationism provides an explanation for the fact that much of our knowledge is more complex than a simple summary of direct experience. For example, you might speculate about someone's personality—even if you have not met them—based on what you know about someone else who belongs to the same social group.

Associative learning is a descriptive term referring to the type of learning that occurs anytime a relationship is detected between multiple concepts or events in an organism's environment (Shanks, 1995). The importance of associationism and associative learning in social psychology is evident in the very definition of an attitude as an "association between a concept and an evaluation—positive or negative, favorable or unfavorable, or desirable or undesirable" (Fazio, 1986, p. 214). This assumption led to the hypothesis that, even if someone consciously rejects the idea of evaluating the new colleague based on evaluations of someone else from the same group, the association between them exists, and may lead to generalization anyway. This hypothesis is consistent with the Associative-Propositional Evaluation

(APE) Model (Gawronski, 2022, this volume; Gawronski & Bodenhausen, 2008, 2011) whereby an associative learning mechanism leads to the automatic formation of mental associations between stimuli that are linked in some way. The APE model also posits a propositional mechanism whereby deliberate reasoning about the truth value of the linked stimuli is possible.

The second assumption guiding our original interpretation of the Ratliff [Ranganath] and Nosek (2008) attitude transfer findings is related to the first—that propositional knowledge is better assessed through direct measures and associations through indirect measures. *Direct measures* of attitudes involve asking participants to self-report their evaluations. *Indirect measures*, on the other hand, either do not alert participants to what is being measured, or reduce participants' deliberative control over their responses, even if they are aware of what is being measured (De Houwer, 2006).

Ratliff [Ranganath] and Nosek's (2008) findings—and subsequent findings replicating the attitude transfer effect (Chen & Ratliff, 2015; Hawkins & Ratliff, 2015; Ratliff & Nosek, 2011)—are consistent with a dual-process argument as follows: An association[4] between the original and new group members forms automatically based on their relationship (e.g., shared group membership, physical resemblance, temporal or spatial proximity of learning about them). This association occurs through generalization and/or second-order conditioning: *generalization* is where a response conditioned to one stimulus generalizes to other stimuli that are similar (see Till & Priluck, 2000), and *second-order conditioning*, sometimes called a *spreading attitude effect*, where an association between Stimulus A (Reemolap) and Stimulus B (negative behavior) and an association between Stimulus A (Reemolap) and Stimulus C (Bosaalap) leads to an association between Stimulus B (negative behavior) and Stimulus C (Bosaalap), despite them never being presented together (Walther, 2002). In either case, the new group members automatically[5] take on the evaluation of the original group members. Thus, one assumption about associations is the formation of attitudes toward novel group members via their relationship with known group members—if a known group member (Reemolap) is viewed negatively, that negativity will transfer to a novel group member (Bosaalap).

We then assume that people can (and usually will; see Table 15.1) apply a rule that it is unfair or illogical to evaluate one member of a large, diverse group based on another person's actions, and so self-reported attitudes reflect differentiation between the new and old group members; that is, we deliberately reject the negative evaluation of group member B. A second assumption about associations now comes into play—that lingering negativity toward group member B can be assessed with the IAT, because that measure is particularly suited to assess association between concepts (e.g., Reemolap, Vabbenif) and evaluative attributes (e.g., good, bad).

To recap the argument, self-reported evaluations of the new people were not influenced by the behavioral descriptions of the original people; however, evaluations of the new people measured with the IAT were equal in strength

and direction to evaluations of the original people. We made two assumptions about associations to explain these findings—that the positivity or negativity of original group members transferred to new group members via their association, and that the association between the new group member and positivity or negativity could be assessed by the IAT. The original data and subsequent studies are consistent with these assumptions, as described below.

Ratliff [Ranganath] and Nosek (2008) found that, after a delay of several days, self-reported evaluations showed evidence of generalization too, suggesting that as resources to prevent generalization (e.g., clear knowledge of who did what) decline, simple associations may play a bigger role in evaluations of new group members. Further support for this interpretation came from subsequent studies showing that there was a stronger correlation between the IAT and self-reported evaluations of the new people at Time 2 than existed immediately after learning about the group members (Ratliff & Storbeck, unpublished data). These findings are consistent with findings that cognitive resources are important for social inference; for example, Kubota et al. (2014) showed that people were more likely to make dispositional judgments about people's behavior when experiencing physiological stress that disrupted executive functioning.

Consistent with other demonstrations of difficulty preventing formation of conditioned attitudes on indirect measures (Gawronski et al., 2014), manipulations to induce intentional control over attitude transfer on the IAT failed to do so. In five studies, Hawkins and Ratliff (2015) demonstrated that IAT scores assessing attitudes toward new consumer products (Studies 1 and 2; see also Ratliff et al., 2012) or new group members (Studies 3–5) showed evidence of attitude transfer even when participants were explicitly instructed to be fair and to avoid generalization. An accountability manipulation telling participants that they would be required to explain their decision at the end of the study (e.g., Thompson et al., 1994) did not prevent implicit attitude transfer, nor did priming egalitarian goals using an objectivity writing task (Moskowitz & Li, 2011), regardless of whether the objectivity manipulation was presented before the attitude induction, after the attitude induction but before introduction to new group members, or after the attitude induction and introduction to new group members but before completing dependent measures. Two additional studies manipulated the relationship between the original and new people to disrupt implicit attitude transfer; however, there were still transfer effects on the IAT when the original and new people were described as enemies (Study 7) or as strangers at a bus stop (Study 8).

Internal meta-analysis confirmed that, in all eight studies, implicit and explicit evaluations formed toward novel consumer products and social group members. Although these attitudes influenced subsequent evaluations of new consumer products, the transfer of evaluations to new people was largely avoided in those studies with social groups. Implicit evaluations, on the other hand, transferred readily between consumer products and social group members, even in those conditions where deliberate control processes should have been engaged.[6]

Challenges to an Associative Learning Explanation for Attitude Transfer and Generalization Effects

In recent years, both assumptions used to explain basic attitude transfer effects on direct and indirect measures—that attitude generalization can be explained through "simple" associative learning mechanisms and that the IAT measures associations between evaluations and attitude objects—have been convincingly challenged (Bading et al., 2020; Corneille et al., 2019; De Houwer, 2014; De Houwer 2019; Högden & Unkelbach, 2020; Van Dessel et al., 2020). Challenges to the assumption that attitude formation and generalization can be explained primarily through associative learning stem largely from studies of *evaluative conditioning*, which can be functionally defined as a change in liking of a neutral stimulus (conditioned stimulus; CS) because of its pairing with a valenced stimulus (unconditioned stimulus; UCS; Gast et al., 2012).

In the attitude transfer paradigm described previously (Chen & Ratliff, 2015; Hawkins & Ratliff, 2015; Ratliff [Ranganath] & Nosek, 2008; 2011), EC-like[7] effects are observed in evaluations of the original people—the initial targets of the attitude formation paradigm used to induce attitudes toward the original group members before the introduction of new people from the same group. Both implicit and explicit attitudes are consistent with the valence of the information presented in the induction; that is, attitudes toward the group described as performing predominantly positive behaviors are more positive than those toward the group described as performing predominantly negative behaviors; these effects are generally very large ($d > 1.0$).

It is generally assumed that conditioning effects are due to the formation of associations between the CS and UCS (De Houwer, 2018). The associative learning hypothesis has strong face validity (Corneille & Stahl, 2019) and empirical support. For example, Hu, Gawronski, & Balas (2017) simultaneously presented participants with pharmaceutical products and negative health conditions, varying whether the product was described as causing or preventing the health condition. Although self-reported evaluations of the pharmaceutical product reflected the relation between the condition, responses on an indirect measure reflected the co-occurrence of the product with the health condition regardless of the relationship (see Moran & Bar-Anan, 2013, for similar findings). More recently, and consistent with the finding from Hawkins and Ratliff (2015), Gawronski and Brannon (2021) found that instructions to counteract stimulus co-occurrence effects were ineffective in preventing evaluative conditioning. Together, these results support the idea that "mere" associations between stimuli could lead to attitude transfer between stimuli that are related in some way.

On the other hand, behavioral and physiological evidence suggests that evaluative conditioning is mediated by propositional representations that specify knowledge about how stimuli are related to one another (relational knowledge; De Houwer, 2018, 2019; Mitchell et al., 2009)—for example, A *predicts* B, A *causes* B, A *goes with* B, A *prevents* B, A *is like* B, A *is opposite* B,

etc. Whether (and how) the valence of one stimulus influences evaluations of another stimulus may depend on this propositional knowledge about how the stimuli are related to one another.

For example, Zanon et al. (2014; Experiment 1) presented participants with one set of non-words paired with positive stimuli and another set paired with negative stimuli, and then told them that the meaning of the non-words was the opposite that of the stimuli with which it was paired. For associative learning, this relational information should not matter—an IAT assessing evaluations of the novel word should reflect the valence of the paired word; however, participants showed an IAT effect demonstrating more positivity toward the non-words paired with negative over those paired with positive, thereby reflecting the instructed pairing (i.e., a proposition) rather than the observed pairing (i.e., an association). Along these lines, Kurdi et al. (2020) found that implicit evaluations of novel moral agents reflected the valence of the outcome their actions produced but also inferences about the actor's mental state; for example, IAT scores reflected more negativity toward someone who *intentionally* caused another person to fall off of a bridge compared to someone who caused the same outcome without meaning to do so. Finally, mere instructions about a learning procedure leads to IAT scores that are at least as large as experiencing the actual learning procedure (e.g., Kurdi & Banaji, 2017; Smith et al., 2019; Van Dessel et al., 2018). Taken together, these and other studies provide strong evidence that implicit measures do not only reflect simple associations (e.g., object + good), but can incorporate additional logical information.

A Propositional Account of Attitude Transfer Effects

It does not require extensive mental gymnastics to reinterpret attitude transfer effects in accordance with prominent propositional models like the Integrated Propositional Model (IPM; De Houwer, 2014, 2018). According to the IPM, and as described previously, a proposition is a mental representation of information about how stimuli are related (e.g., A *predicts* B, A *causes* B, A *goes with* B, A *prevents* B, A *is like* B). Propositions have inherent truth value (i.e, the potential to be true or untrue) which can be accepted or rejected in any given situation through inferential reasoning. For example, propositions about the relationship between group members in the Ratliff [Ranganath] and Nosek (2008) attitude transfer paradigm might be Bosaalap *goes with* Reemolap, or Bosaalap *is like* Reemolap. That proposition would be accepted or rejected based on what one believes to be true about the acceptability of using information about Reemolap to evaluate Bosaalap. For example, the proposition that Reemolap *goes with* or *is like* Bosaalap may be accepted if Reemolap and Bosaalap belong to a group that is more entitative compared to a less entitative group (Hawkins & Ratliff, 2015), an outgroup to the perceiver compared to an ingroup (Chen & Ratliff, 2015), or a group that is joined voluntarily (e.g., choosing to join the military) compared to one that is not (e.g., drafted into the

military; Vitiello & Ratliff, unpublished data). This inferential reasoning about the propositions is then reflected in self-reported evaluations of the new person where it is either accepted (i.e., if Reemolap is like Bosaalap) or rejected (i.e., Reemolap is not like Bosaalap).

So far, there is no divergence in what the IPM and dual-process or associative models, including the APE model (Gawronski & Bodenhausen, 2006, 20011), would predict. But on its surface the IPM would predict sensitivity to deliberate reasoning about group members to be reflected on both self-report *and* the IAT. So why does the IAT reflect guilt-by-association even when inferential reasoning about propositions would say otherwise? De Houwer (2014) describes several ways in which the IPM can account for effects that might appear associative, for example, discrepancies between behavior (in this case, IAT performance) and people's self-reported propositions (in this case, consciously rejecting the idea that information about one group member should be used to evaluate another group member).

The IPM assumes that propositions are activated automatically from memory, which allows propositional models to mimic the behavior of an associative-based network. Thus, it is possible to obtain *partial* automatic retrieval of the proposition; that is, retrieval of propositions form memory may be incomplete. In the case of attitude transfer, someone may have formed and memorized the proposition that *Reemolap is* not *like Bosaalap*, which is what they self-report, but the "not" is dropped during automatic retrieval and what is retrieved from memory during performance of the IAT is instead that *Reemolap is like Bosaalap*. De Houwer (2014) also proposes the possibility that automatically retrieving an older or rejected proposition from memory can lead to effects that run counter to currently held propositions. For example, the general proposition that *things that look similar behave similarly*, or that *things in proximity to one another are similar* may be rejected in favor of a more egalitarian proposition, such as *things that look similar sometimes behave differently*. So, although *Reemolap is like Bosaalap* is an older, rejected proposition, it may be retrieved from memory during IAT performance nonetheless. It could also be that the *things that look similar behave similarly* proposition is more well-rehearsed and, validated more frequently, or is a more general rule than the less practiced, exception-to-the-rule proposition that *things that look similar sometimes behave differently*; these may be other ways in a which one proposition may be feebler—and therefore less likely to be automatically activated—than another.

The IPM has been criticized for conflating the concepts of *consciousness*, *intentionality*, and *effort* within the umbrella of propositional reasoning (Uleman, 2009) and for being too flexible to be falsifiable (Kurdi & Dunham, 2020), and others have argued that the IPM can better explain conditioning though it does offer an alternative explanation for attitude transfer effects that generates novel hypotheses about when we would expect to see attitude transfer effects. Importantly, however, although this model presents a challenge to our original assumptions about the associative nature of attitude

transfer effects, and the ability of the IAT to assess newly formed associations between evaluations and group members, it does not undermine the existence of attitude transfer as an effect that can help us to understand the formation and maintenance of group-based attitudes and stereotypes.

Similarity, Categorization, and Valence Effects in Attitude Transfer

Similarity between Stimuli Impacts Generalization

The role of similarity in generalization is a topic big enough to carry its own chapter, so here I will focus only on select findings that are particularly relevant for impression formation. Fazio et al. (2004) created BeanFest, a computer game where the goal is to accumulate points by making correct decisions about which beans to approach and which to avoid. Approaching a positive bean increases points and approaching a negative bean decreases points. During the game, the beans are viewed on a 10 × 10 matrix where the x-dimension is the shape of the bean (ranging from perfectly circular to oblong) and the y-dimension is the number of speckles the bean has (ranging from 1 to 10). Thirty-six beans, in six regions of the board, were presented to the participants; each selected bean produced a positive or negative outcome after it was presented. Generalization was measured by having participants evaluate 64 novel beans that varied in similarity to the known beans (via Euclidean distance from the novel bean to the nearest bean in the matrix).

The results clearly demonstrated a generalization effect: The more that novel beans visually resembled known beans, the more they were evaluated similarly to the known beans (Fazio et al., 2004). This finding is consistent with early work on generalization gradients in learning, and also with the *shared features principle*, which refers to the idea that when stimuli share one feature, people often assume they share others as well (Hughes et al., 2020). By this logic, if two individuals share some features (e.g., physical resemblance, group membership), people will assume they share others. In this case, if behavior-valence is the only other information one has about one group member, it makes sense that it would be the feature shared between individuals. Attention to shared features is also important for generalization; generalization from Stimulus A to Stimulus B was stronger when the two stimuli were similar on dimension to which participants were instructed to pay attention compared to when the two stimuli were similar on a less salient dimension (Spruyt, et al., 2014). Alves et al. (2020) also found differentiation effects, where generalization occurs more strongly between stimuli that are *distinctly* related to one another. Based on this, we might expect stronger attitude transfer among group members who are members of only one group compared to those with multiple group memberships.

There are also open questions about the role of *actual* versus *perceived* similarity in attitude transfer/generalization. Verosky and Todorov (2010)

demonstrated that attitude generalization increased as a function of objective facial similarity (i.e., novel faces morphed with known faces at 20% or 35%; see also Kraus & Chen, 2010, and Günaydin et al., 2012). This linear effect is consistent with classic learning theory's emphasis on generalization gradients (Klein, 2019). On the other hand, Gawronski and Quinn (2013) showed that—on self-report and an evaluative priming task (Fazio et al., 1995)—the valence of a known individual equally generalized to a novel individual whose face was morphed at 50% or 100% resemblance to the original face. There is also evidence that, for White perceivers, IAT performance indicates that evaluations of a novel group member reflect greater attitude transfer from an original group member when the group members are Black relative to White (Ratliff & Nosek, 2011). This may be due to an outgroup homogeneity effect where the Black group members *seem* to visually resemble one another though they actually do not. It is also possible that, consistent with entitativity effects described previously, groups with members that are perceived as being more visually similar to one another are also seen as more entitative, coherent, and unified (i.e., "group-y"), or as sharing an essence.

There are still other examples where no similarity at all seems to be required for attitude transfer to occur. For example, Hebl and Mannix (2003) found that a male individual seated next to an overweight woman was denigrated in a hiring context substantially more than if he was seated next to an average-sized woman. This effect held even when it was made clear that there was no relationship between the applicant and the overweight woman, suggesting that, at least in some cases, proximity is all that is required to observe attitude transfer effects (see also Hawkins & Ratliff, 2015; Pryor et al., 2012).

One possibility is that generalization and higher-order conditioning are separable processes by which attitudes transfer from one group member to another. A generalization explanation for attitude transfer effects would predict that more objective resemblance between new and old group members should lead to stronger transfer effects. In higher-order conditioning a stimulus only has to be associated *in some way* with another stimulus to take on its valence, for example, by sharing category membership, whether or not it is similar to that original (Walther, 2002). An interesting avenue for future exploration would be to compare the extent of attitude transfer based on different types of group member similarity (e.g., visual resemblance vs. shared beliefs vs. biological relationships). Such tests would be interesting for better understanding transfer effects but could also shed light on the role of associations versus propositions in attitude transfer.

Stimulus Valence

Negative information is generally more influential than positive information in evaluation (Cacioppo et al., 1997). For example, little unfavorable information is needed to confirm a negative stereotype about a group, but quite a bit of favorable information is needed to form a positive stereotype or to

disconfirm a negative stereotype (Rothbart & Park, 1986). Negativity also seems to be more "contagious" than positivity (Rozin & Royzman, 2001; Boydstun et al., 2019). For example, most people refuse to wear a sweater believe to be worn by Adolf Hitler, even if Mother Theresa also wore it. Early evidence that negative evaluations also *generalize* more readily than positive evaluations came from follow-up studies using Fazio and colleagues' BeanFest paradigm. Overall, generalization is moderated by information extremity and valence (Shook et al., 2007). Similarity to beans that produced a more extreme point gain or loss mattered more than similarity to those with a more tempered outcome, but this was less true for negative beans; that is, it takes less similarity to a negative bean to be judged negatively than it takes similarity to a positive bean to produce an equally strong positive evaluative outcome. Ratliff and Nosek (2011, Study 1) observed similar effects using the novel groups attitude transfer paradigm (Ratliff [Ranganath] & Nosek, 2008). Compared to positive information, one group member's negative behavior had a stronger influence on evaluations of a novel group member; this effect was observed on both self-report measures (a small effect) and the IAT (large effect size).

Although there is a general tendency toward negativity bias in attitude generalization, there is also variability (Fazio et al., 2015). Negative attitudes generalize more strongly for some people, positive for some people, and others show no asymmetry in valence weighting. Pietri et al. (2013) demonstrated that this *weighting bias* in generalization is related to behavioral manifestations of rejection sensitivity (level of concern about and perceived likelihood of interpersonal rejection; Downey & Feldman, 1996), threat assessment (judgments of the likelihood that an ambiguous situation will become negative or threatening; Riskind et al., 2000), and risk tolerance (a preference for high-risk/high-reward over low-risk/low-reward options; Wallach et al., 1962). Correlations between these measures and valence weighing are not high (ranging from $r = .22$ to $r = .38$), but it is noteworthy that attitude generalization in a novel computer game would relate to performance-based measures at all.

Stereotyping and Other Related Phenomenon

A variety of other phenomenon may be related to attitude transfer but are just outside the scope of this chapter. Anderson and colleagues' (Andersen et al., 1995) work builds on the Freudian idea of *transference* where a new person activates a representation of a significant other based on one or more common features, and then the new person is assumed to share traits with the significant other. For example, participants evaluate a target person who is physically similar to their romantic partner as having the same personality traits as the romantic partner (Glassman & Andersen, 1999). Work on cognitive balance (Heider, 1958) may also be relevant in person perception. For example, the motivation to maintain coherence among attitudes would

predict that people will like others who are liked by those they feel positively about or who are disliked by those they feel negatively about, and that people will dislike those who are disliked by people they feel positively about or who are liked by those they feel negatively about; Gawronski et al. (2005) found just this. Another related phenomenon is spontaneous trait transfer (STT; Skowronski et al., 1998). In STT, which is frequently described as an associative process (Uleman et al., 2008), the valence of the information that Person A uses to describe Person B transfers to Person A, and thus Person C would evaluate Persons A and B similarly.

And of course, attitude transfer is highly related to stereotyping. Many explanations of group-based stereotyping explain stereotyping as a hierarchical generalization from beliefs about the group to beliefs about the individuals within that group. For example, Secord (1959) defines stereotyping as "a categorical response, i.e., membership is sufficient to evoke the judgment that the stimulus person possesses all the attributes belonging to that category" (p. 309). Others define stereotypes as characteristics that are associated with either a group (Katz & Braly, 1933) or as characteristics that are associated with an individual due to her or his group membership (Hamilton & Trolier, 1986). Stereotyping, like other forms of generalization, can be either accurate or inaccurate. Part of the purpose of a superordinate group is to described features that are shared among members of that group (Medin et al., 1993; Tversky, 1977); however, in the case of stereotyping, those features may be distorted or erroneously applied to group members who do not share them.

The studies I have described here focus primarily on a person-to-person attitude transfer; however, we might also consider the possibility of person-to-group transfer and how such transfer processes might contribute to the formation and maintenance of stereotypes. Hamilton et al. (2015) demonstrated that perceivers draw spontaneous trait inferences about groups just as they do about individuals. Further, evaluations of a single group member influence evaluations of the group itself. Henderson-King and Nisbett (1996) demonstrated that participants in a study with an unkind Black confederate were subsequently less likely to sit near a different Black confederate. Olson and Fazio (2006) used an evaluative conditioning paradigm in which they presented participants with faces of Black or White individuals paired with positive or negative pictures; participants who saw Black faces paired with positive words, and White faces paired with negative words, were subsequently more positive toward different Black people; that is, evaluations of the individual exemplars generalized to the category. Similar exemplar-to-category generalization effects have been demonstrated in adults with fictitious aliens and employees at a company (Glaser & Kuchenbrandt, 2017). Stark et al. (2013) showed that negative attitudes toward an individual outgroup member contributed to students' attitudes toward that person's ethnic group, and Skinner et al. (2020) found that adults' biases in favor of or against one individual can influence children's evaluations of the groups to

which those adults belong. These kinds of person-to-group generalization processes may also contribute to the success (or failure) of intergroup contact, a strategy for prejudice reduction that involved bringing together people from different groups to increase interaction (Allport et al., 1954; see Paluck et al., 2021 for a review).

Concluding Remarks

In *Don Quixote*, Miguel de Cervantes wrote: Tell me what company you keep, and I will tell you what you are (translated from Spanish). You probably noticed that I used a lot of related terms throughout the chapter: attitude transfer, attitude generalization, spreading attitude effect, higher-order conditioning, transference, stigma-by-association, etc. Although each of these has slightly different meanings, and is the preferred nomenclature in different research traditions, they all share the same fundamental idea—that evaluations of one individual may transfer to another who is related in some way. Sometimes this is intentional and people believe it is acceptable to judge people based on the actions of another. Many would agree, say, that choosing a known ax murderer as a best friend says something about you as a person. We would also probably agree that having one time stood next to an axe murderer at a public bus stop probably does not say anything about who you are as a person (other than that you are lucky to have survived the encounter!). But there is plenty of evidence that people's *actual* evaluations and behaviors are less concerned with deliberate judgments of fairness and may be influenced by relationships between group members despite intentions to the contrary.

Attitude transfer can have substantive real-world consequences. Imagine a teacher who judges a student's academic performance based on the grades of their friends, or a jury that judges a criminal based on whether he has family members who have been convicted of a crime (as in Rerick et al., 2021). Prior collaborators of scientists who are guilty of scientific fraud—who themselves have no connection at all to misconduct—face a citation penalty of ~9% in the aftermath (Hussinger & Pellens, 2017). In very tangible ways, attitude transfer matters. Further, transfer of attitudes from one person to another, or from one person to a whole group, could be a key mechanism in the formation of intergroup attitudes and can help to explain how stereotypes and prejudices are maintained. Further, on the positive side, the observation that attitudes toward groups and group members can be changed through evaluations of single group members could be a promising strategy for prejudice and stereotype-reduction interventions.

Acknowledgement

I would like to thank Colin Tucker Smith and Sarah Olshan for their helpful comments on this chapter.

Notes

1 Here I use the term *associated* in a general, lay sense to mean connected, linked, or related in some way. The definition contains no assumption about the presence or absence of deliberate, propositional reasoning about the relationship between the stimuli, a point I will return to in greater detail later.
2 The Attitudes and Social Cognition Lab at the University of Florida has accumulated a considerable amount of data on attitude transfer that never made it into manuscripts; we are using this chapter as an opportunity to present some of those findings that we think might be interesting or useful to others. We set up a project page on the Open Science Framework to share our unpublished study materials and data—https://osf.io/xas3w/. You can find more details about the studies there.
3 Direct measures of judgment generally rely on self-report, while indirect measures, such as the IAT, infer evaluation from participant performance on a task. Further, I am sensitive to calls to for scientists to clarify precisely what they mean when using the term "implicit" or even to abandon the term entirely (Corneille & Hütter, 2020); however, for ease of presentation and continuity, I will retain the original language—implicit attitude transfer—throughout this chapter to indicate transfer that was assessed using indirect measures of evaluations.
4 This is where the term *association* gets confusing. The original and new group members are inherently associated because we have said they are members of the same group (i.e., they are associates). But here I refer to *mental* associations, links between group members, represented in the mind (De Houwer, 2009), whether or not an individual intends or wants for them to form.
5 Recent arguments point to the importance of specifying what constitutes *automaticity* in a given context. Candidate characteristics of automaticity include responses that occur in ways that are uncontrolled, unaware, efficient, or fast (Moors & De Houwer, 2006); in this case, intentionality is the most relevant factor. Intentionality refers to the extent to which performance on a measure can be controlled when motivated to do so. In early attitude transfer studies we assumed a failure of intentionality due to the dissociation between what people self-reported (i.e., no attitude transfer) and what their performance on the IAT (i.e., attitude transfer); in later studies we directly manipulated intentionality (e.g., Hawkins & Ratliff, 2015), a point I return to momentarily, though further work is certainly needed.
6 In these studies, we inferred from the pattern of self-reported evaluations that people intended to avoid attitude transfer; however, it is possible that these online studies with volunteer participants simply did not engage the kind of strong commitment to egalitarian goals that would be necessary to control more spontaneous attitude transfer. Thus, the conclusion from these studies is better summarized as "attitude transfer is difficult to control" than as "attitude transfer cannot be controlled."
7 In a typical evaluative conditioning paradigm, neutral stimuli (UCS) are paired with valenced stimuli (CS) though typically in an incidental way (e.g., they appear on the screen at the same time). In our attitude induction paradigm (adapted from Gregg et al., 2006), participants are given descriptive behavioral information about the group members that directly implies trait characteristics (e.g., Reemolap helped an elderly person cross the street). So the group member is indeed "paired" with the behavioral information, though in a way that is not common for an EC study.

References

Agadullina, E. R., & Lovakov, A. V. (2018). Are people more prejudiced towards groups that are perceived as coherent? A meta-analysis of the relationship between out-group entitativity and prejudice. *British Journal of Social Psychology*, *57*(4), 703–731.

Allport, G. W., Clark, K., & Pettigrew, T. (1954). *The nature of prejudice*. Boston, MA: Addison-Wesley.

Altemeyer, B. (1988). *Enemies of freedom: Understanding right-wing authoritarianism*. Jossey Bass.

Alves, H., Högden, F., Gast, A., Aust, F., & Unkelbach, C. (2020). Attitudes from mere co-occurrences are guided by differentiation. *Journal of Personality and Social Psychology*, *119*(3), 560–581.

Andersen, S. M., Glassman, N. S., Chen, S., & Cole, S. W. (1995). Transference in social perception: The role of chronic accessibility in significant-other representations. *Journal of Personality and Social Psychology*, *69*(1), 41–57.

Bading, K., Stahl, C., & Rothermund, K. (2020). Why a standard IAT effect cannot provide evidence for association formation: The role of similarity construction. *Cognition and Emotion*, *34*(1), 128–143.

Banaji, M. R., & Bhaskar, R. (2000). Implicit stereotypes and memory: The bounded rationality of social beliefs. In D. L. Schacter, & E. Scarry (Eds.), *Memory, brain, and belief* (Vol. 2). Harvard University Press.

Blough, D. S. (1967). Stimulus generalization as signal detection in pigeons. *Science*, *158*(3803), 940–941.

Boydstun, A. E., Ledgerwood, A., & Sparks, J. (2019). A negativity bias in reframing shapes political preferences even in partisan contexts. *Social Psychological and Personality Science*, *10*, 53–61. 10.1177/1948550617733520

Brewer, M. B., Hong, Y., & Li, Q. (2004). Dynamic entitivity: Perceiving groups as actors. In V. Y. Yzerbyt, C. Judd & O. Corneille (Eds.), *The psychology of group perception: Perceived variability, entitativity, and essentialism* (pp. 25–38). New York: Psychology Press.

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1997). Beyond bipolar conceptualizations and measures: The case of attitudes and evaluative space. *Personality and Social Psychology Review*, *1*, 3–25.

Cacioppo, J. T., Marshall-Goodell, B. S., Tassinary, M. L., & Petty, R. E. (1992). Rudimentary determinants of attitudes: Classical conditioning is more effective when prior knowledge about the attitude stimulus is low than high. *Journal of Experimental Social Psychology*, *208*, 207–233.

Campbell, D. T. (1958). Common fate, similarity, and other indices of the status of aggregates of persons as social entities. *Behavioral Science*, *3*, 14–25.

Chen, J. M., & Ratliff, K. A. (2015). Implicit attitude generalization from Black to Black–White biracial group members. *Social Psychological and Personality Science*, *6*(5), 544–550.

Corneille, O., & Hütter, M. (2020). Implicit? What do you mean? A comprehensive review of the delusive implicitness construct in attitude research. *Personality and Social Psychology Review*, *24*(3), 212–232.

Corneille, O., Mierop, A., Stahl, C., & Hütter, M. (2019). Evidence suggestive of uncontrollable attitude acquisition replicates in an instructions-based evaluative conditioning paradigm: Implications for associative attitude acquisition. *Journal of Experimental Social Psychology*, *85*, 103841.

Corneille, O., & Stahl, C. (2019). Associative attitude learning: A closer look at evidence and how it relates to attitude models. *Personality and Social Psychology Review*, *23*(2), 161 189.

Costello, T. H., Bowes, S., Stevens, S. T., Waldman, I., & Lilienfeld, S. O. (2021). Clarifying the structure and nature of left-wing authoritarianism. In press at *Journal of Personality and Social Psychology*.

Crawford, M. T., Sherman, S. J., & Hamilton, D. L. (2002). Perceived entitativity, stereotype formation, and the interchangeability of group members. *Journal of Personality and Social Psychology*, *83*, 1076–1094.

Crump, S. A., Hamilton, D. L., Sherman, S. J., Lickel, B., & Thakkar, V. (2010). Group entitativity and similarity: Their differing patterns in perceptions of groups. *European Journal of Social Psychology*, *40*, 1212–1230. 10.1002/ejsp.716

Dasgupta, N., Banaji, M. R., & Abelson, R. P. (1999). Group entitativity and group perception: Associations between physical features and psychological judgment. *Journal of Personality and Social Psychology*, *77*(5), 991.

De Houwer, J. (2006). What are implicit measures and why are we using them. In R. W. Wiers , & A. W. Stacy (Eds.), *Handbook of implicit cognition and addiction*. Sage.

De Houwer, J. (2014). Why a propositional single-process model of associative learning deserves to be defended. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual processes in social psychology*. (pp. 530–541). New York: Guilford.

De Houwer, J. (2009). The propositional approach to associative learning as an alternative for association formation models. *Learning & Behavior*, 37, 1–20.

De Houwer, J. (2018). Propositional models of evaluative conditioning. *Social Psychological Bulletin*, *13*(3), 1–21.

De Houwer, J. (2019). Moving beyond System 1 and System 2: Conditioning, implicit evaluation, and habitual responding might be mediated by relational knowledge. *Experimental Psychology*, 66(4), 257–265.

Downey, G., & Feldman, S. I. (1996). Implications of rejection sensitivity for intimate relationships. *Journal of Personality and Social psychology*, *70*(6), 1327–1343.

Eagly, A., & Chaiken, S. (1993). *The psychology of attitudes*. Fort Worth, TX: Harcourt Brace Jovanovich.

Fazio, R. H. (1986). How do attitude guide behavior? In R. M. Sorrentino & Higgins, T. (Eds.) *Handbook of motivation and cognition* (pp. 204–243). New York: Guilford Press.

Fazio, R. H., Eiser, J. R., & Shook, N. J. (2004). Attitude formation through exploration: Valence asymmetries. *Journal of Personality and Social Psychology*, *87*, 293–311.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline?. *Journal of Personality and Social Psychology*, 69(6), 1013–1027.

Fazio, R. H., Pietri, E. S., Rocklage, M. D., & Shook, N. J. (2015). Positive versus negative valence: Asymmetries in attitude formation and generalization as fundamental individual differences. In *Advances in experimental social psychology* (Vol. 51, pp. 97–146). Academic Press.

Gast, A., Gawronski, B., & De Houwer, J. (2012). Evaluative conditioning: Recent developments and future directions. *Learning and Motivation*, *43*(3), 79–88.

Gawronski, B., Balas, R., & Creighton, L. A. (2014). Can the formation of conditioned attitudes be intentionally controlled?. *Personality and Social Psychology Bulletin*, *40*(4), 419–432.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*, 692–731.

Gawronski, B., & Bodenhausen, G. V. (2011). The associative–propositional evaluation model: Theory, evidence, and open questions. In *Advances in experimental social psychology* (Vol. 44, pp. 59–127). Academic Press.

Gawronski, B., & Quinn, K. A. (2013). Guilty by mere similarity: Assimilative effects of facial resemblance on automatic evaluation. *Journal of Experimental Social Psychology*, *49*(1), 120–125.

Gawronski, B., Walther, E., & Blank, H. (2005). Cognitive consistency and the formation of interpersonal attitudes: Cognitive balance affects the encoding of social information. *Journal of Experimental Social Psychology*, *41*, 618–626.

Gawronski, B., & Brannon, S. M. (2021). Attitudinal effects of stimulus co-occurrence and stimulus relations: Range and limits of intentional control. *Personality and Social Psychology Bulletin*, *47*, 1654–1667.

Glaser, T., & Kuchenbrandt, D. (2017). Generalization effects in evaluative conditioning: Evidence for attitude transfer effects from single exemplars to social categories. *Frontiers in Psychology*, *8*, 103.

Glassman, N. S., & Andersen, S. M. (1999). Transference in social perception: The role of chronic accessibility in significant-other representations. *Cognitive Therapy and Research*, *23*, 75–91.

Greenwald, A. G., & Lai, C. K. (2020). Implicit social cognition. *Annual Review of Psychology*, *71*, 419–445.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, *74*, 1464–1480.

Gregg, A. P., Seibt, B., & Banaji, M. R. (2006). Easier done than undone: asymmetry in the malleability of implicit preferences. *Journal of Personality and Social Psychology*, *90*(1), 1–20.

Günaydin, G., Zayas, V., Selcuk, E., & Hazan, C. (2012). I like you but I don't know why: Objective facial resemblance to significant others influences snap judgments. *Journal of Experimental Social Psychology*, *48*(1), 350–353.

Hamilton, D. L., Chen, J. M., Ko, D. M., Winczewski, L., Banerji, I., & Thurston, J. A. (2015). Sowing the seeds of stereotypes: Spontaneous inferences about groups. *Journal of Personality and Social Psychology*, *109*(4), 569–588.

Hamilton, D. L., & Trolier, T. K. (1986). Stereotypes and stereotyping: An overview of the cognitive approach. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 127–163). San Diego, CA: Academic Press.

Haslam, N., Bastian, B., Bain, P., & Kashima, Y. (2006). Psychological essentialism, implicit theories, and intergroup relations. *Group Processes & Intergroup Relations*, 9(1), 63–76.

Haslam, N., Rothschild, L., & Ernst, D. (2002). Are essentialist beliefs associated with prejudice?. *British Journal of Social Psychology*, *41*(1), 87–100.

Hawkins, C. B., & Ratliff, K. A. (2015). Trying but failing: Implicit attitude transfer is not eliminated by overt or subtle objectivity manipulations. *Basic and Applied Social Psychology*, *37*(1), 31–43.

Hebl, M. R., & Mannix, L. M. (2003). The weight of obesity in evaluating others: A mere proximity effect. *Personality and Social Psychology Bulletin*, *29*(1), 28–38.

Heider, F. (1958). *Interpersonal relations*. New York: Wiley.

Henderson-King, E. I., & Nisbett, R. E. (1996). Anti-Black prejudice as a function of exposure to the negative behavior of a single Black person. *Journal of Personality and Social Psychology*, *71*(4), 654–661.

Högden, F., & Unkelbach, C. (2020). The role of relational qualifiers in attribute conditioning: Does disliking an athletic person make you unathletic? *Personality and Social Psychology Bulletin*, 0146167220945538.

Hu, X., Gawronski, B., & Balas, R. (2017). Propositional versus dual-process accounts of evaluative conditioning: I. The effects of co-occurrence and relational information on implicit and explicit evaluations. *Personality and Social Psychology Bulletin*, *43*(1), 17 32.

Huckleberry, K. A., Ferguson, L. B., & Drew, M. R. (2016). Behavioral mechanisms of context fear generalization in mice. *Learning & Memory*, *23*(12), 703–709.

Hughes, S., De Houwer, J., Mattavelli, S., & Hussey, I. (2020). The shared features principle: If two objects share a feature, people assume those objects also share other features. *Journal of Experimental Psychology: General*, *149*(12), 2264–2288.

Hussinger, K., & Pellens, M. (2017). Guilt by association: How scientific misconduct harms prior collaborators. ZEW – Centre for European Economic Research Discussion Paper No. 17-051.

Katz, D., & Braly, K. (1933). Racial stereotypes of one hundred college students. *Journal of Abnormal and Social Psychology*, *28*, 280–290.

Klein, S. B. (2019). *Learning principles and applications* (8th Ed.). New York: McGraw Hill.

Kraus, M. W., & Chen, S. (2010). Facial-feature resemblance elicits the transference effect. *Psychological Science*, *21*(4), 518–522.

Kubota, J. T., Mojdehbakhsh, R., Raio, C., Brosch, T., Uleman, J. S., & Phelps, E. A. (2014). Stressing the person: Legal and everyday person attributions under stress. *Biological Psychology*, *103*, 117–124.

Kurdi, B., & Banaji, M. R. (2017). Repeated evaluative pairings and evaluative statements: How effectively do they shift implicit attitudes?. *Journal of Experimental Psychology: General*, *146*(2), 194–213.

Kurdi, B., & Dunham, Y. (2020). Propositional accounts of implicit evaluation: Taking stock and looking ahead. *Social Cognition*, *38*(Supplement), s42–s67.

Kurdi, B., Krosch, A. R., & Ferguson, M. J. (2020). Implicit evaluations of moral agents reflect intent and outcome. *Journal of Experimental Social Psychology*, *90*, 103990.

Lickel, B., Hamilton, D. L., Wieczorkowska, G., Lewis, A., Sherman, S. J., & Uhles, A. N. (2000). Varieties of groups and the perception of group entitativity. *Journal of Personality and Social Psychology*, *78*(2), 223.

Medin, D. L., Goldstone, R. L. & Gentner, D. (1993). Respects for similarity. *Psychological Review*, *100*, 254–278.

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences*, *32*(2), 183.

Moors, A., & De Houwer, J. (2006). Automaticity: a theoretical and conceptual analysis. *Psychological Bulletin*, *132*(2), 297–326.

Moran, T., & Bar-Anan, Y. (2013). The effect of object–valence relations on automatic evaluation. *Cognition & Emotion*, *27*(4), 743–752.

Moskowitz, G. B., & Li, P. (2011). Egalitarian goals trigger stereotype inhibition: A proactive form of stereotype control. *Journal of Experimental Social Psychology*, *47*(1), 103–116.

Olson, M. A., & Fazio, R. H. (2002). Implicit acquisition and manifestation of classically conditioned attitudes. *Social Cognition*, *20*, 89–103.

Olson, M. A., & Fazio, R. H. (2006). Reducing automatically activated racial prejudice through implicit evaluative conditioning. *Personality and Social Psychology Bulletin*, *32*(4), 421 433.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716.

Paluck, E. L., Porat, R., Clark, C. S., & Green, D. P. (2021). Prejudice reduction: Progress and challenges. *Annual Review of Psychology*, *72*, 533–560.

Pavlov, I. P. (1927). *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex*. Oxford, England: Oxford University Press.

Pietri, E. S., Fazio, R. H., & Shook, N. J. (2013). Weighting positive versus negative: The fundamental nature of valence asymmetry. *Journal of Personality*, *81*(2), 196–208.

Pryor, J. B., Reeder, G. D., & Monroe, A. E. (2012). The infection of bad company: Stigma by association. *Journal of Personality and Social Psychology*, *102*(2), 224–241.

Ratliff [Ranganath], K. A., & Nosek, B. A. (2008). Implicit attitude generalization occurs immediately; explicit attitude generalization takes time. *Psychological Science*, *19*, 249–254.

Ratliff, K. A., & Nosek, B. A. (2011). Negativity and outgroup biases in attitude formation and transfer. *Personality and Social Psychology Bulletin*, *37*(12), 1692–1703.

Ratliff, K. A., Swinkels, B. A., Klerx, K., & Nosek, B. A. (2012). Does one bad apple (juice) spoil the bunch? Implicit attitudes toward one product transfer to other products by the same brand. *Psychology & Marketing*, *29*(8), 531–540.

Rerick, P. O., Livingston, T. N., & Miller, M. K. (2021). Guilt by association: mock jurors' perceptions of defendants and victims with criminal family members. In press at *Psychology, Crime & Law*.

Riskind, J. H., Williams, N. L., Gessner, T. L., Chrosniak, L. D., & Cortina, J. M. (2000). The looming maladaptive style: Anxiety, danger, and schematic processing. *Journal of Personality and Social Psychology*, *79*(5), 837–852.

Rothbart, M., & Park, B. (1986). On the confirmability and discomfirmability of trait concepts. *Journal of Personality and Social Psychology*, *50*, 131–142.

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, *5*, 296–320.

Secord, P. F. (1959). Stereotyping and favorableness in the perception of Negro faces. *Journal of Abnormal and Social Psychology*, *59*, 309–314.

Shanks, D. R. (1995). *The psychology of associative learning*. London: Cambridge University Press.

Shimizu, Y., Lee, H., & Uleman, J. S. (2017). Culture as automatic processes for making meaning: Spontaneous trait inferences. *Journal of Experimental Social Psychology*, *69*, 79–85.

Shook, N. J., Fazio, R. H., & Eiser, J. R. (2007). Attitude generalization: Similarity, valence, and extremity. *Journal of Experimental Social Psychology*, *43*(4), 641–647.

Skinner, A. L., Olson, K. R., & Meltzoff, A. N. (2020). Acquiring group bias: Observing other people's nonverbal signals can create social group biases. *Journal of Personality and Social Psychology*, *119*(4), 824–838.

Skowronski, J. J., Carlston, D. E., Mae, L., & Crawford, M. T. (1998). Spontaneous trait transference: Communicators take on the qualities they describe in others. *Journal of Personality and Social Psychology*, *74*(4), 837–848.

Smith, C. T., Calanchini, J., Hughes, S., Van Dessel, P., & De Houwer, J. (2019). The impact of instruction-and experience-based evaluative learning on IAT performance: A Quad model perspective. *Cognition and Emotion*, *34*, 21–41.

Spruyt, A., Klauer, K. C., Gast, A., De Schryver, M., & De Houwer, J. (2014). Feature-specific attention allocation modulates the generalization of recently acquired likes and dislikes. *Experimental Psychology*, *61*(2), 85–98.

Stark, T. H., Flache, A., & Veenstra, R. (2013). Generalization of positive and negative attitudes toward individuals to outgroup attitudes. *Personality and Social Psychology Bulletin*, *39*(5), 608–622.

Thompson, M. M., Naccarato, M. E., Moskowitz, G. B., & Parker, K. J. (2001). The personal need for structure and personal fear of invalidity measures: Historical perspectives, current applications, and future directions. In G. B. Moskowitz (Ed.), *Cognitive social psychology: The Princeton symposium on the legacy and future of social cognition* (pp. 19–40). Mahwah, NJ: Erlbaum.

Thompson, E. P., Roman, R. J., Moskowitz, G. B., Chaiken, S., & Bargh, J. A. (1994). Accuracy motivation attenuates covert priming: The systematic reprocessing of social information. *Journal of Personality and Social Psychology*, *66*(3), 474.

Till, B. D., & Priluck, R. L. (2000). Stimulus generalization in classical conditioning: An initial investigation and extension. *Psychology and Marketing*, *17*, 55–72.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*, 327–352.

Uleman, J. S. (1987). Consciousness and control: The case of spontaneous trait inferences. *Personality and Social Psychology Bulletin*, *13*(3), 337–354.

Uleman, J. S. (2005). On the inherent ambiguity of traits and other mental concepts. In B. F. Malle, & S. D. Hodges (Eds.), *Other minds: How humans bridge the divide between self and others*. (pp. 253–267). Guilford Press.

Uleman, J. S. (2009). Automatic (spontaneous) propositional and associative learning of first impressions. *Behavioral and Brain Sciences*, *32*(2), 227–228.

Uleman, J. S., Adil Saribay, S., & Gonzalez, C. M. (2008). Spontaneous inferences, implicit impressions, and implicit theories. *Annual Review of Psychology*, *59*, 329–360.

Uleman, J. S., Newman, L. S., & Moskowitz, G. B. (1996). People as flexible interpreters: evidence and issues from spontaneous trait inference. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 28, pp. 211–279). Academic Press.

Van Dessel, P., Cummins, J., Hughes, S., Kasran, S., Cathelyn, F., & Moran, T. (2020). Reflecting on 25 years of research using implicit measures: Recommendations for their future use. *Social Cognition*, *38*, s223–s242.

Van Dessel, P., De Houwer, J., & Smith, C. T. (2018). Relational information moderates approach-avoidance instruction effects on implicit evaluation. *Acta Psychologica*, *184*, 137–143.

Verosky, S. C., & Todorov, A. (2010). Generalization of affective learning about faces to perceptually similar faces. *Psychological Science*, *21*(6), 779–785.

Wallach, M. A., Kogan, N., & Bem, D. J. (1962). Group influence on individual risk taking. *The Journal of Abnormal and Social Psychology*, *65*(2), 75–86.

Walther, E. (2002). Guilty by mere association: evaluative conditioning and the spreading attitude effect. *Journal of Personality and Social Psychology*, *82*, 919–934.

Walther, E., Nagengest, B., & Traselli, C. (2005). Evaluative conditioning in social psychology: Some facts and speculations. *Cognition and Emotion*, *19*, 175–196.

Xu, M., Briñol, P., Gretton, J. D., Tormala, Z. L., Rucker, D. D., & Petty, R. E. (2020). Individual differences in attitude consistency over time: The personal attitude stability scale. *Personality and Social Psychology Bulletin*, *46*(10), 1507–1519.

Zanon, R., De Houwer, J., Gast, A., & Smith, C. T. (2014). When does relational information influence evaluative conditioning?. *Quarterly Journal of Experimental Psychology*, *67*(11), 2105–2122.

Part III

The Malleability of First Impressions

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

16 Origins of Impression Formation in Infancy

Brandon M. Woo[1] *and J. Kiley Hamlin*[2]
[1]*Harvard University*
[2]*University of British Columbia*

Imagine being a new student at a school where you do not know anyone. On your first day, you observe a teacher drop a $20 bill in the hallway. Another nearby schoolmate also sees the teacher drop the money, and rather than return it he quickly picks it up, pockets it for himself, and runs away. After witnessing this behavior, what would you think of the schoolmate? What kind of person is he? Would he make a good social partner to you? If you needed help, would he help you?

To determine a person's value as a potential social partner, we must be able to accurately assess their character. One way to learn about someone's character is by interacting with them directly, but if the person is likely to harm you, direct interaction could be costly. An alternative strategy is to form an impression by observing how the person acts toward others. Seeing someone steal another person's money, for example, might lead you to infer that the person who stole is bad, may later treat you poorly, etc. The ability to form such impressions facilitates selecting appropriate social partners, enabling better navigation of the social world.

A large body of work has demonstrated that adults readily and quickly form character impressions from others' behaviors. Indeed, adults assess character traits based on highly minimal behavioral information (Ambady & Rosenthal, 1992; Ambady et al., 1995), without intending to do so (Uleman et al., 2005; Winter & Uleman, 1984), and even when they arguably should not, when there is situational information that readily explains the behavior (Brosch et al., 2013; Gilbert & Malone, 1995; Jones & Harris, 1967; Kubota et al., 2014; Ross, 1977). Because of the spontaneity and speed with which people form behavior-based character impressions, such impressions have been referred to as implicit impressions (Ferguson et al., 2019; Uleman et al., 2005; Uleman et al., 2008).

Importantly, adults' impressions are sensitive not only to the outcomes associated with others' behaviors, but to the intentions underlying them (Heider, 1958; Jones & Davis, 1965; Malle, 1999). A sensitivity to mental states (rather than just outcomes) enables adults to distinguish between, for example, agents who try to help and harm others but fail, or agents who intentionally versus accidentally help and harm (Cushman et al., 2006; Kurdi et al., 2020; Young

DOI: 10.4324/9781003045687-19

et al., 2007). An agent who intends to harm (even if not successful) may be more likely to harm again than an agent who just so happens to help (despite not intending to). That is, a sensitivity to intentions rather than outcomes may enable us to make better social choices as well as more accurate predictions about others' future behavior (see Heider, 1958; Jones & Davis, 1965).

How do adults come to form impressions of others based on others' intentional behaviors so readily and so quickly? One possibility is that humans are very well *practiced* at impression formation by adulthood: Over development, humans have plenty of opportunities to learn about how others' intentional behaviors relate to their character traits, whether it be via personal experience interacting with others over time, observations of others' behavioral regularities, or explicit teaching. A second, non-mutually exclusive possibility, is that humans have somehow *evolved* to form impressions: Over evolutionary history, humans more capable of inferring character from others' intentional behaviors may have outperformed those less capable of impression formation, particularly within the intensely cooperative relationships that characterize the human species. If so, humans' capacity for impression formation may have evolved alongside capacities for cooperation, and may begin to emerge in our species independently of protracted learning processes.

In this chapter, we review a growing body of work suggestive that several key precursors to adults' ability to form impressions are present in infancy, from a few months after birth. Specifically, infants evaluate others based on their behaviors, specifically their intentional ones, and use their impressions to make inferences about others' future social behavior. Although these capacities have been studied for a range of social behaviors (e.g., Mascaro & Csibra, 2012; Powell & Spelke, 2018; Thomas & Sarnecka, 2019; Thomas et al., 2018), here we focus on whether infants form impressions based on morally relevant behaviors. We do so in part because such impressions would presumably deliver the most benefit to developing humans, allowing them to best determine who might treat them well or poorly. Further, most research to date has focused on infants' impressions in the moral domain. Together, this work suggests that tendencies toward impression formation ground humans' earliest interactions with the social world.

Studying Impression Formation in Infancy

Whereas adults can be told how an individual has behaved and then verbally report their impressions and their predictions of the individual's later behavior, infants can neither understand nor produce language. How, then, can we begin to study impression formation in infant populations? Instead of telling infants how an individual has behaved, studies of early impression formation typically present infants with puppet shows or animated events involving agents (sometimes human, but often infant-friendly cartoon characters or puppets) who engage in distinct social behaviors; for example, helping versus hindering another agent (e.g., Hamlin et al., 2007, 2010), and acting fairly versus unfairly

(e.g., Geraci & Surian, 2011). Notably, just as adults can form impressions of unknown individuals based solely on observation (e.g., without being a recipient of an act), infants in these studies are mere bystanders of positive and negative interactions occurring between novel third parties. That is, infants do not experience any positive or negative treatment themselves.

After exposing infants to positive and negative third-party behaviors, infants' impressions are probed nonverbally in two main ways. First, infants' evaluations of prosocial and antisocial agents are examined through various measures of relative preference for one agent versus another, including preferential looking, reaching, and approach. Here, as in much classic work in infant cognition, systematic orienting is taken as indicative of discrimination and differential evaluation (see Fantz, 1961). Second, infants' expectations for how particular agents will behave are examined by measuring the amount of time infants attend following different events. Here, systematic attentional differences are taken as indicative of differential expectations (see Aslin, 2007). Our central claim, that impression formation begins in infancy, is based on evidence that infants both *evaluate* novel agents based on their prosocial and antisocial behaviors, and form *expectations* for agents' future pro- and antisocial behaviors based on their past ones.

Evaluations Based on Morally Relevant Behavior

Although adults evaluate others based on a wide variety of social actions, we are particularly likely to form impressions based on others' morally relevant behaviors (e.g., De Bruin & Van Lange, 1999; Wojciszke et al., 1998). In this section, we review evidence that, like adults, preverbal infants evaluate agents based on whether they help versus hinder or harm others, and whether they act fairly versus unfairly to others.

Evaluations Based on Helping and Hindering/Harming

Helping and Hindering

In order to be sensitive to whether someone has helped or hindered an agent pursuing its goals, one must first represent the agent's goals. Beginning with studies by Woodward (1998), a large body of research has provided evidence that infants in the first year can infer the goals of others' actions (for review, see Woodward, 2009; see also, Sommerville et al., 2005; Woo et al., 2020).[1] Given that infants appear able to represent others' goals, how do infants respond to agents who facilitate versus prevent those goals?

In the first set of studies to examine infants' evaluations of morally relevant behavior, Hamlin et al. (2007) showed 6- and 10-month-olds a protagonist (a colored shape with eyes) who tried but failed to climb a hill on its own. In alternating events, one agent (a helper) pushed the protagonist up the hill, allowing it to achieve its goal, and a second agent (a hinderer) pushed the

protagonist down the hill, preventing it from achieving its goal. Following sufficient familiarization to these events (i.e., infants were "habituated"), Hamlin et al. presented infants with the helper and the hinderer, and examined which agent infants chose to touch first. Both 6- and 10-month-olds preferentially reached first for the helper rather than the hinderer. In follow-up studies, Hamlin et al. found that 6- and 10-month-olds also preferred a helper to a neutral agent, and a neutral agent to a hinderer, suggestive of both positive and negative evaluation; further, infants did not distinguish between agents in an inanimate control condition in which the protagonist was replaced with an inanimate object (a colored shape without eyes that did not move by itself) that could not have had goals. Taken together, these findings suggest that infants saw helping and hindering actions as meaningfully different behaviors, evaluated helping positively and hindering negatively, and that infants' evaluations were based on those actions' social, not physical, characteristics.

This first set of studies has led to what is now a large body of research on infants' capacities for social evaluation. For instance, infants' evaluations of helpers vs. hinderers have been examined in even younger age groups using preferential looking methods (as infants under ~5 months of age cannot reliably reach for objects), and been found to be present as early as 3 months after birth (Hamlin et al., 2010; Hamlin & Wynn, 2011). Infants' evaluations have also been examined in other helping and hindering scenarios; for instance, infants prefer an agent who helps a protagonist (in this case a stuffed animal puppet) in its attempts both to open a box containing an attractive object and to retrieve a ball that it accidentally dropped (Hamlin & Wynn, 2011). Importantly, these studies also included control conditions in which the protagonist was replaced with an inanimate object; here infants did not prefer the characters that either opened the box or returned the ball, providing additional evidence that infants' evaluations only apply to social behaviors. These findings suggest that infants are sensitive to who helps versus hinders agents in possession of a variety of unfulfilled goals.

Other labs have now investigated and replicated these findings (e.g., Scola et al., 2015; Woo & Spelke, 2020a). Although some individual studies have failed to replicate a preference for helpers over hinderers (Cowell & Decety, 2015; Salvadori et al., 2015; Schlingloff et al., 2020), a recent meta-analysis found a significant preference for helpers over hinderers across published and unpublished studies, including when controlling for possible publication bias (Margoni & Surian, 2018).

Harming

A closely related body of work suggests that infants also negatively evaluate aggressive agents who cause others physical harm. For instance, Kanakogi et al. (2013) presented 10-month-olds with an aggressor who hit a victim, and found that infants preferentially reached to the victim over the aggressor, and to a neutral agent over the aggressor, suggestive that infants generally

dislike physically aggressive agents. In a related study using human agents, Buon et al. (2014) found that infants' relative evaluations of two human agents engaging in the very same physical actions (both "comforting" and "hitting") depended on the targets of their actions: 10-month-olds selectively accepted a toy from someone who comforted another human and hit a backpack, over someone who hit the human and comforted the backpack. These findings further support the claim that infants' evaluations are selective to social contexts, and provide additional evidence that infants evaluate others based on their morally relevant behaviors.

Evaluations Based on Fair and Biased Resource Distribution

By some time in the second year, infants' evaluations are not only sensitive to helping, hindering, and harming, but to another kind of morally relevant behavior—whether someone distributes resources fairly or unfairly. In one study, Geraci and Surian (2011) found that 16-month-olds (though not 10-month-olds) preferentially reached for an agent who had previously distributed resources equally over one who had distributed resources unequally. Similarly, Lucca et al. (2018) found that both 13-month-olds and 17-month-olds preferentially approached and accepted a toy from a person who had distributed resources equally over a person who had distributed resources unequally (for similar findings in 15-month-olds, see Burns & Sommerville, 2014). This growing evidence suggests that infants evaluate others based on how they distribute resources.

The Role of Mental States in Infants' Evaluations

The studies reviewed thus far demonstrate that infants may form impressions and evaluate others based on their morally relevant behaviors. However, they do not address a critical factor that is central to adults' impressions (as reviewed above; Heider, 1958, Jones & Davis, 1965; Malle, 1999) and that influences adults' evaluations of moral agents (Cushman et al., 2006; Kurdi et al., 2020; Young et al., 2007): their mental states. An agent's intentions, rather than the outcomes of their behavior, may be particularly informative about their future social behavior and about their value as social partners (see Heider, 1958; Jones & Davis, 1965). Here, we ask: How do infants weigh mental states versus outcomes in their evaluations?

In the infant work reviewed thus far, agents' intentions have been consistent with their associated outcomes; that is, agents who had the intention to help a protagonist successfully facilitated the protagonist's goal, and agents who had the intention to hinder successfully prevented the protagonist's goal. One possibility, then, is that infants merely prefer agents who bring about positive outcomes such as someone achieving their goals over agents who bring about negative outcomes such as someone failing to achieve their goals.

Here we review studies suggesting that infants' evaluations are mentalistic: They incorporate, and even privilege, the mental states that drive others'

behaviors. For instance, by late in the first year, infants' evaluations appear to be based on what others intend to do, whether or not they successfully achieve it. In a set of studies, Hamlin (2013) showed 8- and 5-month-olds agents who demonstrated an intention to either help or to hinder a third party's attempts to open a box, but were either successful or unsuccessful at achieving their helpful and harmful goals. In several conditions pitting different combinations of failed and successful helpers and hinderers, 8-month-olds (though not 5-month-olds, who chose entirely randomly) consistently reached for agents who had intended to help over those who had intended to hinder, regardless of whether or not they successfully did so. Notably, when both agents had the same intention but only one was successful (that is, one agent was clearly associated with the better outcome), 8-month-olds did not distinguish between the agents. Together, these results suggest that holding a positive or negative intention may be both necessary and sufficient for infants to form positive and negative evaluations, and that infants do not evaluate others based solely on the outcomes they are associated with (see Kanakogi et al., 2017, for complementary evidence).

Additional evidence that infants' evaluations are based on intention rather than outcomes comes from Woo et al. (2017), who explored how infants assess agents who cause valenced outcomes for others without intending to (e.g., accidental helping and hindering). Here, all agents pushed over a tall shelf on top of which an attractive toy rested, causing the toy to fall to the ground. Critically, for some infants the toy on the ground was a good outcome: A protagonist had been trying but failing to reach it. For other infants the toy on the ground was a bad outcome: The protagonist had just put the toy away on the top of the shelf. Further, for all infants one puppet knocked the shelf down intentionally, whereas the other knocked it down accidentally, on its way to do something else and without having seen the protagonist demonstrate its goal (of either trying to reach the toy or trying to put the toy on the shelf). Woo et al. found that 10-month-olds consistently distinguished between the intentional and accidental agents, but did so differently depending on whether knocking over the shelf was helping or hindering: Infants preferred an intentional helper to an accidental helper, but preferred an accidental hinderer over an intentional hinderer. Thus, infants seem to positively evaluate agents who have positive/helpful mental states, and negatively evaluate agents who have negative/harmful ones, even when all agents perform basically the same physical acts.

Other studies suggest that infants' impressions are sensitive to a wider range of mental states, such as what others know and believe. In a study examining the role of knowledge and ignorance in infants' evaluations, Hamlin et al. (2013) found that 10-month-olds preferred an agent who caused a positive versus a negative outcome for a protagonist, as long as the agents could have known that they were being helpful and unhelpful when they intervened (because they knew what the protagonist wanted). If the agents could not have known which type of intervention would be helpful versus unhelpful, then

infants chose randomly. More recent studies have built on these findings to examine the role of knowledge in infants' evaluations of accidents (Woo et al., 2017), and the role of true and false beliefs in infants' evaluations of actions that have helpful and unhelpful outcomes (Woo & Spelke, 2020b). Together, these studies suggest that infants' evaluations privilege an agent's helpful intentions (based on their knowledge or ignorance, or their true or false beliefs) over the valence of the outcomes they cause.

Summary: Evaluations Based on Morally Relevant Behavior

This growing body of research demonstrating infants' differential evaluation of individuals who act in various prosocial and antisocial ways provides evidence that a key precursor to adults' capacities for impression formation emerges in infancy. Like adults', infants' evaluations are based not only on others' morally relevant behaviors, but on the mental states—the intentions, knowledge states, and beliefs—that underlie those behaviors. Such an early capacity for mentalistic social evaluation could be adaptive, enabling infants to select better social partners.

Inferences About Future Behavior

In adults, our impressions support not only our evaluations of others, but also our inferences about how others are likely to act in the future. That is, in watching one agent harm another, we both evaluate the agent negatively, and can predict that the agent may continue to harm both that individual and others (Heider, 1958; Jones, 1979; McCarthy & Skowronski, 2011). These predictions play a key role in supporting adults' ability to pursue benefit and avoid harm. Here, we ask: Does infants' capacity for impression formation go beyond mere evaluation, to support inferences about whether an agent will later act prosocially or antisocially? In this section, we first review evidence that infants may form such inferences, and then consider whether infants' inferences demonstrate any of the biases known to influence adults' inferences.

Infants' Behavioral Inferences

Evidence in support of the possibility that infants' impressions support inferences of future behavior comes from studies that have taken advantage of a highly reliable phenomenon in infant research: that infants typically look longer following events that are surprising or unexpected than following events that are unsurprising or expected (for review, see Aslin, 2007). Several studies to date have used such violation-of-expectation paradigms to investigate whether infants expect agents to act in ways that are consistent with their past behaviors. The logic of these studies is in line with adult research on impression formation: that behaviors that are inconsistent with adults' expectations require increased processing (e.g., Sherman et al., 1998).

Here, we reason that if infants' evaluations reflect impressions that go beyond mere evaluation, then infants should look longer when agents act in impression-inconsistent ways.

In one study, Tatone et al. (2015) presented 12-month-olds with one agent who initially consistently gave resources to a protagonist, and a second agent who consistently took resources away from the protagonist. In later trials, the agents either treated the protagonist as they had previously (e.g., gave the target resources when they had previously given the target resources), or behaved in an opposing way (e.g., gave the target resources when they had previously taken resources from them). Infants looked longer when agents changed their behavior, perhaps because it was inconsistent with infants' initial impressions. Following similar logic, Woo and Hamlin (in revision) presented 11-month-olds with agents who first consistently helped or consistently hindered a protagonist, and later changed their behavior from helping to hindering or from hindering to helping. Looking times revealed that once again, infants looked longer following inconsistent acts than following consistent ones, suggestive that they find behavioral changes surprising. Together, these studies suggest that infants notice, and look longer, when agents change from acting prosocially to antisocially or vice versa, which may reflect that they generate impressions whereby they expect prosocial agents to later act prosocially whereas antisocial agents to later act antisocially.

One alternative explanation for these results, however, is that infants merely noticed the physical differences that exist between helping/hindering and giving/taking events, and looked longer following events containing those physical changes. Indeed, infants often look longer at events containing physical inconsistencies (for review, see Colombo & Mitchell, 2009), and impression formation (in adults) presumably reflects expecting someone to continue behaving in the same *general* way, as opposed to the same very specific (i.e., physically identical) way. Critically, two sets of studies argue against this alternative possibility. First, Surian et al. (2018) found that 14-month-olds looked longer when an agent who previously helped a third party later distributed resources to two new agents unequally (an antisocial action) as opposed to equally (a new prosocial action), even though neither type of distribution was physically similar to helping. This suggests that infants may not look longer at all types of inconsistencies, but only those that are inconsistent with their impressions.

A second set of studies asking whether infants expect agents to act in impression-consistent ways approached the question somewhat differently. Specifically, Taborda-Osorio and colleagues (2019) adapted a method traditionally used in the infant object tracking literature, which suggests that infants can generate inferences about the likely number of objects hiding behind a screen (Xu & Carey, 1996). Using this method, they asked: Do infants' impressions of morally relevant behaviors lead them to generate inferences about how many *agents* are present in a scene? In one experiment, infants were familiarized to events in which a protagonist tried but failed to open a box.

On some trials, an agent came out from behind a screen and helped the protagonist; on other trials, a second identical agent came out from behind the screen and hindered the protagonist. Critically, infants only ever saw one agent at a time, and since the two agents were identical, it was unclear whether one or two agents were present behind the screen. During test events, the screen was lowered to reveal either one agent or two, and infants' attention was measured to each outcome. In the condition where infants viewed both helping and hindering, they looked reliably longer when there was only one agent behind the screen, suggesting that they expected there to be two agents: one helper and one hinderer. That is, infants seemed reluctant to assume that an agent who had helped would also hinder, perhaps reflecting that they expect individual agents to act in impression-consistent ways.

To ensure that this pattern of looking was not because seeing any physically distinct actions would lead infants to infer two agents, in a second experiment Taborda-Osorio et al. kept almost everything the same as their first experiment. However, during familiarization events there was no protagonist who tried to open a box, and so the identical agents simply opened versus closed a box without helping or hindering. Thus, the agents' opening and closing acts—though physically different—were neither social nor moral in nature. Here, infants' looking times did not differ to the one-agent versus two-agent outcomes, suggesting infants were unsure whether one agent or two had performed the acts. Together, these results suggest that infants believe that the same agent can engage in distinct physical behaviors, but cannot engage in distinct moral behaviors. The notion that one agent cannot both help and hinder lends further support to the possibility that infants generate impressions of others based on their morally relevant behaviors.

A remaining question is what the depth of infants' impressions is. Infants may just see helping as being positive, and thus expect that a helper will be associated with more positive things. Alternatively, infants may expect that helpers will only be associated with more positive, prosocial acts. Future work should tease apart these possibilities.

Are Infants' Inferences Biased?

Although evidence suggests that infants may expect agents to be consistent in their morally relevant behavior, there is also some evidence that infants' processing of inconsistent agents may be subject to valence biases: that infants may find an agent's prosocial and antisocial behaviors differently informative about that agent's future behavior. That said, to date the evidence is rather mixed as to exactly which valence bias infants may hold. On the one hand, there is some evidence that infants may form inferences about agents' future behavior in ways that are in line with Skowronski and Carlston's (1989) diagnosticity theory, whereby infants may assume that agents who are fundamentally bad may sometimes, or even usually, act prosocially, but that agents who are fundamentally good do not act antisocially (see also, Ferguson

et al., 2019). Consistent with diagnosticity theory, although Surian et al. (2018) found (as reviewed above) that 14-month-olds expected that an agent who had formerly helped a third party would later distribute resources fairly between two new parties, infants at the same age seemed to lack expectations about whether an agent who had formerly hindered a third party would later distribute resources fairly versus unfairly between two parties: Infants looked equally when a former hinderer distributed fairly versus unfairly, suggesting they thought a hinderer might perform either type of act. Similarly, Woo and Hamlin (in revision) found preliminary evidence of a negativity bias in their work on 11-month-olds' expectations of moral consistency. Specifically, although they found (as reviewed previously) that 11-month-olds as a group expected agents to be consistent in their morally relevant behavior, they found that this effect was primarily driven by infants looking longer when an agent hindered a target after first helping the same target, versus when an agent helped a target after first hindering the same target. These findings are in line with evidence that infants' attribution of agency and evaluations of accidents are subject to negativity biases (see Hamlin & Baron, 2014; Woo et al., 2017). Moreover, these findings provide further evidence that infants may view both bad and good agents as being capable of acting prosocially, but good agents as being incapable (or less capable) of acting antisocially.

On the other hand, there is also other evidence for the exact opposite pattern of reasoning in infants. In one study using puppet shows (based on those of Hamlin et al., 2011), Shimizu et al. (2018) showed 6-, 9-, 12-, and 15- to 18-month-old infants alternating events in which one agent consistently helped a protagonist open a box, and a second agent consistently hindered the same protagonist by slamming the box shut.[2] After infants had habituated to these events, Shimizu et al. presented infants with the initially prosocial and antisocial agents and the protagonist in a new scenario, wherein the protagonist lost control of its ball. The initially prosocial and antisocial agents either behaved consistently or inconsistently with their earlier behavior, by either giving the ball back or stealing it away. Shimizu et al. found that infants' looking time suggested that they expected that the initially antisocial agent to take the ball away from the protagonist in this new scenario, but that they did not expect the initially prosocial agent to return the ball to the protagonist in this new scenario. This is the opposite pattern of findings to the work of Surian et al. (2018) and of Woo and Hamlin (in revision), and is inconsistent with diagnosticity theory.

Summary: Inferences About Future Behavior

In sum, infants' impressions appear to support at least some inferences about agents' future morally relevant behaviors. The evidence is mixed, however, as to what kind of valence bias, if any, influences infants' processing. Although some studies have provided evidence that infants expect helpers to continue to be prosocial but do not expect hinderers to continue to be antisocial

(e.g., Surian et al., 2018; Woo & Hamlin, in revision), other studies have not found evidence that infants' processing of inconsistent agents differs for agents who are prosocial and antisocial (e.g., Tatone et al., 2015; see also, Taborda-Osorio et al., 2019), and another study still has found evidence that infants expect hinderers to continue to be antisocial but do not expect helpers to continue to be prosocial (Shimizu et al., 2018). Future work should examine what features of different studies may lead to such differences in infants' processing of inconsistency, and determine whether valence-related null findings within this literature really reflect biases (e.g., rather than not having enough power to detect an effect).

Impression Updating

Given that infants appear to form impressions based on others' morally relevant behavior, both evaluating others and making at least some inferences about their likely future behaviors, the present section turns to the question of what consequences there are for infants' social evaluations when agents act inconsistently. To date, most research examining infants' social evaluations has presented infants with agents who behave consistently (e.g., always helping, always acting fairly). Further, infants seem to notice (e.g., look longer) when others act in impression-inconsistent ways. Upon encountering such inconsistency, do infants subsequently update their impressions? If so, how (see Moskowitz et al., this volume)?

One strategy for impression updating might involve aggregating across a person's many behaviors, calculating a "summary statistic" for how prosocial or antisocial a person is on average, and using that statistic to determine how likely someone will be prosocial in the future and whether they will make a good social partner. For instance, if you have observed a person act prosocially half the time and antisocially half the time, then your summary statistic might lead you to expect them to be prosocial half the time in the future, and neutral as to their general status as a good or bad social partner. In making social choices, you could simply compare different individuals' summary statistics and determine who might make a relatively better social partner. Although generating and comparing summary statistics of individuals' histories of prosocial vs. antisocial behavior may seem fairly straightforward, impression updating may nevertheless be a challenge, for both adults and infants.

(How) Do Adults Update Their Impressions?

A host of research has established that adults' impressions can be long-lasting and difficult to update, particularly adults' implicit impressions (Ferguson et al., 2019; Gregg et al., 2006; Todorov & Uleman, 2004), and that impression updating is subject to various biases. For instance, some studies suggest that adults' impressions are subject to a "first impression bias," or the systematic prioritization of the first information received about an individual

(Anderson, 1965; Anderson & Hubert, 1963). In the case of a first impression bias, having first seen a person help others, adults might privilege this earlier positive information in their overall (good) impression of the person, even if most of the person's subsequent behaviors are negative.

Other studies have documented that first impression biases may interact with a more general "negativity bias," whereby a person's negative behaviors are prioritized relative to their positive ones (Baumeister et al., 2001; Kanouse & Hanson, 1987; Reeder et al., 1982; Rozin & Royzman, 2001; Uleman & Kressel, 2013; but see Boseovski, 2010, for evidence that children may tend toward a positivity bias). In the case of a negativity bias, having seen a person both steal money from another individual and give money to the poor, adults may focus on the bad over the good in their overall impressions of the person. Such a negativity bias would be consistent with Skowronski and Carlston's (1989) diagnosticity theory (as reviewed previously): that fundamentally good and bad agents can both engage in positive, prosocial behaviors, but that only fundamentally bad agents typically engage in negative, antisocial behaviors. In this way, negative behaviors are viewed as more diagnostic of an agent's overall character than positive ones (see Ferguson et al., 2019). In sum, then, research demonstrates that adults often struggle to update their first impressions (see Moskowitz et al., this volume), and that in some cases when they do, negative behaviors may be prioritized.

(How) Do Infants Update Their Impressions?

Given that infants appear able to form at least some form of rudimentary impressions, they necessarily face the same challenge that adults do: How to make sense of inconsistent behaviors. In one sense, by late in the first year after birth, infants appear able to navigate some situations involving inconsistency, namely between agents' intentions and the moral outcomes that they cause or are associated with (as reviewed above; Hamlin, 2013; Woo et al., 2017). By early in the second year, moreover, infants appear to notice inconsistency in an agent's morally relevant behaviors; for example, when an agent initially engages in one behavior (e.g., helping), and then later engages in a behavior of the opposite valence (e.g., hindering, as reviewed previously; Tatone et al., 2015; Woo & Hamlin, in revision). But do infants evaluate agents who behave inconsistently, and if so, how?

One possibility is that infants' evaluations of inconsistent agents may reflect the kinds of biases that influence adults' impressions and infants' processing of inconsistent agents. For instance, if infants' impressions are subject to a negativity bias, then they should not distinguish between an agent who always harms others over an agent who sometimes helps and sometimes harms others, as both would be considered bad. On the other hand, infants whose impressions are subject to a negativity bias should prefer an agent who always helps over an agent who sometimes helps and sometimes harms, as only the occasional harmer would be considered bad. By contrast, if infants'

impressions are instead subject to a positivity bias, prioritizing an agent's positive behaviors over their negative ones, then infants' relative preferences should be flipped: They should not distinguish consistently helpful from inconsistent agents, but prefer inconsistent agents over consistently harmful ones. Finally, if infants fail to update their impressions at all, because they are subject to something like a first impression bias, then their evaluations should default to their initial impression of agents.

Alternatively, infants' expectations that others will behave consistently over time may be strong enough that they are simply *confused* when others behave in impression-inconsistent ways. If such behavioral inconsistency is too confusing, then infants may both fail to update their impressions and struggle to act on their initial impressions. This final possibility is supported by research that has probed infants' evaluations of inconsistent agents to date, which we review below.

Infants' Evaluations of Agents Who Inconsistently Help and Hinder

Several studies from our laboratory have explored infants' evaluations of agents who behave inconsistently. In both Steckler et al. (2017) and Woo and Hamlin (in revision), experimenters presented infants (at 9 and 11 months of age, respectively) with puppet shows depicting either (i) a consistently helpful agent and an inconsistent agent who was sometimes helpful and sometimes unhelpful; or (ii) a consistently unhelpful agent and an inconsistently helpful/unhelpful agent. In both papers, across four studies involving eight conditions (varying whether the consistent agent had helped or hindered the protagonist), infants did not prefer the relatively more prosocial agent within each pair. That is, neither paper found evidence that infants can aggregate across an agent's behaviors in order to determine who is relatively more prosocial. Additionally, in neither paper was there evidence that infants' impressions reflected a negativity bias, a positivity bias, or a first impression bias.

This lack of positive evidence is striking for at least three reasons. First, these two papers collectively made multiple efforts to facilitate the updating of infants' impressions (e.g., increasing the contrast between the two agents, providing a reason for an inconsistent agent's change in behavior, and reducing working memory demands). Despite these successive methodological improvements, infants failed to prefer the relatively more prosocial agent within each pair. Second, these papers tested infants in a situation in which they could have relied on a first impression bias (as adults sometimes do): The consistent and inconsistent agents initially behaved differently from each other, either in the first pair or even in the first few pairs of trials (e.g., one agent helping, and one agent hindering); only in later trials did the inconsistent agent change its behavior. Further, in most of these studies, the inconsistent agent performed only a single inconsistent act (after several consistent ones). Thus, it seems that infants could have rather easily ignored

agents' later behavior, and focused on what they saw first; indeed, another study by Steckler et al. (2017) confirmed that performing just a few consistent acts is sufficient for infants to generate reliable evaluations in the same paradigm. Finally, as described above the 11-month-olds in Woo and Hamlin (in revision) looked longer following the inconsistent acts they observed, suggestive that they noticed the behavioral change, presumably because they had formed initial impressions of the agents based on their behaviors. That is, they were *sensitive* to the behavioral inconsistency that they observed. Despite this sensitivity to inconsistency, however, the same 11-month-olds still failed to incorporate it into their evaluations. Given all this, that infants failed to distinguish between agents in these studies suggests that inconsistency may dramatically impair infants' evaluative capacities.

In a related study (as reviewed previously), Shimizu et al. (2018) similarly presented 6-, 9-, 12-, and 15- to 18-month-old infants with inconsistent agents. In their study, one agent was initially prosocial and one agent was initially antisocial, and, as in the work of Steckler et al. (2017) and of Woo and Hamlin (in revision), one of these agents later acted inconsistently towards the same protagonist. Shimizu et al. (2018) investigated both whether infants noticed this inconsistency (as reviewed previously), and whether infants formed preferences for the initially prosocial agent (who was also the relatively more prosocial agent of the two). The study of Shimizu et al., however, was designed to *also* examine the influence of adult feedback on infants' expectations and evaluations; thus, whereas most, if not all, other research reported in this chapter took place in settings that were designed to minimize adult influence on infants' behavior (e.g., caregivers close their eyes and are instructed to remain silent), Shimizu et al. instructed caregivers to watch events and talk freely with their infants. Here, despite parental input 6-, 9-, and 12-month-olds did not distinguish between the agents, replicating the null findings of Steckler et al. (2017) and consistent with the null findings of Woo and Hamlin (in revision). In contrast, 15- to 18-month-olds did distinguished between them, preferring the initially/more prosocial agent. Although this suggests that older infants may have updated their impressions (or, perhaps, relied on their initial impressions), given that caregivers in the work of Shimizu et al. (2018) could vocalize their own evaluations to infants, it is difficult to know exactly what led infants to distinguish the agents. Indeed, on average, caregivers voiced their own evaluations at least two times, and the more caregivers did so the stronger infants demonstrated preferences for the initially prosocial agent. Thus, infants may not have been relying on their initial impressions based on what they had observed, but based on what their caregivers told them. Therefore, these findings cannot be easily compared to those of previous work, and do not suggest that infants can engage in impression updating independent from their caregivers. However, this work does suggest that feedback from caregivers may be a key factor in helping infants to overcome the limitations of the early systems for impression formation described thus far.

Conclusions and Implications for Future Research

The ability to form impressions of others based on their behavior enables us to predict others' future behavior and determine who may make a better social partner. Classic studies in psychology have demonstrated that adults readily and quickly form such impressions. In the present chapter, we sought to shed light on how adults come to form impressions with such spontaneity and speed. The evidence that we have reviewed here supports the proposal that the capacity for impression formation is early emerging.

Like adults, infants form social evaluations after observing others' morally relevant behavior. In some ways, infants' evaluations appear remarkably mature, privileging the mental states (e.g., intentions, knowledge states, and beliefs) underlying others' behavior over the outcomes that others cause. Moreover, like adults, infants' initial impressions of agents appear to support inferences about the agents' likely future social behavior. Because infants are nonverbal, this body of research has largely relied on indirect methods (e.g., their preferential reaching and preferential looking behavior, or their looking times following events) to explore early impression formation; these methods are necessarily subject to alternative explanations, some of which we have highlighted throughout the chapter. We look forward to future research using other methods (e.g., neuroimaging) that may provide further tests for whether infants engage in impression formation (see Krol & Grossmann, 2020).

Despite infants' impressive capacities, there is one noteworthy challenge to impression formation that both adults and infants face: that people can be inconsistent. Our review highlights two critical differences between adults and infants. First, whereas studies on adults have provided evidence that adults' processing of inconsistency is susceptible to first impression and negativity biases, studies on infants have provided conflicting evidence as to whether infants' processing of inconsistency is susceptible to a negativity bias (Surian et al., 2018; Woo & Hamlin, in revision), a positivity bias (Shimizu et al., 2018), or no bias (Tatone et al., 2015; see also, Taborda-Osorio et al., 2019). Future work will be important to determining why this conflicting evidence exists.

Second, although inconsistency may be difficult for adults, adults nevertheless have impressions of inconsistent agents that they can act on (e.g., that they use to evaluate agents), even if these impressions are susceptible to first impression or negativity biases. By contrast, in two papers, infants (at 9 and 11 months) have not demonstrated preferences when agents have behaved inconsistently. Moreover, at 11 months of age, the same infants who failed to determine which agents were relatively more prosocial also expected agents to be consistent in their helping or hindering behaviors. Despite a growing number of studies that have demonstrated impressive capacities for evaluations and inferences in infancy, then, infants are apparently unable to incorporate inconsistency in their evaluations. Yet, the ability to evaluate people who do not behave consistently is critical to functioning in our everyday social world.

It is an open question as to when and how a sensitivity to inconsistency begins to inform evaluations of inconsistent agents, with some evidence suggesting parental input may play a key role (Shimizu et al., 2018).

As adults, our social life centers around cooperating with other people. How do we come to determine whether someone may cooperate with us? The present chapter makes the case that impression formation has roots in infancy: Beginning in the first year, infants form impressions of others that support their social evaluations and their inferences about others' future behavior. Infants privilege the mental states underlying others' actions over the outcomes of those actions. This early-emerging ability to reason about others' behavior in terms of their character and value as potential social partners may act as a foundation for adults' impressive capacities for impression formation. Future research would be important for understanding the relationship between this foundation and impression formation in older children and adults, and for further characterizing the limitations of this foundation.

Notes

1 Woodward (1998) found that, following sufficient familiarization (i.e., habituation) to events depicting an actress repeatedly reaching to one object (e.g., a bear) over another object (e.g., a ball), 6- and 9-month-old infants looked longer when the actress later reaches for the non-preferred object than for the originally preferred object. These findings support the possibility that infants represented the actress's goal of acting on a particular object. These findings have since been replicated and extended in many labs (e.g., Biro & Leslie, 2007; Choi et al., 2018; Daum et al., 2012; Feiman et al., 2015; Hernik & Southgate, 2012; Luo, 2011; Luo & Johnson, 2009; Sommerville et al., 2005; Woo et al., 2021). This understanding of others' goals may have implications for infants' processing of morally relevant behaviors (Tan & Hamlin, 2022; Woo & Spelke, 2020a).

2 Note that Shimizu et al. (2018) instructed caregivers to watch events and speak freely with their infants while watching events. By contrast, most other studies ask caregivers to minimize their influence on their infants (e.g., closing their eyes, remaining silent). Shimizu et al. were interested in the effects of socialization, however, and their study instructions to caregivers therefore potentially led to infants receiving caregiver feedback about how they should respond to morally relevant behaviors. Indeed, during the study, caregivers on average voiced their own evaluations of agents following at least two helping or hindering actions. That said, Shimizu et al. found no significant relationships between caregiver speech and infants' looking times.

References

Ambady, N., Hallahan, M., & Rosenthal, R. (1995). On judging and being judged accurately in zero-acquaintance situations. *Journal of Personality and Social Psychology*, 69(3), 518–529.

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, *111*(2), 256–274.

Anderson, N. H. (1965). Averaging versus adding as a stimulus-combination rule in impression formation. *Journal of Experimental Psychology*, *70*(4), 394–400.

Anderson, N. H., & Hubert, S. (1963). Effects of concomitant verbal recall on order effects in personality impression formation. *Journal of Verbal Learning and Verbal Behavior*, *2*(5–6), 379–391.

Aslin, R. N. (2007). What's in a look? *Developmental Science*, *10*(1), 48–53.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323–370.

Biro, S., & Leslie, A. M. (2007). Infants' perception of goal-directed actions: Development through cue-based bootstrapping. *Developmental Science*, *10*(3), 379–398.

Boseovski, J. J. (2010). Evidence for "rose-colored glasses": An examination of the positivity bias in young children's personality judgments. *Child Development Perspectives*, *4*(3), 212–218.

Brosch, T., Schiller, D., Mojdehbakhsh, R., Uleman, J. S., & Phelps, E. A. (2013). Neural mechanisms underlying the integration of situational information into attribution outcomes. *Social Cognitive and Affective Neuroscience*, *8*(6), 640–646.

Buon, M., Jacob, P., Margules, S., Brunet, I., Dutat, M., Cabrol, D., & Dupoux, E. (2014). Friend or foe? Early social evaluation of human interactions. *Plos One*, *9*(2), e88612.

Burns, M. P., & Sommerville, J. (2014). "I pick you": The impact of fairness and race on infants' selection of social partners. *Frontiers in Psychology*, *5*, 93.

Choi, Y. J., Mou, Y., & Luo, Y. (2018). How do 3-month-old infants attribute preferences to a human agent? *Journal of Experimental Child Psychology*, *172*, 96–106.

Colombo, J., & Mitchell, D. W. (2009). Infant visual habituation. *Neurobiology of Learning and Memory*, *92*(2), 225–234.

Cowell, J. M., & Decety, J. (2015). Precursors to morality in development as a complex interplay between neural, socioenvironmental, and behavioral facets. *Proceedings of the National Academy of Sciences*, *112*(41), 12657–12662.

Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological Science*, *17*(12), 1082–1089.

Daum, M., Attig, M., Gunawan, R., Prinz, W., & Gredebäck, G. (2012). Actions seen through babies' eyes: A dissociation between looking time and predictive gaze. *Frontiers in Psychology*, *3*, 370.

De Bruin, E. N., & Van Lange, P. A. (1999). Impression formation and cooperative behavior. *European Journal of Social Psychology*, *29*(2–3), 305–328.

Fantz, R. L. (1961). A method for studying depth perception in infants under six months of age. *The Psychological Record*, *71*(1), 194–199.

Feiman, R., Carey, S., & Cushman, F. (2015). Infants' representations of others' goals: Representing approach over avoidance. *Cognition*, *136*, 204–214.

Ferguson, M. J., Mann, T. C., Cone, J., & Shen, X. (2019). When and how implicit first impressions can be updated. *Current Directions in Psychological Science*, *28*(4), 331–336.

Geraci, A., & Surian, L. (2011). The developmental roots of fairness: Infants' reactions to equal and unequal distributions of resources. *Developmental Science*, *14*(5), 1012–1020.

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, *117*(1), 21–38.

Gregg, A. P., Seibt, B., & Banaji, M. R. (2006). Easier done than undone: Asymmetry in the malleability of implicit preferences. *Journal of Personality and Social Psychology*, *90*(1), 1–20.

Hamlin, J. K. (2013). Failed attempts to help and harm: Intention versus outcome in preverbal infants' social evaluations. *Cognition*, *128*(3), 451–474.

Hamlin, J. K., & Baron, A. S. (2014). Agency attribution in infancy: Evidence for a negativity bias. *Plos One*, *9*(5), e96112.

Hamlin, J. K., Ullman, T., Tenenbaum, J., Goodman, N., & Baker, C. (2013). The mentalistic basis of core social cognition: Experiments in preverbal infants and a computational model. *Developmental Science*, *16*(2), 209–226.

Hamlin, J. K., & Wynn, K. (2011). Young infants prefer prosocial to antisocial others. *Cognitive Development*, *26*(1), 30–39.

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, *450*(7169), 557–559.

Hamlin, J. K., Wynn, K., & Bloom, P. (2010). Three-month-olds show a negativity bias in their social evaluations. *Developmental Science*, *13*(6), 923–929.

Hamlin, J. K., Wynn, K., Bloom, P., & Mahajan, N. (2011). How infants and toddlers react to antisocial others. *Proceedings of the National Academy of Sciences*, *108*(50), 19931–19936.

Heider, F. (1958). *The Psychology of Interpersonal Relations*. New York: Wiley.

Hernik, M., & Southgate, V. (2012). Nine-months-old infants do not need to know what the agent prefers in order to reason about its goals: On the role of preference and persistence in infants' goal-attribution. *Developmental Science*, *15*(5), 714–722.

Jones, E. E. (1979). The rocky road from acts to dispositions. *American Psychologist*, *3*, 107–117.

Jones, E. E., & Davis, K. E. (1965). From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 219–266). New York: Academic Press.

Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology*, *3*(1), 1–24.

Kanakogi, Y., Inoue, Y., Matsuda, G., Butler, D., Hiraki, K., & Myowa-Yamakoshi, M. (2017). Preverbal infants affirm third-party interventions that protect victims from aggressors. *Nature Human Behaviour*, *1*(2), 1–7.

Kanakogi, Y., Okumura, Y., Inoue, Y., Kitazaki, M., & Itakura, S. (2013). Rudimentary sympathy in preverbal infants: Preference for others in distress. *Plos One*, *8*(6), e65292.

Kanouse, D. E., & Hanson, L. R., Jr. (1987). Negativity in evaluations. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior*. (pp. 47–62). Lawrence Erlbaum Associates, Inc.

Krol, K. M., & Grossmann, T. (2020). Impression formation in the human infant brain. *Cerebral Cortex Communications*, *1*(1), tgaa070.

Kubota, J. T., Mojdehbakhsh, R., Raio, C., Brosch, T., Uleman, J. S., & Phelps, E. A. (2014). Stressing the person: Legal and everyday person attributions under stress. *Biological Psychology*, *103*, 117–124.

Kurdi, B., Krosch, A. R., & Ferguson, M. J. (2020). Implicit evaluations of moral agents reflect intent and outcome. *Journal of Experimental Social Psychology*, *90*, 103990.

Lucca, K., Pospisil, J., & Sommerville, J. A. (2018). Fairness informs social decision making in infancy. *Plos One*, *13*(2), e0192848.

Luo, Y. (2011). Three-month-old infants attribute goals to a non-human agent. *Developmental Science*, *14*(2), 453–460.

Luo, Y., & Johnson, S. C. (2009). Recognizing the role of perception in action at 6 months. *Developmental Science*, *12*(1), 142–149.

Malle, B. F. (1999). How people explain behavior: A new theoretical framework. *Personality and Social Psychology Review*, *3*(1), 23–48.

Margoni, F., & Surian, L. (2018). Infants' evaluation of prosocial and antisocial agents: A meta-analysis. *Developmental Psychology*, *54*(8), 1445–1455.

Mascaro, O., & Csibra, G. (2012). Representation of stable social dominance relations by human infants. *Proceedings of the National Academy of Sciences*, *109*(18), 6862–6867.

McCarthy, R. J., & Skowronski, J. J. (2011). What will Phil do next? Spontaneously inferred traits influence predictions of behavior. *Journal of Experimental Social Psychology*, *47*, 321–332. 10.1016/j.jesp.2010.10.015

Moskowitz, G. B., Olcaysoy Okten, I., & Schneid, E. (in press). The Updating of First Impressions. In E. Balcetis & G. B. Moskowitz (Eds.), *Handbook of Impression Formation*. New York: Psychology Press/Taylor and Francis.

Powell, L. J., & Spelke, E. S. (2018). Third-party preferences for imitators in preverbal infants. *Open Mind*, *2*(2), 61–71.

Reeder, G. D., Henderson, D. J., & Sullivan, J. J. (1982). From dispositions to behaviors: The flip side of attribution. *Journal of Research in Personality*, *16*(3), 355–375.

Ross, L. (1977). The intuitive scientist and his shortcomings. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 10, pp. 174–220). New York: Academic Press.

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, *5*(4), 296–320.

Salvadori, E., Blazsekova, T., Volein, A., Karap, Z., Tatone, D., Mascaro, O., & Csibra, G. (2015). Probing the strength of infants' preference for helpers over hinderers: Two replication attempts of Hamlin and Wynn (2011). *PloS one*, *10*(11), e0140570.

Schlingloff, L., Csibra, G., & Tatone, D. (2020). Do 15-month-old infants prefer helpers? A replication of Hamlin et al. (2007). *Royal Society Open Science*, *7*(4), 191795.

Scola, C., Holvoet, C., Arciszewski, T., & Picard, D. (2015). Further evidence for infants' preference for prosocial over antisocial behaviors. *Infancy*, *20*(6), 684–692.

Shimizu, Y., Senzaki, S., & Uleman, J. S. (2018). The influence of maternal socialization on infants' social evaluation in two cultures. *Infancy*, *23*(5), 748–766.

Sherman, J. W., Lee, A. Y., Bessenoff, G. R., & Frost, L. A. (1998). Stereotype efficiency reconsidered: Encoding flexibility under cognitive load. *Journal of Personality and Social Psychology*, *75*(3),589.

Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin*, *105*(1), 131–142.

Sommerville, J. A., Woodward, A. L., & Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. *Cognition*, *96*(1), B1–B11.

Steckler, C. M., Woo, B. M., & Hamlin, J. K. (2017). The limits of early social evaluation: 9-month-olds fail to generate social evaluations of individuals who behave inconsistently. *Cognition*, *167*, 255–265.

Surian, L., Ueno, M., Itakura, S., & Meristo, M. (2018). Do infants attribute moral traits? Fourteen-month-olds' expectations of fairness are affected by agents' antisocial actions. *Frontiers in Psychology*, *9*.

Taborda-Osorio, H., Lyons, A., & Cheries, E. W. (2019). Examining infants' individuation of others by sociomoral disposition. *Frontiers in Psychology*, *10*.

Tan, E., & Hamlin, J. K. (2022). Mechanisms of social evaluation in infancy: A preregistered exploration of infants' eye-movement and pupillary responses to prosocial and antisocial events. *Infancy*, *27*(2), 255–276.

Tatone, D., Geraci, A., & Csibra, G. (2015). Giving and taking: Representational building blocks of active resource-transfer events in human infants. *Cognition*, *137*, 47–62.

Thomas, A. J., & Sarnecka, B. W. (2019). Infants choose those who defer in conflicts. *Current Biology*, *29*(13), 2183–2189.

Thomas, A. J., Thomsen, L., Lukowski, A. F., Abramyan, M., & Sarnecka, B. W. (2018). Toddlers prefer those who win but not when they win by force. *Nature Human Behaviour*, *2*(9), 662–669.

Todorov, A., & Uleman, J. S. (2004). The person reference process in spontaneous trait inferences. *Journal of Personality and Social Psychology*, *87*(4), 482–493.

Uleman, J. S., Adil Saribay, S., & Gonzalez, C. M. (2008). Spontaneous inferences, implicit impressions, and implicit theories. *Annual Review of Psychology*, *59*, 329–360.

Uleman, J. S., Blader, S., & Todorov, A. (2005). Implicit impressions. In R. Hassin, J. S. Uleman, & J. A. Bargh (Eds.), *The new unconscious* (pp. 362–392). New York: Oxford University Press.

Uleman, J. S., Kressel, L. M. (2013). A brief history of theory and research on impression formation. In D. E. Carlston (Ed.), *The Oxford Handbook of Social Cognition* (pp. 53–73). New York: Oxford University Press.

Winter, L., & Uleman, J. S. (1984). When are social judgments made? Evidence for the spontaneousness of trait inferences. *Journal of Personality and Social Psychology*, *47*(2), 237–252.

Wojciszke, B., Bazinska, R., & Jaworski, M. (1998). On the dominance of moral categories in impression formation. *Personality and Social Psychology Bulletin*, *24*(12), 1251–1263.

Woo, B. M., & Hamlin, J. K. (2020). 11-month-olds expect others to behave consistently, but fail to evaluate those who do not. Manuscript in revision.

Woo, B. M., Liu, S., & Spelke, E. S. (2021). Open-minded, not naïve: Three-month-old infants encode objects as the goals of other people's reaches. In T. Fitch, C. Lamm, H. Leder, & K. Tessmar-Raible (Eds.), Proceedings of the 43rd Annual Meeting of the Cognitive Science Society. (pp. 514–520). Cognitive Science Society.

Woo, B. M., & Spelke, E. S. (2020a). How to help best: Infants' changing understanding of multistep actions informs their evaluations of helping. In S. Denison, M. Mack, Y. Xu, & B. C. Armstrong (Eds.), Proceedings of the 42nd Annual Conference Cognitive Science Society. (pp. 384–390). Cognitive Science Society.

Woo, B. M., & Spelke, E. S. (2020b). Toddlers' social evaluations of agents who act on false beliefs. *PsyArXiv*. 10.31234/osf.io/eczgp

Woo, B. M., Steckler, C. M., Le, D. T., & Hamlin, J. K. (2017). Social evaluation of intentional, truly accidental, and negligently accidental helpers and harmers by 10-month-old infants. *Cognition*, *168*, 154–163.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69*(1), 1–34.

Woodward, A. L. (2009). Infants' grasp of others' intentions. *Current Directions in Psychological Science*, *18*(1), 53–57.

Xu, F., & Carey, S. (1996). Infants' metaphysics: The case of numerical identity. *Cognitive Psychology*, *30*(2), 111–153.

Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences*, *104*(20), 8235–8240.

17 Around the World in 80 Milliseconds (or Less): Spontaneous Trait Inference across Cultures

Leonard S. Newman and Arthur D. Marsden III
Syracuse University

The graduate student held up the commuters by trying to squeeze through a subway turnstile with a desktop computer slung over his shoulder.

It would be reasonable to guess that the sentence opening this chapter was a stimulus sentence in one of the many studies of spontaneous trait inference conducted over the last few decades by Jim Uleman and his colleagues (see Uleman et al., 1996; Uleman et al., 2008; Uleman et al., 2012). The recall cue would undoubtedly be "thoughtless." In actuality, though, it describes a scene from the first author's attempt to finish collecting data for his dissertation in 1989.[1] He needed a sample of middle-schoolers, and although his plan was to recruit them all from a school on Long Island[2] (where he could borrow a family car to ferry his research materials back and forth), he came up short. His new source of subjects was a school in Brooklyn. He lived in Manhattan, and being a grad student, he did not have a car. And laptops, for all intents and purposes, did not exist back then.[3]

Research on spontaneous trait inference was past its infancy at that point, and the defining features of the phenomenon and its boundary conditions were already coming into focus (see Newman & Uleman, 1989; Uleman, 1987). Left unexplored in that first wave of research, however, was the issue of *why* people seemed to infer personality traits without intention and awareness. Perhaps that shouldn't be surprising. This, after all, was around the same time that Sorrentino and Higgins (1986, p. 3) bemoaned "the almost total exclusion of motivation in research on cognition." More than that, they claimed, "in some areas of social cognition, there is evidence of hostility toward anyone who would dare use the term 'motivation' in anything other than a pejorative manner" (p. 7).

The basic ideas underlying that dissertation (Newman, 1991) were certainly not earth-shattering: the trait inference process, it was assumed, should become relatively automatized the more people engage in it, and the frequency with which people will try to infer traits from behavior will vary as a function of how useful they find that information. Eliot Smith and his

DOI: 10.4324/9781003045687-20

students had already demonstrated that social judgments were made more efficiently with extensive practice (Smith & Lerner, 1986; Smith et al., 1988). But what about "usefulness"?

There was by that time abundant evidence that adolescents and adults were "lay dispositionists"; in other words, they were biased to overemphasize traits and other stable internal factors when describing people, explaining their behavior, and predicting what they were likely to do in the future (Gilbert, 1989; Jones, 1990; Ross & Nisbett, 1991). But there were some hints in the developmental literature that when it came to being hard line dispositionists, adolescents and adults took a back seat to their younger selves. Although children less than around 5 or 6 years of age seemed to have a limited understanding of stable personal dispositions and their implications (Rholes et al., 1990), some research suggested that by middle childhood they became orthodox trait theorists—true believers in the power of traits to determine behavior, lacking even whatever motivation and ability adults had to take contextual information into account when thinking about others (Barenboim, 1981; Leahy, 1976; Ross, 1981). Hence Newman's (1991) main hypothesis: because 9- to 11-year-olds so dearly valued information about people's personality traits, they would be engaging in the trait inference process even more regularly than adults, and as a result, spontaneous trait inference should be even more evident for them.[4]

And that more or less was what was found. There was, however, one wrinkle: that last batch of kids from Brooklyn wouldn't cooperate with the hypothesis. In contrast to the Long Island sample, there was no evidence at all that they were spontaneously inferring traits. Other measures also indicated that when it came to predicting people's behavior, they were much more focused on situational than dispositional information than any other group of participants in the study. What was going on? A few years prior, Joan Miller (1984) had documented what was then a surprising lack of dispositional thinking among research participants in Mysore, India. Similar findings began to emerge from other studies utilizing participants from collectivistic (versus individualistic) cultures (Bond, 1988; Triandis, 1989). Cousins (1989), for example, in a cross-cultural investigation of the self-concept utilizing the Twenty Statements test, found that "Japanese subjects listed fewer abstract, psychological attributes than did American subjects, referring more to social role and behavioral context" (p. 124). Indeed, as Newman (1991, p. 247) noted at the time, "The finding that non-American/European subjects use fewer stable dispositional qualities in their explanations and descriptions is among the most commonly reported social-psychological findings in the cross-cultural literature."

Dispositionism, then, appeared to be culture-bound. Individualists, like middle schoolers, prioritized information about people's personality traits when making sense of their behavior. People in collectivistic cultures, on the other hand (such as East Asians), rather than always assuming that people's behavior is diagnostic of their underlying personal characteristics, also

routinely attended to contextually-specific norms, social obligations, and role expectations as causes of behavior. Later research, in fact, directly showed that people from collectivistic cultures are more likely than people from individualistic cultures to endorse situationist theories of behavior (Norenzayan et al., 2002), and less likely to perceive individuals in social situations as being agentic (Menon et al., 1999).

The Brooklyn participants, as it turns out, were recruited from a school at which almost all of the students identified as "Hispanic" (primarily Puerto Rican)—and according to the extant literature (Hart et al., 1986; Marin & Triandis, 1985), they were more likely to be collectivistic in orientation than the (primarily Anglo) participants from Long Island. Hence, the post-hoc interpretation of the data:[5] collectivists (like that last batch of middle-schoolers in Newman's study) are less wedded than individualists to the idea that a person's behavior is diagnostic of his or her underlying personality traits. As a result, they will be less likely than individualists to engage in the trait inference process frequently and regularly enough to automatize the process. Overall, then, the data seemed to be consistent with the idea that people from collectivist cultures (essentially equivalent to what are also called "interdependent" cultures—Markus & Kitayama, 1991) do not spontaneously infer traits.

But at that point the database relevant to the issue was derived from a total of 19 elementary school students from a public school in Williamsburg, Brooklyn.[6] What have we learned since then? The following narrative review describes the relevant research. It is organized in terms of the three approaches which have been taken to clarify the nature of cross-cultural differences in spontaneous inference: direct cross-cultural comparisons, studies operationalizing cultural constructs as individual differences, and culture priming experiments.

One set of investigations, it should be noted, will not be discussed. A number of cross-cultural researchers have focused on variation in the correspondence bias (Choi & Nisbett, 1998; Knowles et al., 2001; Krull et al., 1999; Lieberman et al., 2005; Miyamoto & Kitayama, 2002). These well-designed studies contribute to a broader understanding of cultural variations in the person perception process more generally, especially the extent to which different phases of the process consume cognitive resources. But almost without exception these studies provide participants with an explicit processing goal ("infer the true attitude of the student"; "estimate the essay writer's actual position on this issue"; "figure out what the student's attitude is"). Spontaneous inferences, however, are those that occur "without intentions or instructions" (Uleman et al., 2012, p. 657). Thus, these studies fall outside the scope of this review.

Direct Cross-Cultural Comparisons

Unquestionably, the most face-valid way to assess cross-cultural differences in the spontaneity of trait inference is to compare groups of people raised (and

even better, still living) in different cultural settings. In the absence of any evidence for such differences, one could reasonably question whether the development of other methods to address the issue would be a worthwhile enterprise.

Zárate et al. (2001) ran two studies with participants who self-identified as either Anglo-Americans or Latinos. They hypothesized that members of the former (individualistic) group would be more likely than members of the second (collectivistic) group to engage in spontaneous trait inference. All participants were students at a university in Texas, and all were fluent speakers of English. Importantly, materials were pretested to ensure that the trait implications of the behaviors presented to participants were the same for Anglos and Latinos (e.g., when asked directly, members of both groups agreed that "She left a 25% tip for the waitress" implied "generous," and "She took the elevator up one flight" implied "lazy").

In Study 1, spontaneous trait inference was assessed via a lexical decision task. Participants viewed a series of sentences presented on a computer screen that they were asked to try to memorize. After some of the sentences, however, a letter string appeared, and participants were to decide as quickly as possible whether it was a word or not. Some of those letter strings were trait words, and they appeared after sentences that either did or did not describe behaviors that were diagnostic of those traits. Thus, traits were either "matched" or "mismatched" with the preceding sentence. If a trait is inferred spontaneously, it should be recognized as a word more quickly when it follows a trait-implying sentence and is a "match." Reaction times were quicker after trait-implying sentences than after control sentences for *both* groups, however, and the hypothesized Ethnic Group × Match interaction was not statistically significant ($p < .129$). But post-hoc tests revealed some support for the hypothesis, as there was reliable evidence for activation of trait concepts among Anglos but not Latinos.

In Zarate et al.'s second study, with a similar sample of participants, spontaneous trait inference was studied using a variant of the savings in relearning paradigm of Carlston and Skowronski (1994; Carlston et al., 1995). Participants examined a series of behavior descriptions, each paired with a photo of a person. Some were asked to form impressions of the people, and some simply to become familiar with the materials. After a brief distracter task, the researchers administered a measure designed to assess the extent to which participants had come to associate the people in the pictures with the personality traits that could have been inferred from their behaviors. Although evidence for that association was stronger among Anglos (regardless of their processing goals), that finding was unexpectedly restricted to multi-sentence behavior descriptions (versus single sentence descriptions). In addition, although the main dependent variable in Carlston et al.'s studies was the ease with which participants could subsequently explicitly pair the target persons with their associated traits (hence, "savings in relearning"), Zarate et al.'s measure was an explicit trait rating of the target person. As Na

and Kitayama (2011, p. 1026) noted, "This procedure makes it difficult to exclude the possibility that participants made the trait inferences on the basis of behaviors that they recalled during the testing phase (as opposed to the traits that had been inferred during the memorization phase)."

Na and Kitayama's (2011) study was the next one to tackle the issue of cross-cultural differences in spontaneous trait inference. Participants, all students at a university in Michigan, were classified as either "European American" (and hence, individualistic in cultural orientation) or "Asian American" (collectivistic), although it is unclear if any were foreign-born. As in Zarate et al.'s Study 2, participants were presented with trait-implying sentences paired with photos of individuals, but in this case they were asked to try to memorize the pairings. In a subsequent lexical decision task, the photos were used as the priming stimuli and the words to be identified included traits that could have been inferred from the behaviors described by the sentences. If a given trait had been associated with a person's picture, then the picture would facilitate lexical decision for that trait, as compared with lexical decision for some other (semantically unrelated) trait word. That pattern of results was in fact found for European Americans (i.e., lexical decisions were significantly faster for implied traits than for control traits), but not for Asian Americans. Na and Kitayama's second study yielded indirect converging evidence for the conclusion that individualists are more likely to spontaneously infer traits than collectivists: when pictures of people were followed by *antonyms* of the traits that their previous behavior suggested, an electrophysiological sign associated with processing of semantically inconsistent information (an ERP component called the N400) was detected among European Americans, but not Asian Americans.

Both of the investigations summarized previously involved college student participants recruited from a single university. The advantages of this research strategy (besides convenience) are obvious. Participants from the two groups will essentially be matched on age, current social/physical environment, and academic achievement. But it is not clear that students from collectivistic cultures enrolled in educational institutions in the United States are necessarily representative or prototypical members of their ethnic or national groups (Austin & Shen, 2016; Kim, 2011). In addition, when research participants are immigrants or temporary residents, their immersion in a new cultural context could have already led them to become acculturated to individualistic norms and attitudes (Murray et al., 2014; Schwartz et al., 2011).

Running laboratory experiments with participants separated by an ocean is obviously a more difficult option. But in a series of investigations, Lee, Shimizu, and Uleman and their colleagues conducted a series of studies on automatic and spontaneous inference processes with participants from North America and Japan, all tested in their home countries, and all presented with materials in either English or Japanese. Lee et al.'s (2015) first investigation focused specifically on the phenomenon of spontaneous trait *transference* (Skowronski et al., 1998). Spontaneous trait transference occurs when people

associate a trait inferred from a behavior not with the individual who enacted the behavior, but with someone else only incidentally related to the event. Lee et al. utilized a modified version of Todorov and Uleman's (2002, 2004) false-recognition paradigm. Participants were presented with a few dozen behavioral descriptions, each one paired with a photo of a person who was clearly identified as someone who was *describing* the behavior (not the person enacting it). In some cases, the description explicitly included the trait associated with the behavior (e.g., "She was rude and left the dinner party without thanking the hostess"). In others, the trait was not mentioned (e.g., "He phoned for help while the others just screamed"—calm; "She arrived to work ten minutes early every morning"—punctual). The key measure was the extent to which participants, when later seeing the photos again, falsely remembered that unmentioned (but implicitly suggested) trait words had been included in the behavioral descriptions that accompanied those photos. Although evidence for spontaneous trait inference was found for both Japanese and American participants, they were more pronounced in the latter than the former group.

Spontaneous trait transference does not require the inference that a person is characterized by a personality trait; all that is necessary is that a behavior be identified as belonging to a particular trait-related category (for more on the distinction, see Moskowitz, 1993; Newman & Uleman, 1993; and Todorov & Uleman, 2002). Thus, it is arguably a more low-level, simple associative process than spontaneous trait inference (cf. Wells et al., 2011). But Lee et al.'s findings suggested that cultural differences in the extent to which people extract trait-related meanings from behavior occur at the earliest stages of social information processing.

Shimizu et al. (2017) used another modified false-recognition procedure to more directly study spontaneous trait inference. In this study, the photos accompanying the behaviors were said to be of the people who actually engaged in those behaviors. Previous research (Todorov & Uleman, 2004) revealed that when behavioral descriptions are about the person in the photograph, as opposed to just reported by that person, the person-trait associations that are formed are stronger. Nonetheless, a cross-cultural difference was found again. In two studies, Shimizu et al. found evidence for spontaneous trait inference among both American and Japanese participants, but the Japanese participants made significantly fewer of them. Study 2 included a separate sample of Asian Americans too, although as in Na and Kitayama's (2011) report, little information was given about the criteria used to assign participants to that group. Interestingly, the results for Asian Americans were indistinguishable from those for European Americans. As Shimizu et al. noted, this result was inconsistent with Na and Kitayama's findings "that spontaneous trait inferences do not occur among Asian Americans" (p. 83).

When observing someone's behavior, there are other kinds of information one can gather and encode other than that person's dispositional characteristics. One could also infer something about the situation in which the

behavior takes place—specifically, the enduring qualities of the situation that would have led to someone engaging in that behavior. If "John gets an A on the test," one could infer that John is smart, but also perhaps that the test is easy. If "Will talks during the lecture," one could infer that Will is impolite, but maybe also that the class is boring (both examples from Ham & Vonk, 2003). Thus, in a number of investigations, the methods developed to assess spontaneous trait inferences have been used to study the extent to which people also make spontaneous *situation* inferences (Ham & Vonk, 2003; Lupfer et al., 1990). Spontaneous trait and situation inferences, it should be noted, are not mutually exclusive; indeed, the two could co-occur in response to an observed behavior (Todd et al., 2011).

In another study involving participants from Japan and North America (from Canada, in this case), Lee et al. (2017) used the savings-in-relearning paradigm to study both spontaneous trait inferences and spontaneous situation inferences. Replicating the findings of their other studies, Lee et al. found that although both North Americans and Japanese participants engaged in spontaneous trait inference, the Japanese participants did so to a lesser degree. There was no difference between the groups in the occurrence of spontaneous situation inferences, however. This finding was not inconsistent with the authors' hypotheses; in addition to predicting that spontaneous trait inference would be more evident among the North American participants than spontaneous situation inference, they predicted and found that Japanese participants would engage in both kinds of spontaneous inference, and do so equally. No predictions were explicitly made about between-group differences. Nonetheless, given that the theoretical background for the study included the observation that "people with high collectivistic/interdependent social orientations describe themselves in terms of their social roles and obligations to others, while being sensitive to contextual factors" (p. 629), it is unclear why one would not expect more spontaneous situation inferences for the Japanese participants.

In their most recent published investigation, Shimizu and Uleman (2021) again studied cross-cultural differences in spontaneous trait and situation inferences, this time with two false recognition paradigm experiments. The results overall replicated those of Lee et al. (2017). A novel finding in Study 2, made possible by the use of eye-tracking technology, was that spontaneous inferences were mediated by overt attention paid to persons relative to situations (as represented by pictures presented along with the behavior descriptions). More attention paid to persons was positively associated with the likelihood of spontaneous trait inference. Notably, both experiments included a group of Asian-American participants, all of whose first language was English. The results for the Asian-American participants were consistent with those for the Japanese participants, not the Canadian participants. This was a different outcome than the one reported by Shimizu et al. (2017), who found similar inferential biases for Asian-Americans and European Americans.

Interim Discussion

It is important to emphasize that in all of the studies described to this point, the behaviors presented to participants were pretested to ensure that they were not only meaningful to all participants, but also suggestive of the same traits. It would of course be inadvisable to present a sentence like "Elizabeth only watched television shows broadcast on PBS" (sophisticated) to research participants in Japan, just as it would be a mistake to expect American participants to understand that "Haruto could not lift the hangiri" implies that the subject of the sentence is "weak."[7]

Nonetheless, this strategy for testing hypotheses about the relationship between individualism, collectivism, and spontaneous trait inference is fraught with perils. Samples recruited from different countries (or even those based on national origin) can differ in any number of ways other than cultural orientation, and some of those differences—for example, socioeconomic status—could themselves be correlated with spontaneous trait inference (Varnum et al., 2012). Researchers generally control for age and gender when making comparisons, and often restrict participation to college students, but the studies reviewed above do not involve rigorously matched samples.

Recent research also casts doubt on an assumption that seems to underlie most cross-cultural comparisons (and not just those focusing on spontaneous trait inference): that if one recruits participants from a nation or group considered "individualistic" (e.g., North Americans, Western Europeans) and another from a locale typically considered to be "collectivistic" (e.g., China, Korea), then one's samples will in fact differ in terms of those worldviews. Talhelm (2020; see also Talhelm, et al., 2014) has demonstrated systematic within-nation variability in cultural orientation. For example, Chinese communities from the wheat-growing North are significantly more individualistic than those in the rice-growing South. Similarly, Kitayama et al. (2006) found that residents of the island of Hokkaido, on Japan's northern frontier, were significantly more individualistic than non-Hokkaido residents of Japan, and no different than European-Americans.

Ambiguity about the nature of the psychological differences between one's samples is even more salient in cases where the group representing collectivism consists of participants recruited from the same setting or cultural milieu as the individualistic group, differing only in terms of ethnicity—such as the "Asian-American" groups in the studies reviewed previously. (The same issue would apply to spontaneous inference studies in which individualist participants were European or American students studying at an Asian university, but the authors are not aware of any such studies). It has not been the norm to provide much information about these participants—where they were born, how long they had lived in the country in which they were residing when the study was conducted, whether they were bilingual, etc. The heterogeneity of these groups (both within a given study, and between labs) could account for the confusing pattern of findings described previously. Investigators using this recruitment

strategy sometimes report pronounced between-group differences, sometimes marginal ones, and sometimes none at all. To confuse matters even further, in some cases, Asian-American and Latinx participants are occasionally lumped with European Americans in an "American" group that is then compared to participants from another country (e.g., Lee et al., 2015). In order to facilitate future reviews of this literature, we urge investigators to provide more richly detailed information about their samples.

Individual Difference Approaches

Cultural differences in spontaneous inference are predicted because it is assumed that people from different cultures have different basic assumptions about themselves, other people, interpersonal relationships, and group processes. In other words, the assumption is that cultural differences are mediated by different cultural worldviews. What if one could measure that mediator directly? In other words, what if one could operationalize "culture" as an individual difference variable? A number of studies have taken this approach to testing hypotheses about the relationship between culture and spontaneous inference.

Triandis and colleagues make a distinction between the cultural level of analysis and the individual level, where individualism and collectivism are reflected, respectively, by the personality dimensions of idiocentrism and allocentrism (Triandis et al., 1995; Triandis et al., 1985). Newman (1993) administered a 29-item measure of idiocentrism developed by Triandis et al. (1988) to participants in two studies of spontaneous trait inference. Study 1 was based on the method first used by Winter and Uleman (1984) to study spontaneous trait inference: the encoding specificity paradigm. Participants were presented with trait-implying sentences for memorization (no mention was made of impression formation, and participants were not instructed to infer traits). After a distractor period, a cued-recall measure was administered. The dependent variable was how well personality traits (specifically, those associated with the behaviors described by the sentences) cued recall of the sentences. Idiocentrism was found to be positively and significantly associated with spontaneous trait inference, but for male participants only. The gender difference was unpredicted.

Study 2 used instead a recognition probe reaction time procedure adapted from McKoon and Ratcliff's (1986) research.[8] Participants were presented with a series of sentences by a computer. After each sentence a word appeared and participants' task was simply to indicate as quickly as possible (with a button press) whether that word had appeared in the preceding sentence. On key trials, (1) the word was a trait word (e.g., "clumsy") and (2) the sentence was either one implying a trait ("He stepped on his girlfriend's feet during the tango") or a scrambled version of that sentence, using as many of the same words as possible ("He and his girlfriend were on their feet and doing the tango"). If the trait is inferred from the behavior, the time to report

its absence from the preceding sentence should be longer (typically, by a few dozen milliseconds) than the time to do so after a semantically similar sentence that is unrelated to the trait. An important aspect of this procedure is that it builds in an incentive *not* to make trait inferences; doing so would only detract from performance on the participant's ostensible main task.

Idiocentrism was again correlated with spontaneous trait inference, and in this case, the relationship was not moderated by gender. Duff and Newman (1997), using the encoding specificity paradigm again, replicated the results of Newman's (1993) first study (with no moderation by gender), and in addition, found that idiocentrism was *negatively* correlated with spontaneous situation inferences. In other words, the higher participants' idiocentrism scores were, the better the cues "fearful" and "hardworking" were for triggering recall of "The mailman avoids the big dog at the house on the corner" and "On the designated day, the electrician is given a raise by his company" (as just two examples). The opposite was true for the situation cues "vicious" and "standard policy."

As already described, Na and Kitayama (2011), in their second study, presented to their participants pictures of people who had previously been described as engaging in trait-diagnostic behavior (in the context of an alleged memorization experiment). When antonyms of those traits appeared immediately after the pictures, an ERP component associated with the processing of semantically inconsistent information was detected among European Americans. This "incongruity effect" was not found for Asian-Americans, indicating that only European-Americans had spontaneously inferred traits. Na and Kitayama had also administered Singelis's (1994) measure of independent and interdependent self-construals. Independent self-construals (indexed by items such as "I am the same person at home that I am at school" and "I enjoy being unique and different from others in many respects") are characteristic of individualistic cultures, and interdependent self-construals ("My happiness depends on the happiness of those around me," "I respect people who are modest about themselves") are more characteristic of collectivistic cultures. The group difference in the incongruity effect found by Na and Kitayama was partially mediated by self-construal—it was stronger for those with higher independent relative to interdependent scores on the Singelis measure. As previously noted, though, the incongruity effect is a relatively indirect measure of spontaneous trait inference.

Finally, Shimizu and Uleman (2021), in their first study (in which they found cross-cultural differences in spontaneous inference with the false recognition paradigm), also had participants complete the Analysis-Holism Scale (Choi et al. 2007). East Asians have been found to make the holistic assumption that every element in the world is somehow interconnected, whereas Westerners tend to view the universe as composed of independent objects. That distinction is arguably consistent with the more general one between collectivism and individualism. The Analysis-Holism Scale is a 24-item measure that assesses analytic versus holistic thinking (e.g., "Everything

in the world is intertwined in a causal relationship"; "It is not possible to understand the parts without considering the whole picture"; "Choosing a middle ground in an argument should be avoided"). Shimizu and Uleman found that their Japanese participants were more holistic in their thinking styles than both the European-American and Asian-American participants, as expected. But there was no evidence that thinking style mediated spontaneous inferences.

The studies by Newman (1993) and Duff and Newman (1997) are the only ones showing a direct correlational relationship between an individual difference measure of cultural constructs and spontaneous trait inference. The findings were not consistently strong, however, and in one case the relationship was unexpectedly moderated by gender. Of even more concern is that the individual difference measure used (Triandis et al., 1988) was never subjected to much in the way of psychometric testing. Very little published data attest to its reliability and validity.

We are tempted, then, to issue a call for research using more established and psychometrically sound individual difference measures. However, although a number of new relevant measures have been developed over the last few decades (e.g., Shulruf et al., 2007), it is not clear which ones we would recommend. Oyserman et al. (2002), in their extensive review of individual-level measures of individualism/idiocentrism and collectivism/allocentrism, found that "a large number of instruments and operationalizations are in current usage" (p. 7), and they "did not find a single standard or most common measure" (p. 9). Oyserman et al. further expressed concern about "the different topics addressed in measurement instruments" (p. 9). As noted by Lee et al. (2017, p. 638), it is still the case that "little consensus has been established among researchers regarding the validity of self-report measures which are supposed to associate with the target mediators" in studies of the social-cognitive implications of different cultural syndromes.

Furthermore, Na et al. (2010) argue persuasively that one should not even necessarily expect that cultural constructs can be conceptualized as individual-level psychological traits. As they explain, "the group-level coherence of variables is independent of their individual-level coherence" (p. 6193). To illustrate, there are clearly systematic differences between societies in terms of whether they are organized as democracies or autocracies. Democratic societies, more than autocracies, will be characterized by economic liberty and safeguarding minority rights, but at the individual level, one might very well find no correlation between support for those two principles. Indeed, Na et al.'s own research revealed very low correlations between individual level measures of various aspects of cognitive style and social orientation typically found to distinguish between individualistic and collectivistic societies.

In sum, it is not clear that administering individual difference measures to members of a given society is the most fruitful approach to understanding cross-cultural differences in spontaneous inference. Results of such investigations

can reveal correlates of the specific measures used, but might be of limited value for addressing the broader issues that motivated the research.

Priming Culture

Validly and reliably measuring the extent to which people endorse individualistic or collectivistic attitudes, beliefs, and values is thus fraught with difficulties. An alternative is to randomly assign people to one cultural syndrome or the other. A number of investigators have developed methods of priming culture (Lechuga, 2008; Oyserman, & Lee, 2007; Trafimow et al., 1991), and one investigation took this approach to studying cultural differences in spontaneous inference.

Saribay et al. (2012) used a technique introduced by Gardner et al. (1999) for priming independent self-construals (as previously discussed, characteristic of individualistic cultures) and interdependent self-construals (characteristic of collectivistic cultures). In three studies, participants read a short story about a trip to the city, either focused on an individual (with frequent use of the pronouns "I," "me," and "mine") or a group (with frequent use of "we," "us," and "our"). Participants were asked to circle all of the pronouns in the story. Thus, independent self-construals were made more cognitively accessible to participants in the first condition, and interdependent self-construals were made more accessible to participants in the second condition. In Saribay et al.'s Experiment 1, spontaneous trait and situation inferences were measured with a lexical decision task. In Experiment 2, the false-recognition method was used to measure spontaneous trait inference. Participants in both studies made spontaneous trait inferences, and those in Experiment 1 also made spontaneous situation inferences. However, in neither case was self-construal a significant moderating variable.

In a third experiment, participants read a number of behavior descriptions (e.g., "Phil got every test question correct") and made explicit, intentional inferences about them ("How smart is Phil?" "How easy is the test?"). In this case, independent self-construal priming led to more trait inferences, but interdependent self-construal priming had no effect on inferences. In sum, the only culture priming effect was found for explicit, intentional inferences. As the authors noted, these findings "stand in apparent contrast to some of the literature on culture and person perception," in which "cultural differences are more profound when perceivers are cognitively busy" (p. 202). For example, although Zárate et al. (2001) found evidence for cultural differences in spontaneous trait inferences, they found no differences between Anglos and Latinos for intentional inferences. But Saribay et al. point out that "culture is a much broader construct" than self-construal, and that perhaps self-construal is not the element responsible for cultural differences in spontaneous inference.

Priming methods could potentially allow researchers to make the kinds of causal connections between culture and social inference processes that are

not possible with other methods. But a salient point raised by Saribay et al.'s research is that any program of research of this kind would need to provide converging evidence for its conclusions with experiments utilizing more than one culture priming procedure. Otherwise, it would be difficult to rule out the possibility that a study's findings could be attributed to the theory-irrelevant idiosyncrasies of any one specific method.

Spontaneous Situation Inferences: We Have a Situation Here

A number of the investigations described in this review, along with testing the hypothesis that spontaneous trait inference is more associated with individualism than collectivism, also tested a second, complementary hypothesis: are spontaneous inferences about the situation more associated with collectivism than individualism? It is generally assumed that trait inferences are made in the service of being able to anticipate and predict other people's behavior, and in that way, to possibly have more control over one's future interactions (McCarthy & Skowronski, 2011; Pittman, & DAgostino, 1985). But people from collectivistic cultures would undoubtedly have the same needs. How might they satisfy them? Given their emphasis on contextual influences, interconnectedness, and more generally, holistic thinking, situational information might be what is perceived by them to be most useful. And over time, situation inferences could possibly become relatively automatized—either instead of or in addition to trait inferences.

There is as yet little evidence for this conjecture, however. Duff and Newman (1997) employed a measure of idiocentrism in their studies, but not allocentrism. And although Lee et al. (2017) and Shimizu and Uleman (2021) found that Japanese participants, unlike their European-American counterparts, did not make more trait than situational inferences, they also did not make more spontaneous situation inferences than trait inferences (nor did they make more situation inferences than North Americans). Unfortunately, even if there were more studies of this kind to cite, we are not sure how useful the results would be for the development of a theoretical understanding of cross-cultural differences in spontaneous social inferences. The issue is the almost totally unbounded nature of "situation cues" as a category. Sometimes the situation cues used by researchers in studies of spontaneous inference refer to a physical characteristic of some object involved in a behavior ("Eric lifts the boulder"—"Light"). In other cases, the situation is defined by an abstract quality ("The secretary solves the mystery halfway through the book"—"Simple plot"). "Situations" are sometimes also operationalized as event descriptions ("The engineer picks up the papers from the floor"—"Dropped them"), and in other cases, even as the personal characteristics of people other than the focal actor ("Paul helps the old lady cross the street"—"Needy"). In extreme cases, principled distinctions between cue types used in experiments to represent dispositional and situational inferences arguably disappear entirely, as seen with "Ed loses the game

of chess from his opponent," where the trait cue is "Bad chess player" and the situation cue is "Good chess player" (all examples from Duff & Newman, 1997; Ham & Vonk, 2003; Lupfer et al., 1990).

These problems apply to all studies of spontaneous situation inferences, not just those investigating cross-cultural differences. But we emphasize them here because of the role they could play in complicating efforts at clarifying social-cognitive similarities and differences between cultures. We urge future researchers to think more systematically about the nature of spontaneous situation inferences, perhaps guided by formal taxonomies of psychological situations (e.g., Parrigon et al., 2017; cf. Reiss, 2018). Alternatively, the focus could narrow to situational cues involving social norms, constraints, and pressures. Those are the kinds of contextual factors that seem most relevant to theoretical analyses of collectivism as a psychological syndrome. Indeed, we are not convinced that any perspective on cross-cultural differences in social inference would lead to the prediction that people from collectivistic cultures are more likely than those from individualistic ones to take into account the difficulty of tests, the weather, or the weight of objects (among other situational variables) when interpreting others' behaviors.

A Plot Twist

Cross-cultural studies of spontaneous trait inference obviously are not required to be framed in terms of the distinction between individualism and collectivism. Conceivably, one could also (as just a few examples) compare research participants recruited from "cultures of honor" to others recruited from cultures of dignity (Nisbett & Cohen, 1996), people from societies characterized with high levels of relational mobility to those from societies with lower levels (Schug et al., 2010), or people from liberal, democratic societies with individuals living in autocracies. Cultures differ socially and psychologically in a wide variety of ways (Oyserman, 2006; Schwartz, 1994). But the hypothesis that spontaneous trait inference is positively associated with individualism and negatively with collectivism has been repeatedly tested for two reasons. First of all, it seems to many to be intuitively compelling. In individualistic cultures, after all, people are said to be motivated by their personal goals; individuality is both valued and assumed in both oneself and others; and people do the things they do, it is believed, because of their unique bundle of personal characteristics. Given all that, it seems reasonable to predict that individualists will develop a tendency to quickly and easily extract trait meanings from behavior. In collectivistic cultures, on the other hand, people define themselves and others in terms of their social roles and relationships; they do the things they do, it is believed, because that is what is expected or required of them in the situations they find themselves in; and no one would necessarily assume that a person's behavior would be highly consistent across situations.

In addition, as previously discussed, studies of how people explicitly describe other people and themselves (e.g., Cousins, 1989; Rhee et al., 1995;

Shweder & Bourne, 1982) and explain behavior (Choi et al., 1999; Miller, 1984) have yielded evidence consistent with the idea that a focus on broad, general personality traits relative to situational factors is more characteristic of people in individualistic cultures than those in collectivistic cultures. So why shouldn't that difference also manifest itself in how individuals spontaneously construe other people and their behavior?

Recent work by Liu et al. (2019), however, might give rise to some rethinking of the relationship between culture and spontaneous trait inference. Liu et al. observe that although collectivistic cultures are characterized by tight interpersonal connections and fuzzy boundaries between self and other, tight social relationships can come with a cost. Surface level intragroup harmony can actually mask a great deal of within-group competition, and people in collectivistic cultures "recognize that others in the group might constrain them and impinge upon their interests" (p. 14539). The result, Liu et al. suggest, is elevated levels of social vigilance, which they define as a social cognitive tendency to anticipate threatening behaviors from members of one's ingroup. In studies with participants from the United States and China, Liu et al. found that Chinese participants were more mindful than American participants of their peers' potential unethical intentions in hypothetical within-group competitions. Another finding was that Chinese participants were more likely than Americans to suspect that a peer's friendly behavior actually masked sinister intentions.

It should be apparent that Liu et al.'s perspective on collectivistic cultures is not consistent with the assumption that collectivists do not find stable dispositions useful for interpreting and predicting behavior, and does not lead to the prediction that they will be less likely than individualists to automatize the trait inference process as a result of regularly engaging in it. Instead, it arguably leads to the prediction that collectivism will be *positively* associated with a certain kind of spontaneous trait inference: the kind involving ingroup members' negative characteristics. Support for that hypothesis would not only provide converging evidence for Liu et al.'s characterization of collectivist cultures: it could also help explain why the evidence for the idea that spontaneous trait inference is more prevalent among individualists than collectivists is not as compelling as one might expect it to be after 30 years.

Summary and Conclusions

What, then, have we learned since the publication of the earliest data suggesting cross-cultural differences in spontaneous trait inference? One thing can be said with certainty: the hypothesis that people in collectivistic cultures do not infer traits spontaneously has been strongly disconfirmed. That much is evident from the results of the studies reviewed in this chapter. It is also apparent in the growing number of investigations not concerned with cross-cultural differences at all, but simply conducted with research participants in both China (e.g., Tong & Chiu, 2012; Wang et al., 2015;

Wang et al., 2018; Wang et al., 2016; Yan et al., 2012; Yang & Wang, 2016; Zhang & Wang, 2013, 2018) and Japan (Shimizu, 2012).[9]

But does the accumulated research suggest that spontaneous trait inferences are more likely in individualistic than collectivistic cultures? Based on the studies reviewed here, our best answer to that question would be a measured "yes." (And no published investigation has ever found the opposite). For reasons already discussed, though, we would not be able to address a similar question about spontaneous situation inferences. Given the lack of constraints on the definition of "situation" in this body of research, the psychological process corresponding to the phrase "spontaneous situation inference" is arguably too underspecified. Similarly, our overall conclusion about cultural differences in spontaneous inference is not informed by the individual differences studies reviewed previously. Such studies can provide evidence relevant to the validity of the specific measures used, and depending on the particular constructs those measures operationalize (self-construal, thinking style, etc.), the findings could be of interest and value to any number of different areas of research. But within-culture variability on those measures will not necessarily shed light on the consequences of between-culture variability in the beliefs, attitudes, or behavioral regularities they assess.

At the very least, existing investigations of culture and spontaneous trait inferences add up to a compelling set of "proof of concept" demonstrations. They indicate that there is in fact meaningful cross-cultural variance in spontaneous social inference to be accounted for. But recent research hints at the complexity of the cultural differences in play (Liu et al., 2019); the picture that ultimately emerges might be more nuanced than previously suspected. In addition, further progress in this area will require moving beyond recruiting convenience samples of participants to represent individualism and collectivism (with the only requirement being that a plausible case can be made for the face-validity of the distinction between the groups). Interpretive problems associated with the use of unrepresentative and systematically biased samples are of course not restricted to the research literature on spontaneous trait inference (see Henry, 2008; McCredie & Morey, 2019). Furthermore, the problems that arise when recruiting participants to represent different cultural syndromes are no different in studies of spontaneous trait inference than they are in other areas of cross-cultural research (Nivette, 2011; Stigler & Miller, 1993). Nonetheless, the stated goals of the studies reviewed here were to recruit representatives of individualistic and collectivistic cultures, compare the extent to which they engaged in spontaneous trait inference, and from those data, generalize about differences between individualism and collectivism. It is now abundantly clear, however, that Japanese college students, American students identifying as "Asian American," students recently arrived from Korea and enrolled in North American universities, and people from wheat-growing regions of China are not interchangeable when it comes to operationalizing cultural worldviews. Adequately testing hypotheses about social-cognitive differences

associated with individualism and collectivism requires detailed knowledge of the populations from which one is sampling—including those assumed to be more individualistic in outlook.

Even with more informed sampling procedures, however, there is no guarantee that differences between a group of individualistic participants and another collectivistic one will not derive from other theory-irrelevant differences between the two. A truly major advance in our understanding of cross-cultural differences in spontaneous inference might require the kinds of large-scale, multi-country, multi-investigator studies that have been carried out to address global variability in other aspects of behavior (see He et al., 2015; Ito et al., 2019; Mõttus et al., 2012; Walter et al., 2020). A project of that kind would be invaluable. Shimizu et al. (2017) suggested that a deeper understanding of culture and cultural differences to a great extent depends on shedding light on "the automatic procedures for imbuing meaning into our own and others' behavior," and further noted that "Particularly revealing are those differences in performance that individuals cannot control, even when they wish to" (p. 80). A more complete picture of cross-cultural differences in spontaneous inference could significantly contribute to that understanding—and to deepening our understanding of the basic social-cognitive processes involved in impression formation.

Notes

1 Okay, it was a Mac Plus, not such a big machine. But still.
2 It was the school where he had been a student 20 years earlier. His second-grade teacher, Miss Tacke, was still there, still teaching the second grade, and unnervingly, looking completely unchanged.
3 Apple's first "laptop," the Macintosh Portable, was released that year. It weighed 16 pounds and cost $6,500.
4 Alternative origin story for this study: the first author was at the time working with both Jim Uleman and Diane Ruble. Diane was studying the development of social attribution in children. Newman put both of their programs of research into a blender, and this is what came out.
5 Would an editor let you get away with this today? Let's see a show of hands … right, I didn't think so either.
6 Just a few blocks away from the location of the first author's grandfather's long-gone Kosher butcher shop.
7 A *hangiri* is a bamboo basket used for mixing and cooling sushi rice.
8 This paper and Newman (1991) were the first publications to feature the recognition probe reaction time procedure as a way to study spontaneous social inferences. Where did Newman get the idea for this approach? From one of Jim Uleman's grant proposals. For the most detailed treatment of the ins and outs of this method for studying spontaneous trait inference, see Uleman, Hon, Roman, and Moskowitz (1996).
9 Wang et al. (2018), Zhang and Fang (2016), and Zhang and Wang (2013) also provided converging evidence for the developmental trends in spontaneous trait inference found by Newman (1991).

References

Austin, L., & Shen, L. (2016). Factors influencing Chinese student's decisions to study in the United States. *Journal of International Students*, 6, 722–739. 10.32674/jis.v6i3.353

Barenboim, C. (1981). The development of person perception in childhood and adolescence: From behavioral comparisons to psychological constructs to psychological comparisons. *Child Development*, *52*, 129–144. 10.2307/1129222

Bond, M. H. (Ed.). (1988). *The cross-cultural challenge to social psychology*. Beverly Hills, CA: Sage.

Carlston, D. E., & Skowronski, J. J. (1994). Savings in the relearning of trait information as evidence for spontaneous inference generation. *Journal of Personality and Social Psychology*, 66, 840–856. 10.1037/0022-3514.66.5.840

Carlston, D. E., Skowronski, J. J., & Sparks, C. (1995). Savings in relearning II: On the formation of behavior-based trait associations and inferences. *Journal of Personality and Social Psychology*, 69, 420–436. 10.1037/0022-3514.69.3.429

Choi, I., Koo, M., & Jong An, C. (2007). Individual differences in analytic versus holistic thinking. *Personality and Social Psychology Bulletin*, *33*(5), 691–705. 10.1177/0146167206298568.

Choi, I., & Nisbett, R. E. (1998). Situational salience and cultural differences in the correspondence bias and actor-observer bias. *Personality and Social Psychology Bulletin*, *24*, 949–960. 10.1177/0146167298249003

Choi, I., Nisbett, R. E., & Norenzayan, A. (1999). Causal attribution across cultures: Variation and universality. *Psychological Bulletin*, *125*, 47–63. 10.1037/0033-2909.125.1.47

Cousins, S. D. (1989). Culture and self-perception in Japan and the United States. *Journal of Personality and Social Psychology*, *56*, 124–131. 10.1037/0022-3514.56.1.124

Duff, K. J., & Newman, L. S. (1997). Individual differences in the spontaneous construal of behavior: Idiocentrism and the automatization of the trait inference process. *Social Cognition*, *15*, 217–241. 10.1521/soco.1997.15.3.217

Gardner, W. L., Gabriel, S., & Lee, A. Y. (1999). "I" value freedom, but "we" value relationships: Self-construal priming mirrors cultural differences in judgment. *Psychological Science*, *10*, 321–326. 10.1111/1467-9280.00162

Gilbert, D. T. (1989). Thinking lightly about others: Automatic components of the social inference process. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 189–211). New York: Guilford Publications.

Ham, J., & Vonk, R. (2003). Smart and easy: Co-occurring activation of spontaneous trait inferences and spontaneous situational inferences. *Journal of Experimental Social Psychology*, *39*, 434–447. 10.1016/S0022-1031(03)00033-7

Hart, D., Lucca-Irizarry, N., & Damon, W. (1986). The development of self-understanding in Puerto Rico and the United States. *Journal of Early Adolescence*, 6, 293–304. 10.1177/0272431686063007

He, J., van de Vijver, F. J. R., Espinosa, A. D., Abubakar, A., Dimitrova, R., Adams, B. G., Aydinli, A., Atitsogbe, K., Alonso-Arbiol, I., Bobowik, M., Fischer, R., Jordanov, V., Mastrotheodoros, S., Neto, F., Ponizovsky, Y. J., Reb, J., Sim, S., Sovet, L., Stefenel, D., … Villieux, A. (2015). Socially desirable responding: Enhancement and denial in 20 countries. *Cross-Cultural Research: The Journal of Comparative Social Science*, *49*, 227–249. 10.1177/1069397114552781

Henry, P. J. (2008). College sophomores in the laboratory redux: Influences of a narrow data base on social psychology's view of the nature of prejudice. *Psychological Inquiry*, *19*, 49–71. 10.1080/10478400802049936

Ito, H., Barzykowski, K., Grzesik, M., Gülgöz, S., Gürdere, C., Janssen, S. M. J., Khor, J., Rowthorn, H., Wade, K. A., Luna, K., Albuquerque, P. B., Kumar, D., Singh, A. D., Cecconello, W. W., Cadavid, S., Laird, N. C., Baldassari, M. J., Lindsay, D. S., & Mori, K. (2019). Eyewitness memory distortion following co-witness discussion: A replication of Garry, French, Kinzett, and Mori (2008) in ten countries. *Journal of Applied Research in Memory and Cognition*, *8*, 68–77. 10.1016/j.jarmac.2018.09.004

Jones, E. E. (1990). *Interpersonal perception*. New York: W. H. Freeman.

Kim, H. (2011). Korean Students in the United States. *International Higher Education*, (64). 10.6017/ihe.2011.64.8557

Kitayama, S., Ishii, K., Imada, T., Takemura, K., & Ramaswamy, J. (2006). Voluntary settlement and the spirit of independence: Evidence from Japan's "northern frontier." *Journal of Personality and Social Psychology*, *91*, 369–384. 10.1037/0022-3514.91.3.369

Knowles, E. D., Morris, M. W., Chiu, C., & Hong, Y. (2001). Culture and the process of person perception: Evidence for automaticity among East Asians in correcting for situational influences on behavior. *Personality and Social Psychology Bulletin*, *27*, 1344–1356. 10.1177/01461672012710010

Krull, D. S., Loy, M. H.-M., Lin, J., Wang, C.-F., Chen, S., & Zhao, X. (1999). The fundamental fundamental attribution error: Correspondence bias in individualist and collectivist cultures. *Personality and Social Psychology Bulletin*, *25*, 1208–1219. 10.1177/0146167299258003

Leahy, R. L. (1976). Developmental trends in qualified inferences and descriptions of self and others. *Developmental Psychology*, *12*, 546–547. 10.1037/0012-1649.12.6.546

Lechuga, J. (2008). Is acculturation a dynamic construct?: The influence of method of priming culture on acculturation. *Hispanic Journal of Behavioral Sciences*, *30*, 324–339. 10.1177/0739986308319570

Lee, H., Shimizu, Y., Masuda, T., & Uleman, J. S. (2017). Cultural differences in spontaneous trait and situation inferences. *Journal of Cross-Cultural Psychology*, *48*(5), 627–643. 10.1177/0022022117699279

Lee, H., Shimizu, Y., & Uleman, J. S. (2015). Cultural differences in the automaticity of elemental impression formation. *Social Cognition*, *33*(1), 1–19. 10.1521/soco.2015.33.1.1

Lieberman, M. D., Jarcho, J. M., & Obayashi, J. (2005). Attributional inference across cultures: Similar automatic attributions and different controlled corrections. *Personality and Social Psychology Bulletin*, *31*, 889–901. 10.1177/0146167204274094

Liu, S. S., Morris, M. W., Talhelm, T., & Yang, Q. (2019). Ingroup vigilance in collectivistic cultures. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(29), 14538–14546. 10.1073/pnas.1817588116

Lupfer, M. B., Clark, L. F., & Hutcherson, H. W. (1990). Impact of context on spontaneous trait and situational attributions. *Journal of Personality and Social Psychology*, *58*, 239–249. 10.1037/0022-3514.58.2.239

Marin, G., & Triandis, H. C. (1985). Allocentrism as an important characteristic of the behavior of Latin Americans and Hispanics. In R. Diaz-Guerrero (Ed.),

Cross-cultural and national studies in social psychology (pp. 69–80). New York: Elsevier Science Pub. Co.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, *98*, 224–253. doi: 10.1037/0033-295x.98.2.224.

McCarthy, R. J., & Skowronski, J. J. (2011). What will Phil do next? Spontaneously inferred traits influence predictions of behavior. *Journal of Experimental Social Psychology*, *47*, 321–332. 10.1016/j.jesp.2010.10.015

McCredie, M. N., & Morey, L. C. (2019). Who are the Turkers? A characterization of MTurk workers using the personality assessment inventory. *Assessment*, *26*, 759–766. 10.1177/1073191118760709

McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 82–91. 10.1037/0278-7393.12.1.82

Menon, T., Morris, M. W., Chiu, C., & Hong, Y. (1999). Culture and the construal of agency: Attribution to individual versus group dispositions. *Journal of Personality and Social Psychology*, *76*(5), 701–717. 10.1037/0022-3514.76.5.701

Miyamoto, Y., & Kitayama, S. (2002). Cultural variation in correspondence bias: The critical role of attitude diagnosticity of socially constrained behavior. *Journal of Personality and Social Psychology*, *83*, 1239–1248. 10.1037/0022-3514.83.5.1239

Miller, J. G. (1984). Culture and the development of everyday social explanation. *Journal of Personality and Social Psychology*, *46*, 961–978. 10.1037/0022-3514.46.5.961

Moskowitz, G. B. (1993). Person organization with a memory set: Are spontaneous trait inferences personality characterizations or behaviour labels? *European Journal of Personality*, *7*, 195–208. 10.1002/per.2410070305

Mõttus, R., Allik, J., Realo, A., Rossier, J., Zecca, G., Ah-Kion, J., Amoussou-Yéyé, D., Bäckström, M., Barkauskiene, R., Barry, O., Bhowon, U., Björklund, F., Bochaver, A., Bochaver, K., de Bruin, G., Cabrera, H. F., Chen, S. X., Church, A. T., Cissé, D. D., … Johnson, W. (2012). The effect of response style on self-reported conscientiousness across 20 countries. *Personality and Social Psychology Bulletin*, *38*(11), 1423–1436. 10.1177/0146167212451275

Murray, K. E., Klonoff, E. A., Garcini, L. M., Ullman, J. B., Wall, T. L., & Myers, M. G. (2014). Assessing acculturation over time: A four-year prospective study of Asian American young adults. *Asian American Journal of Psychology*, *5*, 252–261. 10.1037/a0034908

Na, J., Grossmann, I., Varnum, M. E., Kitayama, S., Gonzalez, R., & Nisbett, R. E. (2010). Cultural differences are not always reducible to individual differences. *Proceedings of the National Academy of Sciences*, *107*, 6192–6197. 10.1073/pnas.1001911107

Na, J., & Kitayama, S. (2011). Spontaneous trait inference is culture-specific: Behavioral and neural evidence. *Psychological Science*, *22*, 1025–1032. 10.1177/0956797611414727

Newman, L. S. (1991). Why are traits inferred spontaneously? A developmental approach. *Social Cognition*, *9*, 221–253. https://doi-org.libezproxy2.syr.edu/10.1521/soco.1991.9.3.221

Newman, L. S. (1993). How individualists interpret behavior: Idiocentrism and spontaneous trait inference. *Social Cognition*, *11*, 243–269. 10.1521/soco.1993.11.2.243

Newman, L. S., & Uleman, J. S. (1989). Spontaneous trait inference. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 155–188). New York: The Guilford Press.

Newman, L. S., & Uleman, J. S. (1993). When are you what you did? Behavior identification and dispositional inference in person memory, attribution, and social judgment. *Personality and Social Psychology Bulletin*, *19*, 513–525. 0.1177/0146167293195004

Nisbett, R. E., & Cohen, D. (1996). *Culture of Honor: The psychology of violence in the South*. Westview Press.

Nivette, A. E. (2011). Cross-national predictors of crime: A meta-Analysis. *Homicide Studies: An Interdisciplinary & International Journal*, *15*, 103–131. 10.1177/1088767911406397

Norenzayan, A., Choi, I., & Nisbett, R. E. (2002). Cultural similarities and differences in social inference: Evidence from behavioral predictions and lay theories of behavior. *Personality and Social Psychology Bulletin*, *28*, 109–120. 10.1177/0146167202281010

Oyserman, D. (2006). High power, low power, and equality: Culture beyond individualism and collectivism. *Journal of Consumer Psychology*, *16*, 352–356. 10.1207/s15327663jcp1604_6

Oyserman, D., Coon, H. M., & Kemmelmeier, M. (2002). Rethinking individualism and collectivism: Evaluation of theoretical assumptions and meta-analyses. *Psychological Bulletin*, *128*, 3–72. 10.1037//0033-2909.128.1.3

Oyserman, D., & Lee, S. W.-S. (2007). Priming "culture": Culture as situated cognition. In S. Kitayama & D. Cohen (Eds.), *Handbook of cultural psychology* (pp. 255–279). The Guilford Press.

Parrigon, S., Woo, S. E., Tay, L., & Wang, T. (2017). CAPTION-ing the situation: A lexically-derived taxonomy of psychological situation characteristics. *Journal of Personality and Social Psychology*, *112*, 642–681. 10.1037/pspp0000111.supp

Pittman, T. S., & DAgostino, P. R. (1985). Motivation and attribution: The effects of control deprivation on subsequent information processing. In J. H. Harvey & G. Weary (Eds.), *Attribution: Basic and applied issues* (pp. 117–142). San Diego, CA: Academic Press.

Reis, H. T. (2018). Why bottom-up taxonomies are unlikely to satisfy the quest for a definitive taxonomy of situations. *Journal of Personality and Social Psychology*, *114*, 489–492. 10.1037/pspp0000158

Rhee, E., Uleman, J. S., Lee, H. K., & Roman, R. J. (1995). Spontaneous self-descriptions and ethnic identities in individualistic and collectivistic cultures. *Journal of Personality and Social Psychology*, *69*, 142–152. 10.1037/0022-3514.69.1.142

Rholes, W. S., Newman, L. S., & Ruble, D. N. (1990). Understanding self and other: Developmental and motivational aspects of perceiving people in terms of invariant dispositions. In E. T. Higgins & R. M. Sorrentino (Eds.), *Handbook of motivation and cognition* (Vol. 2, pp. 369–407). New York: Guilford Publications.

Ross, L. (1981). The "intuitive scientist" formulation and its developmental implications. In J. H. Flavell & L. Ross (Eds.), *Social cognitive development: Frontiers and possible futures* (pp. 1–42). New York: Cambridge University Press.

Ross, L., & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology*. New York: McGraw-Hill.

Saribay, S. A., Rim, S., & Uleman, J. S. (2012). Primed self-construal, culture, and stages of impression formation. *Social Psychology*, *43*, 196–204. 10.1027/1864-9335/a000120

Schug, J., Yuki, M., & Maddux, W. (2010). Relational mobility explains between- and within-culture differences in self-disclosure to close friends. *Psychological Science*, *21*, 1471–1478. 10.1177/0956797610382786

Schwartz, S. H. (1994). Beyond individualism/collectivism: New cultural dimensions of values. In U. Kim, H. C. Triandis, Ç. Kâğitçibaşi, S.-C. Choi, & G. Yoon (Eds.), *Cross-cultural research and methodology series, Vol. 18. Individualism and collectivism: Theory, method, and applications* (pp. 85–119). Sage Publications.

Schwartz, S. J., Weisskirch, R. S., Zamboanga, B. L., Castillo, L. G., Ham, L. S., Huynh, Q.-L., Park, I. J. K., Donovan, R., Kim, S. Y., Vernon, M., Davis, M. J., & Cano, M. A. (2011). Dimensions of acculturation: Associations with health risk behaviors among college students from immigrant families. *Journal of Counseling Psychology*, *58*, 27–41. 10.1037/a0021356

Shimizu, Y. (2012). Spontaneous trait inferences among Japanese children and adults: A developmental approach. *Asian Journal of Social Psychology*, *15*(2), 112–121. 10.1111/j.1467-839X.2012.01370.x

Shimizu, Y., Lee, H., & Uleman, J. S. (2017). Culture as automatic processes for making meaning: Spontaneous trait inferences. *Journal of Experimental Social Psychology*, *69*, 79–85. 10.1016/j.jesp.2016.08.003

Shimizu, Y., & Uleman, J. S. (2021). Attention allocation is a possible mediator of cultural variations in spontaneous trait and situation inferences: Eye-tracking evidence. *Journal of Experimental Social Psychology*, *94*, 1–11. 10.1016/j.jesp.2021.104115

Shulruf, B., Hattie, J., & Dixon, R. (2007). Development of a new measurement tool for individualism and collectivism. *Journal of Psychoeducational Assessment*, *25*, 385–401. 10.1177/0734282906298992

Shweder, R. A., & Bourne, E. J. (1982). Does the concept of the person vary cross-culturally? In A. J. Marsella & G. M. White (Eds.), *Cultural conceptions of mental health and therapy* (pp. 97–137). New York: Reidel.

Singelis, T. M. (1994). The measurement of independent and interdependent self-construals. *Personality and Social Psychology Bulletin*, *20*, 580–591. 10.1177/0146167294205014

Skowronski, J. J., Carlston, D. E., Mae, L., & Crawford, M. T. (1998). Spontaneous trait transference: Communicators take on the qualities they describe in others. *Journal of Personality and Social Psychology*, *74*, 837–848. 10.1037/0022-3514.74.4.837

Smith, E. R., Branscombe, N. R., & Bormann, C. (1988). Generality of the effects of practice on social judgment tasks. *Journal of Personality and Social Psychology*, *54*(3), 385–395. 10.1037/0022-3514.54.3.385

Smith, E. R., & Lerner, M. (1986). Development of automatism of social judgments. *Journal of Personality and Social Psychology*, *50*, 246–259. 10.1037/0022-3514.50.2.246

Sorrentino, R. M., & Higgins, E. T. (1986). Motivation and cognition: Warming up to synergism. In R. M. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition* (Vol. 1, pp. 3–19). New York: Guilford.

Stigler, J. W., & Miller, K. F. (1993). A good match is hard to find: Comment on Mayer, Tajika, and Stanley (1991). *Journal of Educational Psychology*, *85*, 554–559. 10.1037/0022-0663.85.3.554

Talhelm, T. (2020). Emerging evidence of cultural differences linked to rice versus wheat agriculture. *Current Opinion in Psychology*, *32*, 81–88. 10.1016/j.copsyc.2019.06.031

Talhelm, T., Zhang, X., Oishi, S., Shimin, C., Duan, D., Lan, X., & Kitayama, S. (2014). Large-scale psychological differences within China explained by rice versus wheat agriculture. *Science*, *344*(6184), 603–608. 10.1126/science.1246850

Todd, A. R., Molden, D. C., Ham, J., & Vonk, R. (2011). The automatic and co-occurring activation of multiple social inferences. *Journal of Experimental Social Psychology*, *47*, 37–49. 10.1016/j.jesp.2010.08.006

Todorov, A., & Uleman, J. S. (2002). Spontaneous trait inferences are bound to actors' faces: Evidence from a false recognition paradigm. *Journal of Personality and Social Psychology*, *83*, 1051–1065. 10.1037/0022-3514.83.5.1051

Todorov, A., & Uleman, J. S. (2004). The person reference process in spontaneous trait inferences. *Journal of Personality and Social Psychology*, *87*, 482–493. 10.1037/0022-3514.87.4.482

Tong, Y.-Y., & Chiu, C.-Y. (2012). Why do people think culturally when making decisions? Theory and evidence. In R. W. Proctor, S. Y. Nof, & Y. Yih (Eds.), *Cultural factors in systems design: Decision making and action*. (pp. 53–64). CRC Press.

Trafimow, D., Triandis, H. C., & Goto, S. G. (1991). Some tests of the distinction between the private self and the collective self. *Journal of Personality and Social Psychology*, *60*, 649–655. 10.1037/0022-3514.60.5.649

Triandis, H. C. (1989). The self and social behavior in differing cultural contexts. *Psychological Review*, *96*, 506–520. 10.1037/0033-295X.96.3.506

Triandis, H. C., Bontempo, R., Villareal, M. J., Asai, M., & Lucca, N. (1988). Individualism and collectivism: Cross-cultural perspectives on self-ingroup relationships. *Journal of Personality and Social Psychology*, *54*, 323–338. 10.1037/0022-3514.54.2.323

Triandis, H. C., Chan, D. K.-S., Bhawuk, D. P. S., Iwao, S., & Sinha, J. B. P. (1995). Multimethod probes of allocentrism and idiocentrism. *International Journal of Psychology*, *30*, 461–480. 10.1080/00207599508246580

Triandis, H. C., Leung, K., Villareal, M. J., & Clack, F. L. (1985). Allocentric versus idiocentric tendencies: Convergent and discriminant validation. *Journal of Research in Personality*, *19*, 395–415. 10.1016/0092-6566(85)90008-X

Uleman, J. S. (1987). Consciousness and control: The case of spontaneous trait inferences. *Personality and Social Psychology Bulletin*, *13*, 337–354. 10.1177/0146167287133004

Uleman, J. S., Hon, A., Roman, R. J., & Moskowitz, G. B. (1996). On-line evidence for spontaneous trait inferences at encoding. *Personality and Social Psychology Bulletin*, *22*, 377–394. 10.1177/0146167296224005

Uleman, J. S., Newman, L. S., & Moskowitz, G. B. (1996). People as flexible interpreters: Evidence and issues from spontaneous trait inference. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 28, pp. 211–279). Academic Press. 10.1016/S0065-2601(08)60239-7

Uleman, J. S., Rim, S., Saribay, S. A., & Kressel, L. M. (2012). Controversies, questions, and prospects for spontaneous social inferences. *Social and Personality Psychology Compass*, *6*, 657–673. 10.1111/j.1751-9004.2012.00452.x

Uleman, J. S., Saribay, S. A., & Gonzalez, C. M. (2008). Spontaneous inferences, implicit impressions, and implicit theories. *Annual Review of Psychology*, *59*, 329–360. 10.1146/annurev.psych.59.103006.093707

Varnum, M. E. W., Na, J., Murata, A., & Kitayama, S. (2012). Social class differences in N400 indicate differences in spontaneous trait inference. *Journal of Experimental Psychology: General*, *141*, 518–526. 10.1037/a0026104

Walter, K. V., Conroy-Beam, D., Buss, D. M., Asao, K., Sorokowska, A., Sorokowski, P., Aavik, T., Akello, G., Alhabahba, M. M., Alm, C., Amjad, N., Anjum, A., Atama, C. S., Atamtürk Duyar, D., Ayebare, R., Batres, C., Bendixen, M., Bensafia, A., Bizumic, B., … Zupančič, M. (2020). Sex differences in mate preferences across 45 countries: A large-scale replication. *Psychological Science*, *31*, 408–423. 10.1177/0956797620904154

Wang, M., Xia, J., & Yang, F. (2015). Flexibility of spontaneous trait inferences: The interactive effects of mood and gender stereotypes. *Social Cognition*, *33*, 345–358. 10.1521/soco.2015.33.4.1

Wang, M., Yan, B., Yang, F., & Zhao, Y. (2018). The development of spontaneous trait inferences about the actor and spontaneous trait transferences about the informant: Evidence from children aged 8–13 years. *International Journal of Psychology*, *53*, 269–277. 10.1002/ijop.12367

Wang, M., Zhao, Y., Li, Q., & Yang, F. (2016). The effects of mood on spontaneous trait inferences about the actor: Evidence from Chinese undergraduates. *Scandinavian Journal of Psychology*, *57*, 250–255. doi:10.1111/sjop.12283

Wells, B. M., Skowronski, J. J., Crawford, M. T., Scherer, C. R., & Carlston, D. E. (2011). Inference making and linking both require thinking: Spontaneous trait inference and spontaneous trait transference both rely on working memory capacity. *Journal of Experimental Social Psychology*, *47*, 1116–1126. 10.1016/j.jesp.2011.05.013

Winter, L., & Uleman, J. S. (1984). When are social judgments made? Evidence for the spontaneousness of trait inferences. *Journal of Personality and Social Psychology*, *47*, 237–252. 10.1037/0022-3514.47.2.237

Yan, X., Wang, M., & Zhang, Q. (2012). Effects of gender stereotypes on spontaneous trait inferences and the moderating role of gender schematicity: Evidence from Chinese undergraduates. *Social Cognition*, *30*, 220–231. 10.1521/soco.2012.30.2.220

Yang, F., & Wang, M. (2016). Do bosses and subordinates make spontaneous trait inferences equally often? The effects of power on spontaneous trait inferences. *Social Cognition*, *34*, 271–285. 10.1521/soco.2016.34.4.2

Zárate, M. A., Uleman, J. S., & Voils, C. I. (2001). Effects of culture and processing goals on the activation and binding of trait concepts. *Social Cognition*, *19*, 295–323. 10.1521/soco.19.3.295.21469

Zhang, Q., & Fang, N. (2016). The relationship between spontaneous trait inferences and spontaneous situational inferences: A developmental approach. *Social Behavior and Personality: An International Journal*, *44*, 569–578. 10.2224/sbp.2016.44.4.569

Zhang, Q., & Wang, M. (2013). The development of spontaneous trait inferences: Evidence from Chinese children. *Psychological Reports*, *112*, 887–899. 10.2466/21.07.PR0.112.3.887-899

Zhang, Q., & Wang, M. (2018). The primacy-of-warmth effect on spontaneous trait inferences and the moderating role of trait valence: Evidence from Chinese undergraduates. *Frontiers in Psychology*, *9*. 10.3389/fpsyg.2018.02148

18 The Updating of First Impressions

Gordon B. Moskowitz[1], *Irmak Olcaysoy Okten*[2], *and Erica Schneid*[3]
[1]*Lehigh University*
[2]*Florida State University*
[3]*Comcast, Philadelphia, PA*

Changing first impressions is difficult. A long history of research in social psychology on the updating of first impressions reveals them to be "sticky" or resistant to alteration. However, updating can occur despite this propensity for "stickiness." Updating an initial impression (especially an incorrect initial impression) not only is at the heart of important interpersonal dynamics involving psychological reactions including forgiveness, accountability, apology, blame, behavior modification, and self-awareness. It is also at the heart of important intergroup dynamics; social ills such as stereotyping, prejudice, and discrimination often stem from incorrect initial assessments rooted in historical inaccuracies and associations to unwanted negative beliefs and affect. Addressing both interpersonal and intergroup dysfunction and conflict relies on updating first impressions. For this reason, research on updating is located in parallel literatures on stereotyping involving social perceptions about groups and person perception relating to individuals as targets of perception. In this chapter, we focus on how and when updating of groups, group members, and individuals can occur. Why are first impressions resistant to movement, and what unmoors them?

Research on how and when we change first impressions has typically focused on three questions. The first question (actually a set of related questions) centers on the importance of *trait judgments* about others as a form of first impression. The power of traits in establishing a stable first impression has been an assumption in place since Asch's (1946) famous study documenting a primacy effect in impressions that are based on multiple pieces of trait information about a given actor (see also Anderson & Barrios, 1961). Are inferences about, and attributions to, a person's personality traits ubiquitous and automatic? Does this ubiquity make the impression difficult to change? If they do change, how do trait-based impressions change (e.g., Anderson & Barrios, 1961; Asch, 1946; Bargh & Thein, 1985; Gilbert, 1998; Jones, 1979; Nordstrom et al., 1998; Park, 1989; Ross, 1977; Wood & Eagly, 1981)? Will other types of impressions that are not trait-based be more open to change?

The second question is about the persistence of *stereotypes*. Updating of first impressions has been explored through this interest in the social ills that

DOI: 10.4324/9781003045687-21

are caused by first impressions being rooted in negative stereotypes. As such, the alteration of those stereotypes may be a means to affect a reduction of those social ills. This research addresses when and if information that is inconsistent with a stereotype can alter the stereotypic impression initially formed about a person (e.g., Crocker et al., 1983; Locksley et al., 1980; Macrae et al., 1993; Sherman et al., 1998; Stangor & Ruble, 1989). This research has its roots in Allport's (1954) "contact hypothesis" where intergroup contact was believed to be a basis for shattering stereotypes through exposure to the counter-stereotypic information that one would experience during cross-group friendships. This research also addresses how stereotypes are a mental shortcut and how issues of cognitive capacity and motivation lead people to rely on such shortcuts.

A third question that could also be characterized as relating to the changing of first impressions is the question of how and when attitudes change (e.g., Eagly & Chaiken, 1993). However, social influence and persuasion are typically regarded as unique areas of study, separate from work on updating first impressions about particular individuals or groups. We, too, will treat social influence as a separate category of updating that is beyond the scope of this review. An exception will be our inclusion in our review of simple affective impressions of a person (inferences tagging a person's behavior or character as positive versus negative) as a type of first impression that can be updated.

Types of Impressions

To address the issue of updating an impression it must first be recognized that there are a variety of *types* of first impressions that perceivers use to construct meaning in their social world. The updating of first impressions, and the psychological processes through which successful modification is achieved, may be dependent upon the type of first impression formed. One distinction among types of first impressions centers around awareness of the impression—first impressions of which one is aware versus those that are not consciously detected. This is a distinction among *explicit* versus *implicit* impressions. The literature on how and when people update an impression has largely focused on explicit impressions (for reviews, see Moskowitz, 2005, and Hamilton & Stroessner, 2021). There is not a large body of research on the updating of implicit impressions (though see Cone & Ferguson, 2015; Gawronski et al., 2018; Mann & Ferguson, 2017; Olcaysoy Okten et al., 2019), which begs the question, how might the updating of an implicit impression diverge from what we know about the updating of explicit impressions?

To understand such differences in updating of impressions, one must first explore whether there are differences in how explicit versus implicit impressions form. Uleman and Moskowitz (1994) examined whether the accessibility of impression formation goals—having an *explicit* "hypothesis-testing" goal (Park, 1989)—made people form impressions differently than

not having an explicit goal, where the inference process is relatively automatic. A consensus across the studies reviewed was that people formed trait inferences to comparable degrees under different task instructions fostering explicit impression formation goals and in the absence of such explicit instructions (see also Carlston & Skowronski, 1994; McCarthy & Skowronski, 2011). Examinations of active brain regions during explicit and implicit impressions similarly revealed minimal differences between the two types of impressions (other than a few additional areas associated with more elaborative thinking that were activated during explicit impressions; Ma et al., 2011). Despite there being limited differences in *forming* implicit versus explicit impressions, there might be differences in *updating* such impressions. In this chapter, we extend the discussion of the updating of first impressions to include implicit impressions.

A second distinction in types of impression formation centers around the issue of whether the initial inference that is formed refers to a goal, a trait, a state, or an attitude of the person being perceived. What are the referents of our inferences (to what quality of the person being perceived does the inference refer)? A large majority of research in impression formation has been dedicated to one particular type of inference as the basis for a first impression—the *trait inference* (for a review see Moskowitz & Olcaysoy Okten, 2016). In most of this body of work, participants indicate they have been thinking about the traits of a target person either by offering trait words in a free response task when asked to describe a person, or responding to trait words that are provided to them in the experiment (such as in a cued recall or recognition memory task). A trait inference can represent one's impression of a single individual based on an isolated behavior from a stranger (e.g., Winter & Uleman, 1984). It can represent an inference about an individual drawn from prior knowledge about a *social role* that person adopts (Person X belongs to an occupation group that is composed of mean people; e.g., Chen et al., 2014). It can represent an inference about an individual drawn from a group stereotype, based merely on prior knowledge triggered when the person is categorized as a group member from mere physical cues such as facial features or skin tone (e.g., Chen et al., this volume; Wigboldus et al., 2003). Traits can even refer to groups as opposed to individuals (e.g., Hamilton et al., 2015).

However, traits are not the only way we refer to and describe people (see Heider, 1958). The linguistic fact, at least in the English language, is that the same word used to refer to a trait can also refer to a state, a type of action, and a goal (e.g., Bassili, 1989; Fiedler & Schenck, 2001; Fiedler et al., 2005). Regardless of the reasons traits became a central focus, the reality is not only that other types of inference exist (e.g., goal inferences, state inferences, etc.), but it is wholly possible that researchers have been unknowingly studying such inferences all along, rather than traits. That is, a person can be seen as *mean* by virtue of a personality trait, but it is also true that the same word—*mean*—can describe other qualities, which we review next. Beyond trait inferences, what alternative types of inferences exist?

A first type of inference that is not a trait is an inference about the type of behavior that is observed; the category of the thing being described. Psychologists refer to these as a *gist inference* (e.g., Winter & Uleman, 1984), and as behavior identification or *categorization* (Trope, 1986). Such an inference may simply summarize the behavior, serving as a distillation of the gist of the actions observed without implicating the person at all ("that was a mean action"). A second type of inference that is not a trait is what psychologists refer to as a *state inference*. That kind of inference happens when the behavior describes the state of the person who performed that behavior (e.g., Kruse & Degner, 2021). For example, the word *mean* can be a trait, and it can be a mere summary of the action (a mean act), but it can also be an inference that tethers that action summary in the perceiver's memory to the person who performed it. Such an inference is about the state of that individual at that time without suggesting anything about personality (Person X was in a mean state; Person X performed a mean action).

A third type of inference that is not a trait is what psychologists refer to as a *goal inference* (e.g., Moskowitz & Olcaysoy Okten, 2016). Goal inferences are also tethered to the person, but additionally represent the goal or intention of that person, not a trait (Person X had the goal to be mean in this context). A fourth type of inference that is not a trait is what psychologists refer to as a *predictive inference* (McKoon & Ratcliff, 1986). A predictive inference is one that not only describes what the person is doing, but explicitly posits what they will likely do next (Person X is about to do something mean). A fifth type of inference that is not a trait is what psychologists refer to as an *evaluative inference* (e.g., Schneid, Carlston, & Skowronski, 2015).

A third distinction in types of impression formation centers around the issue of whether the initial inference that is formed even refers to a quality one believes to be possessed by the perceiver or simply to one's own affective or evaluative reaction to the person. This is a difference between semantic knowledge being triggered that is ascribed to the person's disposition (goal or trait or atttiude) versus one's own evaluative reaction (positive or negative) being triggered. All of the inferences described up to this point can be grouped together as "semantic inferences" or inferences about a semantic quality, such as "mean." However, an evaluative inference is not about a specific quality, rather it is an inference about the affect assigned to that person (Person X behaved negatively; I have a negative evaluation of Person X). How does updating of a first impression change when it is based on semantic information versus an evaluative response? Is it dependent on the type of semantic information that is inferred (a trait versus a goal, for example)? Our chapter shall address this issue of inference type as well.

Given that people tend to form a variety of semantic inferences (states, goals, traits, predictions) in parallel (e.g., Ham & Vonk, 2003), it is critical to understand how the formation of these different inferences impacts the revision "potentials" of one another. For instance, goal inferences can operate to enhance or suppress trait inferences in different circumstances, and

vice versa. Inferring the goal of socializing can foster inferring the trait of extraversion, whereas inferring the goal of *seeming* sociable for strategic reasons may not relate to or even suppress such trait inference. Again, inferring the trait of extraversion can impact the goal inferred from the behavior and its potential to be revised in light of new information.[1]

Why Is the Updating of First Impressions Difficult?

The research on *explicitly* formed impressions identifies multiple reasons as to why such first impressions are resistant to change. A first reason explicitly formed first impressions resist updating is the ease and efficiency of relying on what is already known. A second reason first impressions resist updating is an aversion to effortful processing in cases where the motivation to do so is lacking. A third reason first impressions resist updating is there are often cases in which renouncing a socially shared first impression (e.g., a stereotype, a group norm) becomes tantamount to dissent and makes one subjected to public scrutiny for non-normative expression. A fourth reason first impressions resist updating is that perceivers desire a sense of control that is most efficiently achieved by being able to predict what another person is like, and what that person is likely to do, as delivered by the existing impression. A fifth reason first impressions resist updating is the comfort one feels in having a perceived ability to prepare appropriate action based on those predictions. Finally, a sixth reason that explicitly formed first impressions resist updating is that people have a basic cognitive bias to anchor on a belief that has already been formed (with that anchor preventing subsequent judgment from drifting away).

However, explicit processes are merely the surface of how people make sense of others. Given the necessity for sense-making when humans are encountered by the events, people, and objects that constantly bombard them in the stimulus world (e.g., Asch, 1946; Bruner, 1957; Heider, 1944), the impression formation process becomes routinized and habitualized over experience so that it can happen without conscious intent or awareness. Winter and Uleman (1984) introduced this influential hypothesis using Asch (1946, p. 258) as a leaping off point: "we look at a person and immediately a certain impression of his character forms itself in us … We know that such impressions form with remarkable rapidity and great ease … we can no more prevent its rapid growth than we can avoid perceiving a given visual object." Focusing on the ease with which such processes occur, Uleman and colleagues posited and demonstrated that we categorize the behaviors we observe others perform in terms of their corresponding traits, inferring enduring qualities in them without our awareness or intent. We form *spontaneous trait inferences* (STIs) to attain meaning and prepare for action, without any manifest purpose in forming these inferences. Their functionality breeds their automaticity.

This was an important adjustment to our understanding of the inference-making process, one that allowed for it to be functional for guiding behavior, yet invisible to the person who is being guided. Garcia-Marques et al.

(Chapter 7, this volume) described it as the special ingredient that had been missing from prior decades of theorizing about impression formation. Melnikoff and Bargh (Chapter 10, this volume) described it as a discovery that transformed our understanding of the unconscious. Here, we extend the importance of this idea of the "unconscious inference" further. Though not discussed at the time by Winter and Uleman (1984), STIs were among the first types of *implicit bias* systematically studied (see Moskowitz & Roman, 1992). As described by Sherman (Chapter 12, this volume), the concept of implicit inference opened the door to the study of implicit stereotyping.

In the more specific context of the current chapter, *implicitly* formed first impressions are also resistant to change, perhaps more difficult to change. For example, Gregg et al. (2006) showed that people's initial implicit (unlike explicit) impressions of two fictional groups remained the same even when the experimenter said they accidentally flipped the information about the groups in the first place (Experiment 3) or that the groups' moral characteristics changed over time (Experiment 4). Peters and Gawronski (2011) also found that perceivers updated explicit but not implicit evaluations after receiving information that undermined the validity of the initial information. The reasons that explicit impressions are difficult to change apply to implicit inferences as well. For example, they have the utility of allowing people to predict what a person is likely to do in the future (e.g., McCarthy & Skowronski, 2011; Olcaysoy Okten et al., 2019). This affords one the desired feeling of control over their environment—one knows what to expect and how to act even if one is unaware of where that feeling comes from. However, in addition to the challenges to updating revealed by research on explicit inference, there are even further difficulties for updating implicit inferences. We detail three further challenges to updating that we identify as made apparent by research on implicit inferences.

First, implicit impressions are those that the individual often does not even recognize having formed, even if directly asked (e.g., Todorov & Uleman, 2002; Uleman & Moskowitz, 1994). It would be more unusual, relative to an explicit impression that one intended to form and is aware exists, to seek to update and alter an impression that one did not intend to form and does not know exists. Its invisibility provides an enhancement to the typical durability seen in explicit first impressions. Lack of awareness is the cognitive reason why implicit impressions are especially hard to update, a fact illustrated by research on stereotyping. Stereotypes are nothing more than an inference drawn about a person. They often take the form of a trait inference, one that is shaped by existing beliefs about a group to which the individual belongs. When this inference is an implicit one, perceivers fail to see that it is they who are *imposing* a trait on the person, rather than the person *revealing* a trait to them. If this process of implicitly inferring stereotypes is somewhat ubiquitous, as the literature on implicit bias suggests, it would mean the impression of the individual and group is constantly being reinforced through an invisible gathering of supportive evidence. As such, the ubiquity of

forming stereotypic inferences about a person's traits makes updating of the impression more difficult because the quality becomes linked silently not only to the individual's enduring qualities, but also to the enduring qualities believed to be held by the group.[2] Both the heightened stability of the inference, and its shared nature, are features bestowed upon the inference by its implicit nature that make it more resistant to change.

A second heightened obstacle to updating introduced by implicit impressions is motivated in nature. Even if one ultimately becomes aware of an implicitly formed first impression, it should seem to the perceiver to be accurate and veridical because of the ease with which it formed. Such ease creates the sense that if one did not need to work to form an inference it must be because it reflects a natural property of the person about whom this inference so fluently arose. This heightened sense of confidence in the inference, ironic since less work was actually done to achieve this greater confidence, should leave one less motivated to alter the impression (e.g., Chaiken et al., 1989; Wilson & Brekke, 1994). Thus, any motivational factors that make the formation of STIs more fluent and easy should also make them harder to change. Such fluency is the basic nature of STI, but it is shown to be enhanced by motives that promote specific inferences. For example, Uleman et al. (1986) showed that authoritarian individuals made different types of implicit first impressions than non-authoritarians. Olcaysoy Okten and Moskowitz (2020a) showed that conservatives formed different types of implicit impressions than liberals. Moskowitz (1993) showed that people high in need for structure were more likely to tie their inferences to the person who was being perceived, making the inference more stable. We argue that these motivated STIs make updating more difficult. First, because they have the quality of seeming especially easy to produce and to tie to the person. Second, because they serve a goal of the perceiver, unknowingly promoting goal pursuit.

A third heightened obstacle to updating introduced by implicit impressions is linked to the issue of types of inferences, reviewed above. In the course of examining how implicit trait inferences differ from implicit evaluative inferences, research has suggested that different types of cognitive processing are invoked; perhaps even different processing *systems*. This allows for the possibility that each type of implicit impression is updated via different guiding factors that introduce further obstacles to impression updating.

How Is Impression Updating Defined, Especially for Implicit Impression Updating?

To determine whether an impression has been updated, researchers should look to see if perceivers' implicit or explicit judgments of a person have changed. But the question actually is not as simple as it seems, and that is because updating can be indicated through examining cognitive processes other than judgment. An impression is not merely determined by one's

judgment of a person, particularly since explicit and implicit assessments of impressions can diverge (e.g., Rydell & McConnell, 2006). Such divergence offers very different views of what perceivers' impressions actually are and whether they have indeed been updated. Moreover, an impression is also revealed by memory measures. What an individual believes about a person can be assessed, perhaps with even greater accuracy, by examining how information about a person is organized and structured in memory (e.g., Hastie et al., 1980; Srull & Wyer, 1986). To assess whether updating has occurred is not as straightforward as it seems on its face.

Importantly, there is research that reveals how memory measures and judgment measures diverge in what they suggest about the perceiver's impression; these two forms of examining impressions do not always yield the same conclusions (e.g., Sherman et al., 1998). The judgment might suggest nothing has changed, whereas the memory structure may have been reorganized substantially. Research evidence shows that after forming a first impression from observing a set of behaviors, perceivers have superior recall for newer behaviors they learn that contradict the first behaviors (e.g., Hastie & Kumar, 1979). This suggests that people overturned the first impression as evidenced by the superior recall for the newer behavior. However, research also shows that judgments of a person are typically consistent with the original information and are not in line with the new information learned. This can even be true when they have superior recall for the newer information (e.g., Bargh & Thein, 1985; Chartrand & Bargh, 1996; Hastie & Kumar, 1979). This evidence from one's judgment suggests that people do not usually update their impressions. Is it updating even if the person's *judgment* does not reflect the new information stored in the memory structure?

We turn now to discussing ways that memory can be used to reveal impression updating, even if judgment suggests the impression has not been updated. Is the initial inference *replaced* in the memory representation with an entirely new inference? Or is the representation merely altered so that additional associations make the memory structure different and more complex? If the original implied trait remains present in memory, still able to be retrieved, does this constitute updating of the impression? Is the original impression entirely overwritten in memory so no trace of it remains? Rather than remain agnostic as to whether updating must be defined by *replacement* of the original inference in memory and the simultaneous change in the judgment to reflect this replaced inference, we instead offer a firm definition of impression updating as rooted in memory updating. We define updating as: a change in the representation in memory of a person/group, even if that change to the representation does not impact judgment of the person/group at that point in time, or in that particular context. A change in the representation creates the opportunity for new, updated aspects to the representation to be triggered and to impact judgment at some other point in time, or in some other context, where processing goals might deem it relevant to draw on that information. We now review three dominant ways in which memory updating has been treated

in the literature and allow for each of these processes to be included as part of our definition of impression updating.

Addition

Perhaps the most common form of change to a memory structure that may result in an updating of how one consciously describes their impression of a person is the addition of new information. *Addition* refers to the initial associations remaining intact in memory, but to these associations are added associations to new behaviors and inferences about those new behaviors (e.g., Anderson, 1965; 1974). For example, learning that a person engaged in a kind behavior after already having inferred, from a prior behavior, that the person was cruel, would lead to the memory structure now including associations to both behaviors and to inferences that the person has been both kind and cruel. Such an addition does not over-write the initial inference; the fact that one knows the person had been cruel is unchanged. Instead, both inferences exist and would be capable of guiding how one responds to that individual in the future—with a likelihood that the first impression (cruel) could be reinstated or the new inference (kind) might predominate.

Negation

A second way in which the updating of an impression has been examined is through a process of *negation* in which the initially formed impression is undermined or weakened by either new information or training (e.g., Calanchini et al., 2013; Kawakami et al., 2000; Wyer, 2010). This could occur in the manner described by Petty and colleagues (e.g., Petty & Briñol, 2010; Petty et al., 2007; Petty et al., 2006) in which updating to the memory structure occurs through falsifying or invalidating the initial inference and altering the memory structure by adding a "tag" to the inference that labels it as false. Therefore, the inference is intact, but marked as incorrect. A tag that identifies an inference as false (negating it) was described as being particularly difficult to retrieve from memory when implicit measures of attitudes are being used, making it more difficult to update an implicit as compared to an explicit attitude. Negation can also occur through information that changes the meaning of the initial behavior, thus negating the original inference. In such instances, a new inference replaces the original one as the strongest association to the person (e.g., Kawakami et al., 2000; Mann & Ferguson, 2015; 2017; Wyer, 2010).

Memory Reconsolidation

The last way to define updating of first impressions has been subject to far fewer empirical examinations, and it involves a process known as *memory reconsolidation*. Similar to negation, reconsolidation involves a restructuring of the associations. However, the process through which this occurs is based

on models of memory updating in cognitive psychology in which the memory structure has the original impression overwritten so that it cannot be reinstated; the new information will be integrated with the original. (e.g., Hupbach et al., 2007; Hupbach et al., 2008; Nader et al., 2000). The restructuring of the associations in memory begins with the *reactivation* of the initial associations while one is receiving the new information that can serve to update that first impression. Reactivation of the old impression as the new impression is being formed will create two nodes of information that are activated simultaneously, creating an opportunity for an association to be formed among them (e.g., Read & Miller, 1998). Next, for the new impression to replace the old one, there must be a period in which reconsolidation of the memory structure can occur, such as overnight during sleep (see Bray et al., Chapter 20, this volume). Evidence for such memory updating is currently limited to memory for objects, but it holds promise for its applications to first impressions of people. Lupo and Zárate (2019) have explored how a first impression can be extended from an individual who belongs to a particular group to other members of the group as a result of reconsolidation. However, this is not about the updating of first impressions, but the generality of first impressions. Several labs are now exploring the possibilities for updating of impressions via reconsolidation, and hope to demonstrate ways in which stereotypic impressions may even be changed by processes of memory reconsolidation (e.g., Ferguson et al., 2020).

Having defined the types of updating that may occur, we can now explore the evidence for updating that exists in the literature. The collective research suggests that there are information-driven factors (behavior extremity and consistency) and perceiver-driven factors (cognitive capacity and motivation) that dictate the updating of a first impression. With information-driven factors, the burden for a change to the perceiver's impression is placed on the targets of that impression. The perceiver's impression will update if the target of that impression acts in a fashion that manipulates the type of information the perceiver receives and thus compels updating. People can compel updating through confronting others (see Chaney et al., Chapter 21, this volume), by acting inconsistently with stereotypes and expectations, by displaying diagnostic acts (see Shen & Ferguson, Chapter 19, this volume), etc. Unlike with information-driven factors where the impetus for change is initiated from a source external to the perceiver, with perceiver-driven factors the burden for change is placed on the biased perceiver. The perceiver assumes the responsibility, requiring them to take the initiative to alter motivations and cognitive processing.

We begin with the information-driven factors that promote updating—factors such as diagnosticity of the new information that could serve to update the impression, the consistency of the new information with what was expected given the first impression, and the ability of the new information to alter the meaning of the original behavior by providing contingencies and context. We then transition to discussing two categories of perceiver-driven factors

that support updating of first impressions—the motives the perceiver adopts when considering other people and the processing capacity of the individual (and the effort exerted given the constraints on the perceiver's processing capacity).

Evidence for Impression Updating: Aspects of the New Information Encountered

What does the literature inform us about when impressions are updated to reflect new information about a person because of the qualities of that new information that compel the person to update? When the impetus to change comes from outside the perceiver, one common source is a confrontation. Confrontation of an existing impression that is in need of updating can come in the form of an intervention or workshop that raises awareness about bias, a media presentation that creates such awareness, social protest movements that demand justice, and interpersonal contact. There is an excellent review on the working of confrontation to change stereotypic impressions in the current volume, and we will refer the reader to Chaney et al. (Chapter 21, this volume). There is also literature warning of the dangers of confrontation at producing updating, yielding backlash and defensiveness instead (e.g., Howell & Ratliff, 2017; Stone et al., 2011; Vitriol & Moskowitz, 2021). Because there are these excellent recent reviews about when confrontation will or will not lead to updating, we focus instead on the learning of new behaviors that challenge an existing impression, rather than on explicit attempts to change the impression through a confrontation with a person or group.

One quality of new behavior we learn about a person is its consistency with what we already expect. Those expectations can come from stereotypes that tie an existing belief about a group to a new person, for whom the first impression reflects these "similar" others (e.g., Kashima, 2000). Expectations can also come from recently learned, new behaviors and traits that establish the impression. When subsequently encountered information is incongruent with these expectations, if the behavior is inconsistent with the first impression, will the impression change? Evidence suggests that the answer to this question depends on whether the first impression is explicitly formed or implicitly formed, with the factors that determine the impact of inconsistent information being slightly different for each of these types of impressions. We review explicit updating in the face of inconsistent information first, then the very recent evidence on implicit impression updating in the face of inconsistent information.

Updating Explicit Impressions from Inconsistent Information

The earliest research on first impressions gave people lists of traits to use when forming an impression. Some of the lists contained traits that were

incongruous among the set (Asch, 1946). In such experiments the list of traits were all learned simultaneously, so it is less clearly an examination of updating, as opposed to integrating a set of new information into a coherent structure. However, this can be construed as relevant to updating if the initial traits in the list are leading to a first impression that then needs to be updated by the later traits in the list. Asch described such a situation when describing primacy effects in the emergent impression. And what he found was no evidence for updating; the overall impression either ignored or explained-away the items that were incongruous with the primacy effect. However, Asch used only a measure of judgment and perhaps, had he included memory measures, may have seen that the inconsistent items were better recalled and encoded more strongly (given that great efforts were made to explain them away, providing them with processing effort).

Asch and Zukier (1984) moved from incongruous traits to specifically asking participants to form impressions of a person who displayed traits *antagonistic* to one another. They illustrated updating of a first impression using judgment measures through what we have identified as processes of "addition." They discussed three distinct types of addition that led to the judgment being updated. One type of addition they describe is through participants assigning goals to the person to explain the inconsistent behavior. The inconsistent behavior is not treated as a trait of the person, but a means to which a current goal could be achieved. This renders the behavior no longer inconsistent with a trait of the person, but consistent with a goal of the person. It might be argued that this is not an updating of the impression, but the use of attributional logic to dismiss the items that are inconsistent by attributing those behaviors to temporary goals. However, research from our own lab reveals that participants do not dismiss the behaviors they assign to goals, or fail to update their memory structure. We find that when people form goal inferences to inconsistent behavior, they have strong memory for those behaviors and associate the goals with the person. The impression has been added to and updated. The type of semantic inference formed—be it a trait inference or goal inference—can allow for updating of the impression.

Asch and Zukier illustrated a different type of addition by participants interpolating; the perceivers added new information beyond that which had actually been provided to them. By interpolating in this way perceivers were able to explain the inconsistency. Here the updating occurs by the perceiver fabricating new information rather than relying on facts they have learned. Finally, a third type of addition described participants as "segregating" the inconsistent traits to a "separate sphere" of the person. Perceivers make an exception of the inconsistent behavior; it is still acknowledged to be descriptive of the person, despite being inconsistent with the general impression held. A similar type of segregating inconsistent behavior is seen when the initial expectation comes from a stereotype. Kunda and Oleson (1995) showed that updating the impression of the person and the group occurs through the creation of a subtype for whom the inconsistent behavior

makes sense. Rather than negate the initial impression (a stereotype in this case), an update is made that allows for exceptions to the stereotype.

The more common way to study impression updating is not Asch's (1946) method of presenting a set of traits simultaneously, but through first establishing an expectation that serves as a first impression and then learning about behaviors that are inconsistent with that impression. As mentioned previously, the initial expectation can come from an existing stereotype or from initially learned behaviors. Stangor and Ruble (1989) argued that the ability of behaviors that are inconsistent with expectations to cause impression updating is tied to the strength of the original expectation; how well-developed it is. If weak or not fully formed, defensive strategies may not yet have been formed to block and disregard any inconsistent information. More strongly held stereotypes may block stereotype-inconsistent information, or engage processing to explain-away the inconsistencies, discounting them (e.g., Crocker et al., 1983). For example, stereotype-consistent acts may be attributed to a trait, but inconsistent acts attributed to pressure in the situation that does not warrant altering one's impression.

Yzerbyt et al. (1998) make a similar point relating to expectation strength with their work on group entitativity. Entitativity is the degree to which a group is seen as coherent and uniform; in a sense it is the "groupiness" of the group (e.g., Allison & Messick, 1988; Hamilton et al., 2011). Yzerbyt et al. (1998) propose that people expect more similarity among highly entitative groups, and thus the more attention-grabbing it is when stereotype-inconsistent behavior is encountered. The inconsistency seems incoherent and in need of explaining when coming from such a group. This could lead to more updating if it forces the perceiver to process the inconsistencies deeply in order to justify why they do not truly describe the person. However, it is often the case that, as Crocker et al. (1983) showed, people simply attribute the inconsistent behavior to something external to the person and need not bother processing it any further. Avoiding and ignoring inconsistencies are not elements of updating.

What this discussion introduces is an important mediating variable. It is not the consistency or inconsistency of the behavior per se that causes impression updating, but the depth of information processing in which the perceiver must engage (e.g., Rogers et al.,1977). If an inconsistency is either ignored or attributed to something irrelevant to the target, little attention is afforded it, and its impact is scant. For example, when comparing memory for a single piece of old and new information about an actor, Rothbart et al. (1979) showed a memory advantage for expectancy-consistent information—perceivers avoided and ignored behaviors that violated an expectation. Consistent information has an advantage at capturing and keeping attention, and the impression is not updated. However, if the perceiver focuses on the inconsistency and wrestles with explaining it, they expend effort to process it in order to keep the first impression intact. These efforts to keep the initial impression intact will ironically confer

associative strength to the inconsistent items. This is ironic because it results in the impression having been modified to reflect the unwanted new information, adding new associations to such inconsistent information to the representation of the person. These additions render the original impression more fragile by making the representation more complex (e.g., Bargh & Thein, 1985; Hastie et al., 1980; Macrae et al., 1993; Stangor & McMillan, 1992; see for a review Moskowitz, 2005). Memory evidence points to hope that updating can occur, but it requires not merely being exposed to information that is inconsistent with an expectation, but allocating processing depth to that information.

Hastie and Kumar (1979), for example, showed that people remember expectancy-incongruent information better than expectancy-congruent information, as long as the inconsistent behaviors stay distinct (e.g., they are low in number; but see also Bargh & Thein, 1985; Hemsley & Marmurek, 1982). Research in the stereotyping literature supports this finding. For example, illusory correlation research shows that perceivers recall stereotype-confirming information better than disconfirming information when the two are equally available. When all else are equal, a quiet librarian (stereotype-consistent) is remembered better than a quiet salesman (stereotype-inconsistent). However, Hamilton and Rose (1980) provide evidence that updating of the stereotype is possible when the inconsistent behavior is frequently repeated. If the salesman is not equally quiet to the librarian, but described by behaviors suggesting a quiet disposition even more frequently than a librarian, participants show a recall advantage for the inconsistent behaviors that imply the salesman has the trait of "quiet." The behavior should be particularly salient in this context. Not only does the salesman act in a stereotype-inconsistent way, but he acts that way more often than someone for whom such behavior is expected. Srull's (1981) findings also show superior recall for inconsistent information due to its violation of a strong expectation about the individual. The effort needed to process the inconsistency provides it with a richness of encoding that yields a memory advantage.

Research on whether people update when they encounter stereotype-inconsistent information also often uses both measures of judgment (what is your impression of the individual) and measures of memory (what behaviors of the individual do you recall) in the same experiment. This research reveals that despite the memory advantage often seen for inconsistent information, the resulting impression is often consistent with the original impression. That is, one type of evidence suggests updating and the other does not. A complex relationship among the processing of information, memory for information, and the change in judgment exists. Sherman (Chapter 12, this volume) describes how these relationships can be understood by considering that processing the details and characteristics of a behavior can be different than the processing of the meaning that one might infer from those details. In an important set of experiments, Sherman et al. (1998) showed that memories

for behaviors that are relevant to a stereotype can be *qualitatively* (and not necessarily quantitatively) different. Specifically, Sherman et al. suggested that stereotype-incongruent (vs. congruent) information is likely to be perceptually encoded (i.e., with greater attention to details) rather than conceptually encoded (i.e., with greater attention to the gist). The *details* of stereotype-inconsistent information were recognized better in a word identification task than the *traits* implied by that same stereotype-inconsistent information. These results were important in showing that encoding new expectancy-incongruent information with all its details would likely not be enough to change the judgment, as the perceiver may still be blind to the gist of this new information, i.e., from whence the inconsistency originates (see also Jerónimo et al., 2015). Cognitive effort may result in the heightened memory for stereotype inconsistent information and all its detail, but judgment measures might suggest that no updating has occurred. We contend that the memory restructuring alone is sufficient evidence for updating.

Processing depth is essential to updating an explicit impression. This is best illustrated by work that focuses on the fact that sometimes *consistent* behavior leads to updating, if the consistency can trigger greater depth of processing. Moscovici (1976) performed an historical analysis of minority groups who caused social change and came to the conclusion that those achieving success at transforming how they were viewed and treated in the larger culture typically did so with intransigence. That is, they remained steadfast in repeating their actions and beliefs, maintaining the consistency of their unpopular positions and beliefs. An example is consistent nonviolent protest in the American civil rights movement of the 1960s. These insights led to experimental work to illustrate the updating of beliefs about minorities when the minority behavior is consistently expressed. While much of this research is about social influence (agreement with a minority on some perceptual event), some of this research explores how beliefs about the qualities of the minority group can be updated due to their clear and unyielding expression of their minority view. The mechanism revealed for such minority influence at belief updating is depth of processing. Whereas stereotypes allow perceivers to dismiss and ignore minority views, minority consistency in the face of strong social disapproval can create surprise, threat, and curiosity that opens the perceiver to switching from heuristic dismissal of the minority view to engaging it. At first, such engagement is focused on counter-arguing, rationalizing, and explaining-away the minority position. Later, such engagement may reveal something as simple as common ground, points of agreement, making the wall that divides the groups seem less impenetrable (e.g., Wood et al., 1996). These greater processing efforts have created new associations to the minority that have updated the initial belief about the group (e.g., Baker & Petty, 1994; Moskowitz, 1996). Even if the belief itself does not shift to favor the minority view, the overall impression has changed to represent their views and newly discovered commonalities (and belief could eventually change).

Diagnosticity may make the consistent information grab attention. Persistence in the face of normative pressure to yield reveals something diagnostic about the group. Diagnostic information is attention grabbing as it is salient and distinct; as such, it receives extra processing focus, resulting in a recall advantage for it (e.g., Sherman & Hamilton, 1994). Consider negative information. Because negative information is highly distinctive, negative impressions are more easily formed and more difficult to revise than positive ones (Baumeister et al., 2001; Hess & Pullen, 1994; Skowronski & Carlston, 1989). A similar pattern has also been observed in attributions; perceivers were less likely to update initial negative (vs. positive) attributions over time (Carlston, 1980) and in response to contradictory information (e.g., Anderson, 1965; Briscoe et al., 1967; Hess & Pullen, 1994; Reeder & Coovert, 1986; Park, 1986; Skowronski & Carlston, 1992; Ybarra, 2001). Extreme information is also diagnostic, such as when the behavior is immoral (e.g., Cone & Ferguson, 2015; Peters & Gawronski, 2011; Rydell, et al., 2006). Brambilla et al. (2019) showed that morality-related inferences are stronger in degree and morality-related (vs. competence-related) new information is more likely to update impressions. This suggests that when the initial impression is one of immorality it will be hard to update, but when it is the new information that suggests the highly diagnostic case of immoral behavior, the first impression may be easier to update.

Locksley and colleagues (Locksley et al., 1980; Locksley et al., 1982) provided evidence that people readily use information to update a stereotype if the behavior is diagnostic. Stereotypes are resistant to being changed and are likely to be used when encountering new information, especially if that new information is ambiguous and somewhat open to interpretation (e.g., Allport, 1954; Devine, 1989; Hamilton & Stroessner, 2021). Locksley and colleagues illustrated that when the new information is clear and unambiguous behavior that is non-stereotypic it can promote impression updating. Their research hits perceivers over the head with acts that are diagnostic in their clear violation of a strong expectation. This triggers cognitive processing that will not allow the inconsistencies to be ignored or rationalized-away.[3]

To summarize, when does updating of an explicit impression occur in response to inconsistent information? First, it occurs when the first impression is weak. It also occurs when the person has time to examine the perceptual details of the information. It occurs when the gist of the inconsistent information is allowed to be extracted. It also occurs when the inconsistent behavior is frequently repeated; and when the relative frequency of the inconsistent behavior is compared against how frequently that behavior occurs for someone for whom it is expected (base rates are taken into account). Finally, updating of an explicit impression occurs when new information is diagnostic. In general, a variety of aspects of the new information that is encountered that promote processing depth also promote updating. Of course, such updating is not always a complete reconsolidation of the memory

that eliminates all initial inferences; one could update a negative explicit impression of a target but still harbor the negative affect from the initial inference.

Updating Implicit Impressions from Inconsistent Information

Will inconsistent behavior also update an implicit impression? As individuals are not aware of initial implicit impressions, perceivers do not consciously detect that the new information is inconsistent. They are unaware of even having a first impression, let alone that anything can be inconsistent with it. Does this make updating harder or easier? It is true that defenses may be weaker since one does not have a conscious impression to defend or protect. However, it is also true that the perceiver is not intentionally forming impressions, and thus updating may be harder if one does not intend to do it. These are very new questions, and evidence is limited to work in the last 15 years. Ma et al. (2012) showed that trait-inconsistent new information about a target activated different brain regions depending on whether the perceiver initially formed their first impression of this target explicitly or implicitly. Two regions that have been associated with conflict monitoring, posterior medial frontal cortex (pmFC) and right prefrontal cortex (rPFC), were more strongly activated during explicit (vs. implicit) inference-making. Such findings also suggest the potential for system-level differences between implicit and explicit processing that could impact updating (for reviews, see Smith & DeCoster, 2000; Satpute & Lieberman, 2006). We shall return to this possibility later. At the very least it suggests that updating will differ for implicit versus explicit impressions.

For instance, in a narrative comprehension task, Rapp and Kendeou (2007) showed that impressions of story characters are more likely to be updated to incorporate trait-inconsistent information about a character as the story unfolds if the readers were explicitly told to evaluate the characters' actions carefully, rather than left to form impressions implicitly. Consistent with what we reviewed above regarding explicit impression updating, the inconsistent information receives extra processing and the first impression is modified (especially if the person is motivated to be careful or accurate). However, Rapp and Kendeou also found that in the absence of explicit instructions, readers still showed evidence of recognizing trait-inconsistent information about a character in the story (they spent a longer time reading such information). As in Sherman et al. (1998), they also found that participants did not use their heightened processing of such information to update their judgment accordingly. Despite this divergence with judgment, the memory advantage for inconsistent information suggests to us that updating of the implicit impression still occurred.

Research by Gawronski and colleagues on the updating of impressions is a second example of updating of an implicit impression, in their case through *addition* processes (e.g., Gawronski et al., 2010, 2015; 2018). In their work on evaluative inference, impressions are formed about a single target after

learning multiple pieces of consistent information about that target, following a classic procedure used in research on implicit measures of attitudes (e.g., Boucher & Rydell, 2012; Rydell et al., 2006; Wilson et al., 2000). Later, participants learn new information about the person in either the same context or a different context. Since these are evaluative inferences, inconsistent information would imply that an initial inference of "good" would be updated by an inference of "bad" while an initial inference of "bad" would be updated by an inference of "good." Gawronski and colleagues argue that first impressions are more likely to be context-free and encoded in a more abstract fashion, whereas information learned later is contextualized, tied to the learning environment. The new impressions are formed and added to the memory structure with the original impression, with the associations differing in the degree to which they tie the inference to the learning context. They posit that an initial inference that a person is "bad" would be formed in a generalized fashion, but information learned later that is inconsistent with this initial impression—the added inference that the person is "good"—would be contextualized. That is, the perceiver infers that the inconsistent behavior only occurred due to the presence of a situational pressure such as someone the actor needed to impress, or an immediate goal that compelled the action. The initial behavior is untethered to the context in this way, and thus is hard to overwrite. It can be added to, but not overturned or negated. If both evaluations exist, which will be used when making judgments and decisions about that person? Gawronski et al. (2014) argue that the new associations predominate when the evaluation occurs in a context similar to the context in which that new information was learned. In all other contexts, the first impression will predominate.

Winter et al. (2021) found that implicit bias, in the form of negative evaluations of outgroups, were reduced without any explicit attempt to control those attitudes. This was accomplished by inducing more flexible processing through having participants contemplate inconsistent information. They assessed cognitive flexibility with a categorization task in which participants evaluate the fit to a category (e.g., vehicles) of objects (e.g., car). Higher cognitive flexibility reduced the degree to which judgments aligned with the dominant outgroup attitude. This research is consistent with other areas of research that explore motivations that make people more flexible and, similarly, reduce implicit bias. Such research will be reviewed shortly when we review person-driven factors that lead to updating. In this case, the motivation to be more flexible arises from encountering stereotype-inconsistent acts.

Wyer (2010) also provides evidence that an implicit evaluation can be updated after stereotype-inconsistent acts. However, updating occurred if, and only if, the initial information that served as the basis for the first impression was reactivated immediately after the new information had been presented. For example, one may have initially received behavioral information about a bald-headed person that indicated that person was a

skinhead. Later, one might learn that the initial information received was not accurate, and that the person was actually a cancer patient, not a skinhead. The explicit impressions were easily updated, the implicit impressions were not. However, updating of the implicit impression did occur if participants were able to revisit the original behaviors and reassess the person in light of the new information. Wyer interpreted this finding as support for "reinterpretation" of the initial impression in light of later information that allowed for re-elaboration of that evidence. Wyer (2016) further argued that if implicit evaluations are unlikely to reverse without elaborative or integrative processing, then people who are less likely to engage in such processing should be less likely to update the implicit impression. Individuals high in need for structure who are unlikely to engage in the revisiting of initial associations in the presence of new contradictory information (e.g., Moskowitz, 1993) were found in these studies to be less likely to reverse implicit evaluations. Wyer did not speculate as to why reinterpretation (and updating) of the initial information would be more likely following the inconsistent information when a reminder was present. We suggest this might be a good example of the reconsolidation account of updating reviewed earlier; as the older network of information was reactivated while receiving the new information, integration of the two (and thus restructuring) may have become more likely. Perhaps also in support of this model is work by Gawronski et al. (2010) that illustrated that the updating of an initially formed impression was more likely when initially learned information shared salient features with the new information. It is possible that this shared feature cued or reactivated the first impression as the new information was being learned.

The clearest way to test a reconsolidation account of the updating of first impressions after learning new information that was inconsistent with that first impression would be to explicitly manipulate conditions of reactivation and memory reconsolidation. Olcaysoy Okten and Moskowitz (2020b) were the first to do so.[4] They found support for the idea that reconsolidation allowed for a change in the first impression by examining the formation and updating of implicit inferences—STIs and spontaneous goal inferences (SGIs). They found that an initially formed STI was not replaced by a new STI that considered new information learned; initial STIs were preserved in the face of contradictory information. However, this does not mean the impression failed to update. There was evidence of an addition process. While the initial trait associations were preserved, new STIs were formed and added to the memory structure as well. Perhaps providing some evidence for restructuring of the impression, participants not only formed new STIs, they also formed new *goal inferences*. Specifically, SGIs got stronger when memory was reconsolidated and trait-*inconsistent* new information was learned. SGIs got weaker when memory was reconsolidated and trait-consistent new information was learned. This suggests that context-dependent semantic inferences like SGIs are easier to move than context-independent semantic inferences like STIs. It also suggests that while the initial inference may

remain intact, the impression has been modified to reflect not only new traits, but something less stable about the person—a malleable goal that ties the action to their current situation.

Although implicit first impressions may be harder to overturn than explicit ones, the early evidence suggests that they can be overturned when one encounters information inconsistent with the implicitly held impression. This may require reactivation of the initial impression. It may require the initial impression being less decontextualized, or the newer impression being less contextualized. However information that is inconsistent with an impression that one does not even know exists can lead to the updating of the invisible impression.

Updating Impressions from Information-Driven Factors Other Than Inconsistency

The inconsistency of new information is not the only quality of information that makes it garner attention and elaboration. Often it is the case that new information does not contradict what was initially learned, but reimagines what was learned. New information can describe contingencies that alter how we interpret what the initially learned behaviors mean. Can we update an impression, even an implicit impression, in the light of such contingencies?

Mann and Ferguson (2015, 2017) found that updating of the evaluation required not only that the new information provided a contingency to the original behavior, but that the perceiver was able to engage in effortful deliberation about the original information. Participants needed to reinterpret the meaning of the original behavior *in light of the contingency* information in order to update their impression. For example, an initial behavior of damaging a neighbor's house led to a negative inference. However, that was replaced or updated to be a positive inference after that negative inference was met with later learning that introduced a contingency. For example, the neighbor's house had been damaged because the target actor was rescuing children from the house while it was on fire. This is an example of updating through negation, where a contingency allows one to overturn the original negative impression. Of course, it is possible that the updating occurred not through negation and the total reversal of the initial affect, but through addition. People could be preserving both evaluations (maintaining positive and negative associations to the person) but to different degrees, with one association being stronger than the other. Though the precise mechanism of updating is not clear, the updating of the evaluation is clear.

Mann and Ferguson (2015) assessed explicit evaluations; their participants *intended* to form inferences and were aware of having formed inferences.[5] Can contingency information allow people to update impressions that were formed spontaneously, with impression formation and updating happening without awareness and without intent? We know of only one experiment to

address this question. Olcaysoy Okten et al. (2019), explored both evaluative inferences and trait inferences (SEIs versus STIs). They exposed participants to two pieces of information that painted two different portraits of an actor: an original piece plus a second, contingency piece. The contingency in this research is additional information that changes the meaning of the behavior on which the initial inference was based. For example, participants are asked to memorize a sentence such as "Dave screamed at the child," which was immediately followed by the contingency "because the child was about to be burned by the hot stove." In such sentences, one might form an STI that the man was cruel and then form an STI that the man was protective or kind. However, the second STI is not only additional information, it contradicts the first STI and could serve to overwrite it (one could argue that logically it should overwrite it). The impressions formed are implicit since the procedures follow the methods of spontaneous inference paradigms—the participant is never asked to form a judgment, merely told to memorize the items, and the behaviors are single actions that describe a series of single individuals, not multiple behaviors about the same individual. The experiment allowed for the testing of both the initial STI and the subsequent STI. The questions of interest are 1) is the implicit impression that is formed altered after processing the contingency, and 2) is the existence of updating dependent on whether the inference in question is an SEI or an STI?

This experiment showed that SEIs were updated through an apparent negation; the evaluation of the target person reversed upon learning contingency information. Participants updated their SEIs (e.g., from positive to negative) in accordance with the full scenario. As discussed above in regard to the Mann and Ferguson (2017) finding, it is possible that the updating occurred not through negation but through addition, with both evaluations being preserved. SEIs were updated, but the mechanism is not yet clear. In contrast, the semantic inference (STIs) formed did not reverse. Participants did not switch from inferring the man to be cruel to him being kind. Thus, there is no negation as an updating mechanism of STIs. However, this does not mean the impression was not updated. Instead, updating of STIs occurred in the form of addition; new inferences based on the contingency information were formed just as powerfully as those formed based on the original information. Participants formed an inference of "cruel" but then added to the representation the subsequent inference of "kind" tied to the same actor. If they later have the capacity and motivation, it may be possible that there is some type of consolidation or reconfiguration that might occur so that the newer trait negates the initial inference. Future studies should examine this question.

The findings in Experiment 2 of Olcaysoy Okten et al. (2019) were consistent with this suggestion. While an implicit measure again showed that both the original and contingency-based inferences existed in memory, this study also included another method of assessing impressions—predictions about future behavior. With an addition process, multiple inferences about the person are stored. The question then arises as to which inference would

be the one used for making predictions about the person and preparing to interact with that person. In particular, if the inferences contradict one another, it is important to know which one will guide the perceiver. When asked explicitly how the person will act in the future, the prediction made by participants was based on the new/additional associations, not the initial STI. Therefore, perceivers did indeed consider contingencies of the actors' behaviors and revise their trait inferences accordingly *when given the time to do so* in a more elaborative, effortful task. This question of restructuring due to memory reconsolidation processes should be explored further with follow-up experiments. Issues of which inference guides judgment and behavior when multiple inferences are stored, and when reconsolidation of the entire memory structure occurs, could depend on the strength of the association as determined by the quantity (e.g., Rydell & McConnell, 2006; Rydell et al., 2007) and the quality (e.g., extremity, negativity, persuasiveness; Cone & Ferguson, 2015; Petty & Brinol, 2010) of information on which the inference is based. It could depend on the extent to which contextual cues or goals make one versus the other association more accessible (as discussed previously in the work of Gawronski).

In summary, features of the information itself can garner attention and spur more elaborate processing. Indeed, when learning expectancy-incongruent information, people need more time and greater cognitive capacity (as they are typically slower and more subject to cognitive load; Hemsley & Marmurek, 1982; Srull, 1981). The information itself can promote updating if it has qualities that shake perceivers from their attachment to first impressions and opens them to processing this new information more elaborately, or reprocessing earlier learned information. Even implicit impressions are more easily updated if one is engaged in more elaborate processing, despite not knowing they are re-assessing a former impression. The features of information that promote such elaboration of implicit impressions include salient acts, extreme acts, diagnostic acts, acts inconsistent with expectations, acts that remain consistent in the face of great pressure to yield, acts that heighten the accessibility of negation "tags," acts that reactivate the first impression and allow for memory reconsolidation, and acts that provide contingencies or alter the interpretation of what was learned originally. Though less is known about this newer discipline of implicit impression updating, it may depend on the type of implicit inference that is made as a first impression. We shall discuss this at the end of the chapter.

Finally, the notion that information-driven factors can spur more elaborate processing so that updating can occur opens the door to a complementary matter. Do people have the capacity and willingness to elaborate? For inconsistent behavior to change an impression, the perceiver must be able to perform the effortful processing that allows for the encoding of all the details that suggest one should update. Further, even if able to put those details together, one must *want* to update. That is, the inconsistent information, processed so diligently, must introduce doubt that relying on the existing impression is

acceptable. It is possible that people exert a great deal of effort to shield themselves from experiencing such doubt. This is a competing form of effort surrounding an impression that blocks updating, belying the effort spent on analyzing the contradictory details. For updating to occur, this competition must be resolved so that types of processing effort work in unison. Updating relies on one being motivated to use deeply processed information to either add new impressions or negate existing ones. We turn now to exploring how perceiver-driven factors such as the capacity to elaborate and the motivation to elaborate determine if impression updating will occur.

Evidence for Impression Updating: The Perceiver's Capacity and Motivation to Update

As described at the start of this chapter, people are resistant to updating a first impression for a variety of reasons. Some of these are tied to the limited processing capacity that constrains what perceivers are able to think about at any given point in time. For example, one must detect and attend to new information for it to be able to motivate updating. Yet capacity limitations may prevent people from paying attention to information inconsistent with an initial expectation. This can result in memory for such information never being encoded and having the power to change the representation or the judgment of the person. Impressions can change when a person exerts the processing effort to consider new information.

Anderson and Hubert (1963) suggested that one type of impression stickiness—a primacy effect—may stem from attentional deficits when processing new information about others. Consistent with this suggestion, increasing attention to new information through continuous responding (by repeatedly testing perceivers' impressions subsequent to each piece of new information) created a recency effect instead of a primacy effect, a form of updating, in explicit impressions (Briscoe et al., 1967; Luchins, 1958; Stewart, 1965). Other studies documented that primacy effects are especially likely when perceivers are under time pressure (Kruglanski & Freund, 1983), stress (Nordstrom et al., 1998), or fatigued (Webster et al., 1996). In general, updating an explicit impression (in these examples, a primacy effect) is believed to require mental effort or capacity, and hence any limits to processing capacity reduce the ability to engage in the mental work needed for updating to occur (e.g., Brewer, 1988; Chaiken et al., 1989; Fiske & Neuberg, 1990; Wilson & Brekke, 1994). Increasing capacity, or removing such constraints, allows one to exert the effort to update the impression.

Park (1989; see also Trope & Bassok, 1983) called the impression formation process one of "hypothesis testing." The initially encountered information operates as a "trait hypothesis" during the impression formation process. Observed actions are then tested against the initial trait hypothesis. That is, explicit impressions would constantly be subject to an anchoring-adjustment process, where first inferences are used as anchors (Tversky &

Kahneman, 1974) and adjusted over time as long as certain conditions are met. Once again, cognitive capacity and attentional deficits are seen as a key to whether adjustments to first impressions can be made. Evidence for this comes from explorations of the correspondence bias (Jones, 1979). In this bias, the perceiver uses a trait to explain an observed behavior, even when that initial trait inference should be updated to take into account situational pressures that may have compelled the behavior.

Jones (1979), and later his student Gilbert (1998), argued that these initial "trait" anchors are hard to overturn, with limits to cognitive capacity undermining the updating of such a trait-biased first impression (e.g., Gilbert et al., 1988; 1989). Under time constraints and conditions of divided attention, the initial trait "anchor" is used as the impression, even when information that suggests updating is present (e.g., Jones & Davis, 1965). With the removal of these constraints, the perceiver adjusts the initial inference (see also Quattrone, 1982) to consider the situation in which the behavior occurred and the goals of the individual. Only then does updating of the initial correspondent inference occur to allow the individual to factor in the force of the situation that compels people to act (just as Jones & Davis, 1965, described). Updating of an initial trait inference occurs in the manner specified by the theory of correspondent inference if the perceiver has the cognitive capacity to engage in updating. In the 50 years from Asch (1946) to Gilbert (1998), we see many types of evidence for first impressions formed around traits, and updating of those impressions requires cognitive capacity and mental effort. We have reviewed evidence for this role of cognitive capacity in updating from the correspondence bias, primacy effects, minority influence, and how people respond to information inconsistent with a first impression. Rather than provide an exhaustive review of the relationship between capacity and updating in other domains of person perception (such as stereotyping, priming effects, nonverbal behavior, etc.), let us shift to the next point, one regarding motivation and updating. Even when capacity is available, people still often fail to update an impression (e.g., Gilbert, 1998).

Motives That Facilitate Explicit Updating

As noted previously, the stickiness of first impressions is multiply determined, and one such determining factor is people find them efficient, easy, and useful as predictions. That is, they are motivated to use them as a means of providing meaning, and unmotivated to change them as it would mean sacrificing both control and meaning. Thus, one needs not only the capacity to update a first impression, but the motivation to do so. A variety of motives internal to the perceiver can engender this motivation to process information that may update prior beliefs. There is far too large a literature on the topic of motivated cognition to provide an exhaustive review here, even if we limit the discussion to motivated changes to first impressions. We provide some representative samples here, and then spend the rest of this section discussing newer work on

motives that drive updating in implicit impressions. But first, a brief overview of work from the updating of explicit impressions, where the motives examined share an ability to provoke flexibility in how one thinks about an existing impression, opening one to contemplating if it needs updating. Such motives include: accountability (Tetlock, 1983), accuracy goals (Neuberg, 1989), promotion focus (Liberman et al., 1999), need for cognition (Cacioppo & Petty, 1982), creativity (Sassenberg et al., 2017), emotional ambiguity (e.g., Rothman & Melwani, 2017), empathy or perspective taking goals (e.g., Galinsky & Moskowitz, 2000), interdependence (e.g., Fiske & Neuberg, 1990; Gaertner & Dovidio, 2000), egalitarianism (e.g., Moskowitz, 2010), guilt alleviation (e.g., Monteith et al., 2002), impression management (e.g., Plant & Devine, 1998), dissonance reduction (e.g., Moskowitz & Vitriol, 2022), open-mindedness (e.g., Gollwitzer, 1993), identity management (e.g., Van Bavel & Packer, 2021), and a general desire to have confidence in one's judgment about the person (e.g., Chaiken et al., 1989).

Some of the motives, such as empathy and identity enhancement, provoke flexibility in how one thinks by drawing the target person into one's identity circle through specifying a shared group membership or experience. For example, asking one to take another's perspective (Galinsky & Moskowitz, 2000) can increase the degree of self-other overlap that is perceived to exist, which softens the reliance on the first impression when shaping one's impression of the other. Identifying a shared identity (e.g., Gaertner & Dovidio, 2000) can move a person from being an untrusted outsider to a valued insider. A student of Lehigh University may be viewed scornfully by a student from a rival university, such as Bucknell, due to first impressions suggested by stereotypes. But, if both students are re-categorized not as from opposing groups, but as representatives of the same collegiate sports conference (the Patriot League), the stereotype is inhibited from inclusion in the impression and the shared features become salient.

Other motives, such as to be creative and to be open-minded, create what Sassenberg et al. (2022) call a flexibility mindset in which people become more open in the types of information they consider as relevant to their impression. They engage in greater deliberation surrounding a wider array of information. Yet other motives, such as egalitarianism (Moskowitz, 2010), accuracy (Neuberg & Fiske, 1987), interdependence (Fiske & Neuberg, 1990), and accountability (Tetlock, 1983), lead people to strive for heightened fairness that can only be achieved by scrutiny of the veracity of the inferences on which their impression is based. These motives provide a willingness to make changes to the impression when the existing beliefs and attitudes are called into doubt by this scrutiny. Interdependence goals and accountability goals that raise concerns about the accuracy of judgments lead to a decrease in primacy effects and an increase in using new information in one's ultimate judgment (e.g., Webster et al., 1996; Tetlock, 1983). In the absence of such explicit motivating conditions, perceivers are unlikely to engage in the effort to update a first impression. However, if motivated to be

accurate (through goals such as being interdependent or being held accountable) perceivers can counter-argue with their prior beliefs, examine the contradictory evidence, dedicate the time to resolve the inconsistencies, and deliberate about the meaning of the new information. (e.g., Chaiken et al., 1989; Fiske & Neuberg, 1990).

Impression management goals lead people to be concerned with the appearance of being biased (e.g., Plant & Devine, 1998), which forces them to, at least when in public, alter their impression to fit with what is socially acceptable and normative. And as research on dissonance reduction goals shows, such public commitment to a belief can lead to the alteration or updating of the internalized belief. The individual is motivated to have consistency among what they privately believe and what they do and say in public. Finally, as reviewed by Chaney et al. (Chapter 21, this volume), guilt is a powerful motivator of compensatory behavior and cognition. When one feels remorse or regret over actions taken or beliefs held, change is motivated so that new actions can seek to undo past wrongs, and belief updating can replace unwarranted impressions. Updating in this view is a form of self-regulation (e.g., Monteith, et al., 2002; Moskowitz, 2010).

Motives That Facilitate Implicit Updating

Consistent with work on explicit updating, implicit updating is also possible through motivation, but not because the motives make one desire to change the impression itself (such as a goal to be accurate or accountable). Rather, it is because goals impact the flexibility and depth of cognitive processing more broadly, having an indirect impact on how people consider and weigh information relevant to an impression. One research domain to address questions of implicit updating is research on implicit bias.[6] Perhaps the first example of an implicit impression being updated comes from work on implicit stereotyping. Lepore and Brown (1997) showed that White individuals who were *low in prejudice* had altered their personal beliefs about the group "Blacks." They had learned the social stereotype about this group, but updated this stereotype to establish a personal belief about the group that no longer associated negative traits with the group. When the category label was triggered for such individuals, the resulting impressions formed did not reflect the socially shared stereotype.

Moskowitz (2010) discusses a similar type of updating of a stereotype that arises when a person has *chronic egalitarian goals*. Not only do such individuals have different associations to a group such as "women" or "Black men" that have replaced the social stereotype, they also exhibit negation processes in which the stereotype itself is inhibited. Inhibition provides evidence that the associations to the original inferences that dominated the first impression still exist, but they are weaker, relative to other associations to the group, in the updated impression. The updated impression instead has strong associations to egalitarian goals rather than negative stereotypes. Moskowitz (2010)

makes a distinction among chronic egalitarians, who have *updated* their impression, versus people with a temporary egalitarian goal who are *controlling* their impressions. The result in the moment is the same—implicit bias is not merely curtailed, but inhibited. However, the reason for the bias having been mitigated is subtly different. Chronic egalitarians have updated their impression in memory. Temporary egalitarian goals allow for the momentary triggering of processes that inhibit stereotypes and negative affect, but the structure remains unaltered, and once the goal to be egalitarian is released, the bias returns.

The conclusion to be drawn from such findings is not that a temporary goal is an ineffective tool for updating an implicit stereotype or prejudice. As reviewed by Chaney et al. (Chapter 21, this volume) they are necessary first steps that can curtail bias in the short term. With time and commitment, the goals can be triggered repeatedly in similar situations and eventually lead to more permanent control and the updating of the memory structure. However, it is also true that this often requires conscious and explicit commitment to the goal. Goals that promote control over bias—such as to be egalitarian, accurate, or accountable—explicitly focus the individual on becoming aware of the implicit impression and changing it. For implicit bias to be updated through *implicit* processes, the goals must spur processing that allows for the consideration and integration of information inconsistent with the bias in a way that does not make updating explicit. This can happen when the goals are implicit (such as with the chronic goals mentioned previously, or primed goals), or when the goals are explicit but seemingly unrelated to impression updating. Examples of the latter include perspective-taking goals (e.g., Galinsky & Moskowitz, 2000), creativity goals (Sassenberg et al., 2022), goals that promote emotional ambivalence (e.g., Rothman & Melwani, 2017), and goals to find a common identity. While not an exhaustive list, such goals alter implicit bias by changing the reliance on snap judgments and broadening the range of information considered about the target person (either by including the self-concept in the analysis, or by widening the range of information deliberated upon).

Another example of stereotype updating comes from the research of Kawakami et al. (2000) in which negation training was used to alter the associations to a stereotype. Through repeatedly saying "no" to qualities once associated with the group, the memory structure was altered to reflect the weakened association of these qualities in the updated impression. Further, Dasgupta and Asgari (2004) also showed that counter-stereotypic exemplars (as opposed to specific counter-stereotypic behaviors) could negate the implicit association of negative affect to a stereotyped group. While negative associations may not turn positive in these examples, the group stereotypes were, nonetheless, updated so that positive associations to the group were added to the representation and the existing negative associations were weakened. A similar updating through weakening negative association has been shown to result after participants in applied settings have completed an

intervention workshop (e.g., Forscher et al., 2017; Stone et al., 2020). Such interventions appear to be successful when they do more than just raise awareness of bias in the moment, but teach strategies to update the associations to the targeted group (e.g., Moskowitz & Vitriol, 2022). As noted previously, this is a key for distinguishing among merely controlling the expression of bias in the moment and the updating of the impression so that the impression has changed. Training, counter-stereotypic exemplars, and awareness-focused workshops may merely be teaching people to control the expression of a bias rather than update it. However, when these strategies are focused on negating the bias and teaching new and more appropriate ways to respond, we see them as a form of updating.

There is not a large body of research examining the motivations that promote the updating of an implicit impression, outside of the work on implicit bias. If we were to treat the ease with which an impression can be updated as a type of motivation to update, then different types of implicit impressions might motivate people to update to different degrees. For example, it is possible to expect different types of semantic inferences to be updated at different rates. Given that states and goals are assumed to be more situation-dependent and, therefore, more temporary than traits, it is possible to expect that a perceiver might be more motivated to update these more malleable inferences than the more stable trait inference. We are not aware of any study that has compared the updating rates of different types of spontaneous semantic inferences. A similar prediction might be made about updating evaluative inferences relative to semantic inferences. There are clear differences in the dimensional complexity of semantic and evaluative inferences. Evaluative inferences generally lie across the "good-bad" or "positive-negative" continuum, while semantic trait or goal inferences can lie across a number of different continua. This would also be a rich area for future investigation, to determine whether this complexity matters and how much it affects the motivation and ability to update. In one relevant study, reviewed earlier, Olcaysoy Okten at al. (2019) found the negating of SEIs in the moment, with little need for time or effort. The negating of STIs seemed to occur only with time and effort. Differences in the updating of evaluative and semantic inferences aligns with the possibility that there are different systems for processing evaluative versus semantic information. A distinction has been drawn since the time of Asch (1946) between a general evaluation and the semantic concepts that may be attached to a person, and how each could independently organize information about the person in distinct ways and through different mechanisms. What can a different systems approach tell us about the possibilities for updating different types of inferences?

Evaluative and Semantic Systems

The single memory system argument states that an individual's response is a product of a single memory system housing linked semantic and evaluative

concepts (e.g., Devine, 1989; Gawronski & Bodenhausen, 2006; Greenwald & Banaji, 1995; Smith & DeCoster, 2000). In these models, once a concept is activated, that activation spreads through these links to related concepts. By contrast, multiple systems models propose that semantic and evaluative processes occur via independent, parallel processes. While evidence is not yet definitive, nearly all of the work on evaluative vs. semantic implicit impression formation and updating has supported the latter.

Carlston's (1992, 1994) Associated Systems Theory (AST) proposed that trait inferences occur in a verbal/cognitive system, whereas evaluative inferences occur in an affective system, making the two types of inferences relatively independent—although they may also both contribute to an overall impression. Interestingly, it appears as though these components are most likely to be independent when the impression formation process is more spontaneous or implicit. Once the process transitions to a more conscious or explicit impression formation, the different representations work together as the individual analyzes all available information to determine their impression(s).

In addition to the AST, Amodio and Ratner (2011) developed the Memory Systems Model (MSM), which proposes that there are multiple systems within implicit memory, all with different functional properties. They suggest that the system for semantic processes is separate and distinct from the systems that process affective information. Amodio and his colleagues have related implicit stereotypes to a semantic associative memory system, and implicit evaluations (prejudice) to affective memory systems such as those involved in classical conditioning (Amodio & Devine, 2006; Amodio & Hamilton, 2012). Their neural research provides evidence that evaluative processes increase activity in the amygdala and orbital frontal cortex, but semantic processes increase activation in the inferior frontal gyrus and the medial prefrontal cortex (e.g., Amodio, 2014; Gilbert et al., 2012; Itkes et al., 2017; Rissman et al., 2003). Other neurological evidence also shows that impression formation tasks that promote trait inferences increase activation in the medial prefrontal cortex (e.g., Mitchell et al., 2005; Mitchell et al., 2006).

Therefore, if evaluative and semantic impressions do develop though separate systems, the processes followed and conditions necessary for updating within one system would not necessarily map onto change in the other (e.g., Olcaysoy Okten & Moskowitz, 2020b; Olcaysoy Okten et al., 2019). As noted, this view has been supported by the work done in the area of implicit impression formation. Specifically, these studies have examined whether semantic and evaluative processes operate independently (and in parallel). For example, Schneid, Carlston, and Skowronski (2015) found that perceivers simultaneously make spontaneous inferences about traits implied by a behavior and the evaluative implications of the behavior. However, the evaluations of the person seemed to play no role in informing the semantic processes through which inferences about traits are associated with the same person. They suggested that future studies might shed further light on the

matter if they explored whether spontaneous evaluations occur in the absence of explicit memory for the trait implications of the behaviors. While not addressing this issue directly, Olcaysoy Okten and Moskowitz (2020b) found that spontaneous inferences did occur without explicit memory for the behaviors on which they were based. Interestingly, updating of the impression through the incorporation of spontaneous goal inferences did require explicit memory for the behaviors. These findings suggest several questions yet to be explored about the relationship between explicit memory for the behaviors and explicit memory for initial trait inferences on both the formation of other inference types and on the updating of inferences. For example, Hupbach et al. (2022) found that when participants were told to "forget" the behaviors they learned (cued to be "forgotten" right after exposure), recall of behaviors was reduced, but recognition of traits implied from the behaviors was not affected; evidence for trait inferences' independence from explicit behavior memory.

Finally, Crawford et al. (2007) found that giving perceivers the goal of detecting a lie while being exposed to behavioral information disrupted trait inferences from being formed but had no effect on the formation of evaluative inferences. Schneid, Carlston and Skowronski (2015; Experiment 3) used the same manipulation to determine whether this manipulation also disrupted evaluative inferences (SEIs). They found a reduction in the formation of STIs—however, SEIs were *not* affected by the instruction set. These findings suggest that the formation of an evaluative inference is not reliant on the formation of a trait inference.

The separate systems idea has also been explored through research on the semantic and evaluative components of racial bias. Research using both the "Shooter task" (e.g., Correll et al., 2014; Glaser & Knowles, 2008) and interpersonal interaction (e.g., Amodio & Devine, 2006; Amodio & Hamilton, 2012) show that implicit stereotyping and implicit prejudice have distinct predictive abilities. For example, implicit stereotyping, which involves semantic processing, predicts the association of Black men with weapons, whereas implicit prejudice, which involves affective processing, does not. Similarly, expecting an interaction with an outgroup member increases nonverbal displays of bias and arousal, which involve affective processing, but is unrelated to stereotype accessibility and verbal displays of bias, which involve semantic processing.

Taken together, these behavior and neural findings support both the independence of evaluative and semantic processes, but also provide support for Bob Zajonc's (1980) much-debated argument for the primacy of affect—which holds that affective evaluations are inescapable, independent of, and separable from, cognitive responses. The MSM supports this in the argument that affective associations typically develop faster than semantic associations (Amodio & Ratner, 2011), and classical learning studies have shown that threat-related affective associations are formed faster than semantic associations (Bechara et al., 1995; March et al., 2018). Finally, Bargh

(1989) showed that participants presented with trait words at speeds below the threshold of conscious awareness were able to provide the evaluative meaning of the words on test, but unable to report the semantic meaning.

Of course, even though such evidence for a distinction among evaluative and semantic impression formation is consistent with a model of distinct cognitive and evaluative processing systems, it is not yet conclusive. Van Dessel et al. (2019) argued that distinctions seen in judgments of evaluative and semantic tasks (i.e., a change in performance in one task but not the other) could be explained by a single storage system rather than distinct systems by simply positing differential rates of memory retrieval for each processing type. These distinctions in performance could also be explained by context demands that favor the retrieval of one or the other type of information. It may likewise be the case that evaluative and semantic associations differ in their sensitivity to different types of contextual variables and interventions. For example, a semantic association could be more sensitive to information that conveys truth value than an evaluative association, and this need not implicate a system-level distinction (e.g., Lai et al., 2014).

Conclusion

When an impression is explicitly formed, updating may be difficult and fraught with obstacles rooted both in capacity limitations and lack of motivation to abandon a first impression. However, both the qualities of the new information received and the perceiver's motivations can at times impel the perceiver to engage the effort needed to update the impression. Can one update an impression they are not aware exists? Research suggests this too is possible. This is especially important when considering the updating of an implicit bias because the implicit impression has implications beyond the interpersonal domain but to greater societal issues.

However, there is no guarantee that an updated first impression will be more veridical than the first impression. The perceiver may seek to update to achieve greater accuracy, or the perceiver may be motivated to shape the impression, consciously managing it in a desired direction. The resulting impression can be inaccurate in different ways, or even more inaccurate. When thinking about what motivates a perceiver to update an impression, an important consideration is that accuracy need not be the guiding principle. When motivated and capable, people follow *theories* about how to update an impression. These theories do not always yield an accurate impression (e.g., Moskowitz & Skurnik, 1999; Wegener & Petty, 1995; Wilson & Brekke, 1994). While updating is often in the service of eliminating bias to produce an accurate impression (it is often referred to as "correction" in the psychological literature), it is not achieved if the theory about how to correct one's biased impression is flawed. This is a case of one wanting to update an impression in an accurate way, but not knowing how to do so. Accuracy in updating is also unlikely to be achieved if it was never even the goal of the

updating process; if objective truth was never the standard in the first place. Motives to produce specific biased updates to an impression can be in place, with the perceiver motivated to reach a specific conclusion rather than an accurate conclusion (e.g., Kunda, 1987; Chaiken et al., 1989).

Our concern has not been with the issue of accuracy in updating, but merely with what spurs attempts at updating.

Understanding the updating of first impressions would be significant if it merely helped us to regulate important interpersonal dynamics and have more accurate information about the individuals with whom we interact. Can we overcome correspondence bias? When are primacy effects mitigated? Are first impressions forever imprinted? When will new information be impactful? Can inferences drawn from misinformation be negated? However, the study of impression updating is not limited to the study of impressions of individual strangers about whom only limited behavioral information is known. Because of social stereotypes, individuals are also imbued with traits that go beyond the observed behaviors. The inferred stereotypes become part of the impression, shaping the very meaning assigned to the behaviors that are observed. Given the propensity for stereotyped impressions to cause harm, researchers have naturally examined how to update these types of impressions. Stereotyped inferences are tied not merely to the observed behavior, but are interconnected with a set of socially shared beliefs, providing the expectations they yield with even greater strength and resistance to change. The fact that stereotypes and prejudice are an important source of first impressions creates another level of urgency for understanding impression updating, especially the updating of implicit stereotypes and prejudice.

Like many before us, we came to social psychology as a discipline to better understand and potentially address human stereotyping and prejudice. The lead author can recall sitting in Don Taylor's social psychology class at McGill University at age 17 and having the epiphany that there was a way to formally think about these issues through academic work and scientific rigor. By age 19, Wally Lambert's year-long seminar on social psychology had shown him that work on the unconscious was evolving, and that perception itself was guided by unconscious motives. How we think about people and make attributions was determined by self-interest, unseen goals, and invisible inferences. A realization about stereotyping and prejudice was reached—it was a complex problem impacted by all the variables described by Allport (1954), but at the heart of it, changing biases required understanding the broader processes through which people make inferences, are motivated to reason in prescribed ways, and an understanding of the implicit nature of both motivation and cognition. Stereotyping was a type of inference, and like other inferences shaped by motives. The inferences and the motives could be invisible to people. Smarter folks had already had this 19-year old's epiphany, and a group of young scholars at NYU were doing the type of work that could provide this broad understanding of implicit social cognition that was

essential to updating stereotypes. Back home to New York City our lead author traveled.

The elder statesman of the NYU group was a relatively young (in his 40s) scientist named James Uleman. He was doing the only work on the planet on the implicit nature of the inferences we draw about people, a key to understanding how first impressions are formed and updated. Uleman, however, was so engrossed in the phenomenon of unconscious inference itself that he was not yet seeing the key role it played in addressing broader cultural phenomena such as stereotyping and stereotype change. As Chen, Quinn, and Maddux (this volume) point out, that cross pollination of STI and stereotyping has barely been explored to this day. However, the amazing group of even younger scholars Uleman was surrounded with (in their 20s and 30s) were similarly doing work on the implicit nature of impression formation and attitude change. At NYU there was also Susan Andersen, John Bargh, Shelly Chaiken, Madeline Heilman, Tory Higgins, Diane Ruble, Jeff Tanaka, and Yaacov Trope all doing foundational work to shape how we understand impressions and impression change. There were their students and postdocs such as Felicia Pratto, Len Newman, Tim Strauman, Chuck Stangor, Gerd Bohner, Eun Rhee, Akiva Liberman, Eva Pomerantz, and Abigail Panter. As a result of the inroads made by social cognition as a field of study, a new way of thinking about stereotyping was emerging in the discipline more broadly. What a privilege to be arriving at a time and a place where the entire discipline was being reimagined. Being in the heart of Greenwich Village with this young and collegial group of friends/scholars was the greatest environment for receiving the breadth of understanding of the underlying factors that shape our impressions and impression change. Although not any one of these scholars is likely primarily defined as a stereotyping researcher, our understanding of stereotyping, and stereotype change, would not be possible without their collective efforts to understand the nature of social cognition more broadly.

Notes

1 Chen et al. (this volume) make a similar point about how stereotypic inferences can impact the formation of trait inferences drawn from behavior. They ask whether trait inferences drawn from behavior, that are likely slower to accumulate and form than stereotypic inferences drawn from physical cues, can over-ride stereotypic inferences. Can dispositional inferences compete with stereotypic inferences as an impression forms? Rather than correcting or updating a stereotype, how do different inference types unfold over time during the formation of the first impression?

2 Moskowitz and Uleman (1987) argued that stereotypes were STIs formed from observing the behavior of a group member. Devine (1989) showed that behavior was not even necessary for stereotypic inferences to be drawn; that perceivers merely needed to perceive group membership for the stereotype to be triggered.

3 Locksley et al. (1982) posit that this may not be due to heightened depth of processing but to perceiver error. The *base-rate fallacy* occurs when perceivers know

prior probabilities, but when making predictions they fail to use those probabilities and make predictions relying on recently encountered information. Base-rates are neglected. This may be why the recently encountered, non-stereotypic behavior is used rather than the stereotype. Alternatively, it could be that giving people information about a stereotype that is clearly in violation of the expectation, and then asking people what they think about the person, is too fraught. This would almost by necessity force perceivers to use such information regardless if they truly had altered their impression.

4 On the first day of the experiment, research participants form implicit inferences from a single piece of behavioral information about each of a series of target actors (for whom images were provided). They then returned on the second day when the impression was reactivated (through seeing the image of each person) and a second piece of behavioral information was learned. This new information was either consistent with, or inconsistent with, the original behavior learned at time one. Finally, participants were given time overnight for such information to reconsolidate, and were brought back on the third day to test if their impressions were updated from what was learned on day one.

5 Both Gawronski and colleagues' work on updating "implicit" impressions and the Ferguson and colleagues' work on "implicit" impressions are not examples of implicit inference. In those two lines of research, the inferences formed by participants are intentional and made with conscious awareness. Participants read a series of behaviors about the same person (Bob) that all suggest the same impression. It is improbable that, even though not asked to form an impression, that participants are not ultimately aware of doing so given the multiple pieces of consistent information about the same person. Repeating consistent information in this way is very likely to encourage conscious processes of impression formation. The implicit aspect of the research is in the use of indirect *measures* of attitudes. Rather than illustrating the updating of an *implicit* impression, such findings reveal the updating of an explicit impression through the use of an implicit measure. An implicit measure of attitudes allows one to show that updating is real by providing a spontaneous assessment of one's current impression unspoiled by biases in reflection and retrospection (it is a more genuine impression rather than one that is an artifact of being explicitly asked for an impression). But while the measure is implicit, the inference itself is not. In contrast, tasks used to assess implicit impression formation and implicit updating avoid explicitly prompting the participant to form an impression or to reinterpret the original action. For example, research on spontaneous trait inferences: a) asks participants to simply review or familiarize themselves with stimuli rather than providing explicit impression formation instructions, and b) provides participants with limited information about multiple actors as opposed to providing many pieces of information about a single actor (which would compel conscious reasoning to fit the many pieces of information into a coherent narrative).

6 Many research teams are concerned with the issue of overturning or controlling bias. Some seek to do this by means that we would not characterize as "updating" of the implicit stereotype or attitude. For example, one line of research on controlling implicit bias seeks to modify the goals perceivers have when forming an impression, creating a motivation to curtail the expression of bias (e.g., Monteith et al., 2002; Plant & Devine, 1998). Here, external pressures to suppress the expression of a bias exist, and the bias is not changed, the individual is merely succumbing to normative pressure to not speak it. Other research seeks to control bias not by updating the impression, but by again changing the goals of the perceiver so that different aspects of the impression become salient versus inhibited. The associations stay the same, but their triggering versus inhibition is altered from moment-to-moment

according to the goals of the perceiver (e.g., Macrae et al., 1994; Moskowitz & Li, 2011). As Monteith et al. argue, such alterations are a precursor to impression change, and can lead, over time, to new associations to the person being formed. But the inhibition and activation of associations that already exist, while a form of control over what is expressed, are not what we would define as updating.

References

Allison, S. T., & Messick, D. M. (1988). The feature-positive effect, attitude strength, and degree of perceived consensus. *Personality and Social Psychology Bulletin*, *14*(2), 231–241.

Allport, F. H. (1954). The structuring of events: outline of a general theory with applications to psychology. *Psychological Review*, *61*(5), 281–303.

Amodio, D. M., & Devine, P. G. (2006). Stereotyping and evaluation in implicit race bias: evidence for independent constructs and unique effects on behavior. *Journal of Personality and Social Psychology*, *91*(4), 652–661.

Amodio, D. M., & Hamilton, H. K. (2012). Intergroup anxiety effects on implicit racial evaluation and stereotyping. *Emotion*, *12*(6), 1273–1280.

Amodio, D. M., & Ratner, K. G. (2011). A memory systems model of implicit social cognition. *Current Directions in Psychological Science*, *20*(3), 143–148.

Amodio, D. M. (2014). The neuroscience of prejudice and stereotyping. *Nature Reviews Neuroscience*, *15*(10), 670–682.

Anderson, N. H. (1965), Averaging versus adding as a stimulus-combination rule in impression formation. *Journal of Experimental Psychology*, *70*, 394–400.

Anderson, N. H. (1974), Algebraic models in perception. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception* (Vol. 2, pp. 215–298). New York: Academic Press.

Anderson, N. H., & Barrios, A. A. (1961). Primacy effects in personality impression formation. *The Journal of Abnormal and Social Psychology*, *63*(2), 346–350. 10.1037/h0046719

Anderson, N. H., & Hubert, S. (1963). Effects of concomitant verbal recall on order effects in personality impression formation. *Journal of Verbal Learning and Verbal Behavior*, *2*(5–6), 379–391.

Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology*, *41* (3), 258–290.

Asch, S. E., & Zukier, H. (1984). Thinking about persons. *Journal of Personality and Social Psychology*, *46*(6), 1230–1240.

Baker, S. M., & Petty, R. E. (1994). Majority and minority influence: Source-position imbalance as a determinant of message scrutiny. *Journal of Personality and Social Psychology*, *67*(1), 5–19.

Bargh, J. A., & Thein, R. D. (1985). Individual construct accessibility, person memory, and the recall-judgment link: The case of information overload. *Journal of Personality and Social Psychology*, *49*(5), 1129–1146. 10.1037/0022-3514.49.5.1129

Bargh, J. A. (1989). Conditional automaticity: Varieties of automatic influence in social perception and cognition. In J. S. Uleman , & J. A. Bargh (Eds.), *Unintended thought*, (pp. 3–51). New York, NY:Guilford.

Bechara, A., Tranel, D., Damasio, H., Adolphs, R., Rockland, C., & Damasio, A. R. (1995). Double dissociation of conditioning and declarative knowledge relative to the amygdala and hippocampus in humans. *Science*, *269*(5227), 1115–1118.

Bassili, J. N. (1989). Traits as action categories versus traits as person attributes in social cognition. In J. N. Bassili (Ed.), *On-line cognition in person perception* (pp. 61–89). Hillsdale, NJ: Erlbaum.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323–370.

Boucher, K. L., & Rydell, R. J. (2012). Impact of negation salience and cognitive resources on negation during attitude formation. *Personality and Social Psychology Bulletin*, *38*(10), 1329–1342.

Brambilla, M., Carraro, L., Castelli, L., & Sacchi, S. (2019). Changing impressions: Moral character dominates impression updating. *Journal of Experimental Social Psychology*, *82*, 64–73.

Brewer, M. B. (1988). A dual process model of impression formation. In T. K. Srull & R. S. Wyer (Eds.), *Advances in social cognition* (Vol. 1, pp. 1–36). Hillsdale, NJ: L. Erlbaum.

Briscoe, M. E., Woodyard, H. D., & Shaw, M. E. (1967). Personality impression change as a function of the favorableness of first impressions. *Journal of Personality*, *35*(2), 343–357.

Bruner, J. S. (1957). On perceptual readiness. *Psychological Review*, *64*(2), 123–152.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*, 116–131.

Calanchini, J., Gonsalkorale, K., Sherman, J. W., & Klauer, K. C. (2013). Counterprejudicial training reduces activation of biased associations and enhances response monitoring. *European Journal of Social Psychology*, *43*, 321–325. 10.1002/ejsp.1941.

Carlston, D. E. (1992). Impression formation and the modular mind: The associated systems theory. In L. Martin & A. Tesser (Eds.), *The construction of social judgments* (pp. 301–341). Hillsdale, NJ: Lawrence Erlbaum.

Carlston, D. E. (1980). The recall and use of traits and events in social inference processes. *Journal of Experimental Social Psychology*, *16*(4), 303–328.

Carlston, D. E. (1994). Associated systems theory: A systematic approach to cognitive representations of persons. In R. S. Wyer, Jr. (Ed.), *Associated systems theory: A systematic approach to cognitive representations of persons* (Vol. 7, pp. 1–78). Hillsdale, NJ: Lawrence Erlbaum.

Carlston, D. E., & Skowronski, J. J. (1994). Savings in the relearning of trait information as evidence for spontaneous inference generation. *Journal of Personality and Social Psychology*, *66*(5), 840–856.

Chaiken, S., Liberman, A., & Eagly, A. H. (1989). Heuristic and systematic information processing within and beyond the persuasion context. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 212–252). New York: Guilford.

Chartrand, T.L., & Bargh, J.A. (1996). Automatic activation of impression formation and memorization goals: Nonconscious goal priming reproduces effects of explicit task instructions. *Journal of Personality and Social Psychology*, *71*, 464–478. 10.1037/0022-3514.71.3.464.

Chen, J. M., Banerji, I., Moons, W. G., & Sherman, J. W. (2014). Spontaneous social role inferences. *Journal of Experimental Social Psychology*, *55*, 146–153.

Cone, J., & Ferguson, M. J. (2015). He did what? The role of diagnosticity in revising implicit evaluations. *Journal of Personality and Social Psychology*, *108*(1), 37–57.

Correll, J., Hudson, S. M., Guillermo, S., & Ma, D. S. (2014). The police officer's dilemma: A decade of research on racial bias in the decision to shoot. *Social and Personality Psychology Compass*, *8*(5), 201–213.

Crawford, M. T., Skowronski, J. J., Stiff, C., & Scherer, C. R. (2007). Interfering with inferential, but not associative, processes underlying spontaneous trait inference. *Personality and Social Psychology Bulletin*, *33*(5), 677–690. 10.1177/0146167206298567

Crocker, J., Hannah, D. B., & Weber, R. (1983). Person memory and causal attributions. *Journal of Personality and Social Psychology*, *44*(1), 55–66. 10.1037/0022-3514.44.1.55

Dasgupta, N., & Asgari, S. (2004). Seeing is believing: Exposure to counterstereotypic women leaders and its effect on the malleability of automatic gender stereotyping. *Journal of Experimental Social Psychology*, *40*(5), 642–658.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*(1), 5–18. 10.1037/0022-3514.56.1.5

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Fort Worth, TX: Harcourt Brace Jovanovich.

Ferguson, M., Moskowitz, G. B., Hupbach, A., Krosh, A., & Olcaysoy Okten, I. (2020). *Using behavioral, computational, and neural approaches to understand correction of first impressions*. National Science Foundation Grant.

Fiedler, K. & Schenck, W. (2001). Spontaneous inferences from pictorially presented behaviors. *Personality and Social Psychology Bulletin*, *27*, 1533–1546. 10.1177/01461672012711013

Fiedler, K., Schenck, W., Watling, M., & Menges, J. I. (2005). Priming trait inferences through pictures and moving pictures: The impact of open and closed mindsets. *Journal of Personality and Social Psychology*, *88*, 229 –244.

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in Experimental Social Psychology*, *23*, 1–73.

Forscher, P. S., Mitamura, C., Dix, E. L., Cox, W. T. L., & Devine, P. G. (2017). Breaking the prejudice habit: Mechanisms, timecourse, and longevity. *Journal of Experimental Social Psychology*, *72*, 133–146.

Galinsky, A. D., & Moskowitz, G. B. (2000). Perspective taking: Decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of Personality and Social Psychology*, *78*, 708–724.

Gaertner, S. L., & Dovidio, J. F. (2000). The aversive form of racism. In C. Stangor (Ed.), *Stereotypes and prejudice: Essential readings*, (pp. 289–304). Psychology Press.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*, 692–731. 10.1037/0033-2909.132.5.692

Gawronski, B., Hu, X., Rydell, R. J., Vervliet, B., & De Houwer, J. (2015). Generalization versus contextualization in automatic evaluation revisited: A meta-analysis of successful and failed replications. *Journal of Experimental Psychology: General*, *144*(4), e50–e64.

Gawronski, B., Rydell, R. J., & De Houwer, J., Brannon, S. N., Ye, Y., Vervliet, B., & Hu, X. (2018). Contextualized Attitude Change. *Advances in Experimental Social Psychology*, *57*, 1–52.

Gawronski, B., Rydell, R. J., Vervliet, B., & De Houwer, J. (2010). Generalization versus contextualization in automatic evaluation. *Journal of Experimental Psychology: General*, *139*(4), 683–701.

Gawronski, B., Ye, Y., Rydell, R. J., & De Houwer, J. (2014). Formation, representation, and activation of contextualized attitudes. *Journal of Experimental Social Psychology*, *54*, 188–203.

Gilbert, D. T. (1998). Ordinary personology. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology*. (pp. 89–150). Boston: McGraw-Hill.

Gilbert, D. T., Pelham, B. W., & Krull, D. S. (1988). On cognitive business: When person perceivers meet person perceived. *Journal of Personality and Social Psychology*, *54*, 733–740.

Gilbert, S. J., Swencionis, J. K., & Amodio, D. M. (2012). Evaluative vs. trait representation in intergroup social judgments: Distinct roles of anterior temporal lobe and prefrontal cortex. *Neuropsychologia*, *50*(14), 3600–3611.

Glaser, J., & Knowles, E. D. (2008). Implicit motivation to control prejudice. *Journal of Experimental Social Psychology*, *44*(1), 164–172.

Gollwitzer, P. M. (1993). Goal achievement: The role of intentions. *European Review of Social Psychology*, *4*(1), 141–185.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*, 4–27. 10.1037/0033-295X.102.1.4

Gregg, A. P., Seibt, B., & Banaji, M. R. (2006). Easier done than undone: Asymmetry in the malleability of implicit preferences. *Journal of Personality and Social Psychology*, *90*(1), 1–20.

Ham, J., & Vonk, R. (2003). Smart and easy: Co-occurring activation of spontaneous trait inferences and spontaneous situational inferences. *Journal of Experimental Social Psychology*, *39*(5), 434–447.

Hamilton, D. L., Chen, J. M., Ko, D. M., Winczewski, L., Banerji, I., & Thurston, J. A. (2015). Sowing the seeds of stereotypes: Spontaneous inferences about groups. *Journal of Personality and Social Psychology*, *109*(4), 569–588. 10.1037/pspa0000034

Hamilton, D. L., Chen, J. M., & Way, N. (2011). Dynamic aspects of entitativity: From group perceptions to social interaction. In R. M. Kramer, G. J. Leonardelli, & R. W. Livingston (Eds.), *Social cognition, social identity, and intergroup relations: A Festschrift in honor of Marilynn B. Brewer*. (pp. 27–52). Psychology Press.

Hamilton, D. L., & Sherman, S. J. (1996). Perceiving persons and groups. *Psychological Review*, *103*(2), 336–355.

Hamilton, D. L., & Stroessner, S. J. (2021). *Social cCognition: Understanding people and events*. Sage.

Hamilton, D. L., & Rose, T. L. (1980). Illusory correlation and the maintenance of stereotypic beliefs. *Journal of Personality and Social Psychology*, *39*(5), 832–845.

Hastie, R., & Kumar, P.A. (1979). Person memory: Personality traits as organizing principles in memory for behaviors. *Journal of Personality and Social Psychology*, *37*, 25–38. 10.1037/0022-3514.37.1.25.

Hastie, R., Ostrom, T. M., Ebbesen, E. B., Wyer, R. S. Jr., Hamilton, D. L., & Carlston, D. E. (Eds.). (1980). *Person memory: The cognitive basis of social perception*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Heider, F. (1958). The naive analysis of action. In F. Heider (Ed.), *The psychology of interpersonal relations*, (pp. 101–124). New York, NY: Wiley.

Heider, F. (1944). Social perception and phenomenal causality. *Psychological Review*, *51*(6), 358–374.

Hemsley, G. D., & Marmurek, H. H. (1982). Person memory the processing of consistent and inconsistent person information. *Personality and Social Psychology Bulletin*, *8*(3), 433–438.

Howell, J. L., & Ratliff, K. A. (2017). Not your average bigot: The better-than-average effect and defensive responding to Implicit Association Test feedback. *British Journal of Social Psychology*, *56*(1), 125–145.

Hess, T. M., & Pullen, S. M. (1994). Adult age differences in impression change processes. *Psychology and Aging*, *9*(2), 237–250.

Hupbach, A., Gomez, R., Hardt, O., & Nadel, L. (2007). Reconsolidation of episodic memories: A subtle reminder triggers integration of new information. *Learning & Memory*, *14*(1–2), 47–53. 10.1101/lm.365707

Hupbach, A., Hardt, O., Gomez, R., & Nadel, L. (2008). The dynamics of memory: Context-dependent updating. *Learning & Memory*, *15*, 574–579. 10.1101/lm.1022308.

Hupbach, A., Olcaysoy Okten, I. , & Horn, P. (2022). Directed forgetting in the social domain: Forgetting behaviors but not inferred traits. *Journal of Applied Research in Memory and Cognition*. Advance online publication.

Itkes, O., Kimchi, R., Haj-Ali, H., Shapiro, A., & Kron, A. (2017). Dissociating affective and semantic valence. *Journal of Experimental Psychology: General*, *146*(7), 924–942.

Jones, E. E. (1979). The rocky road from acts to dispositions. *American Psychologist*, *34*, 107–117.

Jerónimo, R., Garcia-Marques, L., Ferreira, M. B., & Macrae, C. N. (2015). When expectancies harm comprehension: Encoding flexibility in impression formation. Journal of Experimental Social Psychology, 61, 110–119.

Jones, E. E., & Davis, K. E. (1965). From acts to dispositions: the attribution process in social psychology. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 219–266), New York: Academic Press.

Kawakami, K., Dovidio, J. F., Moll, J., Hermsen, S., & Russin, A. (2000). Just say no (to stereotyping): Effects of training in the negation of stereotypic associations on stereotype activation. *Journal of Personality and Social Psychology*, *78*(5), 871–888. 10.1037//0022-3514.78.5.871.

Kashima, Y. (2000). Maintaining Cultural Stereotypes in the Serial Reproduction of Narratives. *Personality and Social Psychology Bulletin*, *26*, 594–604. 10.1177/0146167200267007.

Kruglanski, A. W., & Freund, T. (1983). The freezing and unfreezing of lay inferences: Effects on impressional primacy, ethnic stereotyping, and numerical anchoring. *Journal of Experimental Social Psychology*, *19*, 448–468.

Kruse, F., & Degner, J. (2021). Spontaneous state inferences. *Journal of Personality and Social Psychology*, *121*(4), 774–791.

Kunda, Z. (1987). Motivated inference: Self-serving generation and evaluation of causal theories. *Journal of Personality and Social Psychology*, *53*(4), 636–647.

Kunda, Z., & Oleson, K. C. (1995). Maintaining stereotypes in the face of disconfirmation: Constructing grounds for subtyping deviants. *Journal of Personality and Social Psychology*, *68*(4), 565–579.

Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J. -E. L., Joy-Gaba, J. A., Ho, A. K., Teachman, B. A., Wojcik, S. P., Koleva, S. P., Frazier, R. S., Heiphetz, L., Chen, E. E., Turner, R. N., Haidt, J., Kesebir, S., Hawkins, C. B., Schaefer, H. S., Rubichi, S., Sartori, G., Dial, C. M., Sriram, N., Banaji, M. R., & Nosek, B. A. (2014). Reducing implicit racial preferences: I. A comparative investigation of 17 interventions. *Journal of Experimental Psychology: General*, *143*(4), 1765–1785.

Lepore, L., & Brown, R. (1997). Category and stereotype activation: Is prejudice inevitable? *Journal of Personality and Social Psychology*, *72*(2), 275–287.

Liberman, N., Idson, L. C., Camacho, C. J., & Higgins, E. T. (1999). Promotion and prevention choices between stability and change. *Journal of Personality and Social Psychology*, *77*(6), 1135–1145.

Locksley, A., Borgida, E., Brekke, N., & Hepburn, C. (1980). Sex stereotypes and social judgment. *Journal of Personality and Social Psychology*, *39*(5), 821–831. 10.1037/0022-3514.39.5.821

Locksley, A., Hepburn, C., & Ortiz, V. (1982). Social stereotypes and judgments of individuals: An instance of the base-rate fallacy. *Journal of Experimental Social Psychology*, *18*, 23–42.

Lupo, A. K., & Zárate, M. A. (2019). Guilty by association: Time-dependent memory consolidation facilitates the generalization of negative–but not positive–person memories to group and self-judgments. *Journal of Experimental Social Psychology*, *83*, 78–87.

Luchins, A. S. (1958). Definitiveness of impression and primacy-recency in communications. *Journal of Social Psychology*, *48*(2), 275–290.

Macrae, C. N., Hewstone, M., & Griffiths, R. J. (1993). Processing load and memory for stereotype-based information. *European Journal of Social Psychology*, *23*, 77–87.

Ma, N., Vandekerckhove, M., Van Overwalle, F., Seurinck, R., & Fias, W. (2011). Spontaneous and intentional trait inferences recruit a common mentalizing network to a different degree: Spontaneous inferences activate only its core areas. *Social Neuroscience*, *6*(2), 123–138.

Ma, N., Vandekerckhove, M., Baetens, K., Van Overwalle, F., Seurinck, R., & Fias, W. (2012). Inconsistencies in spontaneous and intentional trait inferences. *Social Cognitive and Affective Neuroscience*, *7*(8), 937–950.

Macrae, C. N., Milne, A. B., & Bodenhausen, G. V. (1994). Stereotypes as energy-saving devices: A peek inside the cognitive toolbox. *Journal of Personality and Social Psychology*, *66*(1), 37–47.

Mann, T. C., & Ferguson, M. J. (2015). Can we undo our first impressions? The role of reinterpretation in reversing implicit evaluations. *Journal of Personality and Social Psychology*, *108*(6), 823–849.

Mann, T. C., & Ferguson, M. J. (2017). Reversing implicit first impressions through reinterpretation after a two-day delay. *Journal of Experimental Social Psychology*, *68*, 122–127.

McCarthy, R. J., & Skowronski, J. J. (2011). What will Phil do next?: Spontaneously inferred traits influence predictions of behavior. *Journal of Experimental Social Psychology*, *47*(2), 321–332.

March, D. S., Gaertner, L., & Olson, M. A. (2018). On the prioritized processing of threat in a dual implicit process model of evaluation. Psychological Inquiry, 29(1), 1-13.

McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(1), 82–91. 10.1037/0278-7393.12.1.82

Mitchell, J. P., Banaji, M. R., & Macrae, C. N. (2005). The link between social cognition and self-referential thought in the medial prefrontal cortex. *Journal of Cognitive Neuroscience*, *17*(8), 1306–1315. 10.1162/0898929055002418

Mitchell, J. P., Cloutier, J., Banaji, M. R., & Macrae, C. N. (2006). Medial prefrontal dissociations during processing of trait diagnostic and non-diagnostic person information. *Social Cognitive and Affective Neuroscience*, *1*(1), 49–55. 10.1093/scan/nsl007

Monteith, M. J., Ashburn-Nardo, L., Voils, C. I., & Czopp, A. M. (2002). Putting the brakes on prejudice: On the development and operation of cues for control. *Journal of Personality and Social Psychology*, *83*(5), 1029–1050. 10.1037/0022-3514.83.5.1029

Moscovici, S. (1976). *Social influence and social change*, London: Academic Press.

Moskowitz, G. B. (1993). Individual differences in social categorization. *Journal of Personality and Social Psychology*, *65*, 164–174.

Moskowitz, G. B. (1996). The mediational effects of attributions and information processing in minority social influence. *British Journal of Social Psychology*, *35*, 47–66.

Moskowitz, G. B. (2005). *Social cognition: Understanding self and others*. (A. Tesser, Ed.). Guilford Press.

Moskowitz, G. B. (2010). On the control over stereotype activation and stereotype inhibition. *Social and Personality Psychology Compass*, *4*(2), 140–158. 10.1111/j.1751-9004.2009.00251.x

Moskowitz, G. B., & Li, P. (2011). Egalitarian goals trigger stereotype inhibition: A proactive form of stereotype control. *Journal of Experimental Social Psychology*, *47*(1), 103–116.

Moskowitz, G. B., & Olcaysoy Okten, I. (2016). Spontaneous goal inference (SGI). *Social and Personality Psychology Compass*, *10*(1), 64–80.

Moskowitz, G. B., & Roman, R. J. (1992). Spontaneous trait inferences as self-generated primes: Implications for conscious social judgment. *Journal of Personality and Social Psychology*, *62*, 728–738.

Moskowitz, G. B., & Skurnik, I. W. (1999). Contrast effects as determined by the type of prime: Trait versus exemplar primes initiate processing strategies that differ in how accessible constructs are used. *Journal of Personality and Social Psychology*, *76*, 911–927.

Moskowitz, G. B., & Uleman, J. (1987, August). The facilitation and inhibition of spontaneous trait inferences at encoding. Poster presented at the 95th Annual Convention of the American Psychological Association, New York.

Moskowitz, G. B., & Vitriol, J. A. (2022). A social cognition model of bias reduction. In A. Nordstrom & W. Goodfriend (Eds.), *Innovative Stigma and Discrimination Reduction Programs* (pp. 1–39), Oxon, UK: Taylor and Francis.

Nader, K., Schafe, G., & Le Doux, J. (2000). Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature*, *406*, 722–726. 10.1038/35021052

Neuberg, S. L. (1989). The goal of forming accurate impressions during social interactions: Attenuating impact of negative expectancies. *Journal of Personality and Social Psychology*, *56*, 374–386.

Neuberg, S. L., & Fiske, S. T. (1987). Motivational influences on impression formation: Outcome dependency, accuracy-driven attention, and individuating processes. *Journal of Personality and Social Psychology*, *53*(3), 431–444.

Nordstrom, C. R., Hall, R.J., & Bartels, L. K. (1998). First impressions versus good impressions: The effect of self-regulation on interview evaluations. *The Journal of Psychology*, *132*(5), 477–491, 10.1080/00223989809599281

Olcaysoy Okten, I., & Moskowitz, G. B. (2020a). Spontaneous goal versus spontaneous trait inferences: How ideology shapes attributions and explanations. *European Journal of Social Psychoogy*, *50*, 177– 188. 10.1002/ejsp.2611

Olcaysoy Okten, I., & Moskowitz, G. B. (2020b). Easy to Make, Hard to Revise: Updating Spontaneous Trait Inferences in the Presence of Trait-Inconsistent Information. *Social Cognition*, *38*(6), 571–624.

Olcaysoy Okten, I., Schneid, E. D., & Moskowitz, G. B. (2019). On the updating of spontaneous impressions. *Journal of Personality and Social Psychology*, *117*(1), 1–25. 10.1037/pspa0000156

Park, B. (1989). Trait attributes as on-line organizers in person impressions. In J. Bassilli (Ed.), *On- line cognition in person perception* (pp. 39–60). Hillsdale: Lawrence Erlbaum Associates.

Park, B. (1986). A method for studying the development of impressions of real people. *Journal of Personality and Social Psychology*, *51*(5), 907–917.

Petty, R. E., & Brinol, P. (2010). Attitude change. In R. F. Baumeister , & E. J. Finkel (Eds.), *Advanced social psychology: The state of the science*, (pp. 217–259). Oxford University Press.

Petty, R. E., Briñol, P., & DeMarree, K. G. (2007). The Meta–Cognitive Model (MCM) of attitudes: Implications for attitude measurement, change, and strength. *Social Cognition*, *25*(5), 657–686.

Petty, R. E., Tormala, Z. L., Brinol, P., & Jarvis, W. B. G. (2006). Implicit ambivalence from attitude change: An exploration of the PAST model. *Journal of Personality and Social Psychology*, *90*(1), 21–41.

Peters, K. R., & Gawronski, B. (2011). Are we puppets on a string? Comparing the impact of contingency and validity on implicit and explicit evaluations. *Personality and Social Psychology Bulletin*, *37*(4), 557–569.

Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, 75, 811–832.

Quattrone, G. A. (1982). Overattribution and unit formation: When behavior engulfs the person. *Journal of Personality and Social Psychology*, *42*(4), 593–607.

Rapp, D. N., & Kendeou, P. (2007). Revising what readers know: Updating text representations during narrative comprehension. *Memory & Cognition*, *35*(8), 2019-2032.

Reeder, G. D., & Coovert, M. D. (1986). Revising an impression of morality. *Social Cognition*, *4*(1), 1–17.

Read, S. J., & Miller, L. C. (Eds.). (1998). *Connectionist models of socialreasoning and social behavior*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Rissman, J., Eliassen, J. C., & Blumstein, S. E. (2003). An event-related fMRI investigation of implicit semantic priming. *Journal of Cognitive Neuroscience*, *15*(8), 1160–1175.

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology*, *35*(9), 677–688.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L., Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10, pp. 173–220). New York: Academic Press.

Rothman, N. B., & Melwani, S. (2017). Feeling mixed, ambivalent, and in flux: The social functions of emotional complexity for leaders. *Academy of Management Review*, *42*(2), 259–282.

Rothbart, M., Evans, M., & Fulero, S. (1979). Recall for confirming events: Memory processes and the maintenance of social stereotypes. *Journal of Experimental Social Psychology*, *15*(4), 343–355.

Rydell, R. J., & McConnell, A. R. (2006). Understanding implicit and explicit evaluation change: A systems of reasoning analysis. *Journal of Personality and Social Psychology*, *91*(6), 995–1008.

Rydell, R. J., McConnell, A. R., Strain, L. M., Claypool, H. M., & Hugenberg, K. (2007). Implicit and explicit evaluations respond differently to increasing amounts of counterattitudinal information. *European Journal of Social Psychology*, *37*(5), 867–878.

Rydell, R. J., McConnell, A. R., Mackie, D. M., & Strain, L. M. (2006). Of two minds: Forming and changing valence-inconsistent implicit and explicit attitudes. *Psychological Science*, *17*(11), 954–958.

Sassenberg, K., Moskowitz, G. B., Fetterman, A., & Kessler, T. (2017). Priming creativity as a strategy to increase creative performance by facilitating the activation and use of remote associations. *Journal of Experimental Social Psychology*, *68*, 128–138.

Sassenberg, K., Winter, K., Becker, D., Ditrich, L., Scholl, A., & Moskowitz, G. B. (2022). Flexibility mindsets: Reducing biases that result from spontaneous processing. *European Review of Social Psychology*, *33*(1), 171–213.

Satpute, A. B., & Lieberman, M. D. (2006). Integrating automatic and controlled processes into neurocognitive models of social cognition. *Brain Research*, *1079*(1), 86–97.

Schneid, E. D., Carlston, D. E., & Skowronski, J. J. (2015). Spontaneous evaluative inferences and their relationship to spontaneous trait inferences. *Journal of Personality and Social Psychology*, *108*(5), 681–696.

Schneid, E. D., Crawford, M. T., Skowronski, J. J., Irwin, L. M., & Carlston, D. E. (2015). Thinking about other people: Spontaneous trait inferences and spontaneous evaluations. *Social Psychology*, *46*(1), 24–35.

Sherman, J. W., Lee, A. Y., Bessenoff, G. R., & Frost, L. A. (1998). Stereotype efficiency reconsidered: Encoding flexibility under cognitive load. *Journal of Personality and Social Psychology*, *75*(3), 589–606. 10.1037/0022-3514.75.3.589

Sherman, J. W., & Hamilton, D. L. (1994). On the formation of interitem associative links in person memory. *Journal of Experimental Social Psychology*, *30*(3), 203–217.

Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin*, *105*(1), 131–142.

Skowronski, J. J., & Carlston, D. E. (1992). Caught in the act: When impressions based on highly diagnostic behaviours are resistant to contradiction. *European Journal of Social Psychology*, *22*(5), 435–452.

Smith, E. R., & DeCoster, J. (2000). Dual process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, *4*, 108–131. 10.1073/pnas.93.24.13515

Srull, T. K. (1981). Person memory: Some tests of associative storage and retrieval models. *Journal of Experimental Psychology: Human Learning and Memory*, *7*(6), 440–463.

Srull, T. K., & Wyer, R. S., Jr. (1986). The role of chronic and temporary goals in social information processing. In R. M. Sorrentino, & E. T. Higgins (Eds.), *Handbook of motivation and cognition: Foundations of social behavior*, (pp. 503–549). Guilford Press.

Stangor, C., & Ruble, D. N. (1989). Strength of expectancies and memory for social information: What we remember depends on how much we know. *Journal of Experimental Social Psychology*, *25*(1), 18–35. 10.1016/0022-1031(89)90037-1

Stangor, C., & McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: A review of the social and social developmental literatures. *Psychological Bulletin*, *111*(1), 42–61.

Stewart, R. H. (1965). Effect of continuous responding on the order effect in personality impression formation. *Journal of Personality and Social Psychology*, *1*(2), 161–165.

Stone, J., Moskowitz, G. B., Zestcott, C., & Wolsiefer, K. (2020). Testing active learning workshops for reducing implicit stereotyping of Hispanics by majority and minority group medical students. *Stigma and Health*, *5*(1), 94–103. 10.1037/sah0000179

Stone, J., Whitehead, J., Schmader, T., & Focella, E. (2011). Thanks for asking: Self-affirming questions reduce backlash when stigmatized targets confront prejudice. *Journal of Experimental Social Psychology*, *47*(3), 589–598.

Tetlock, P. E. (1983). Accountability and the perseverance of first impressions. *Social Psychology Quarterly*, 285–292.

Todorov, A., & Uleman, J. S. (2002). Spontaneous trait inferences are bound to actors' faces: evidence from a false recognition paradigm. *Journal of Personality and Social Psychology*, *83*, 1051–1065. 10.1037/0022-3514.83.5.1051

Trope, Y. (1986). Identification and inferential processes in dispositional attribution. *Psychological Review*, *93*, 239–257.

Trope, Y., & Bassok, M. (1983). Information-gathering strategies in hypothesis-testing. *Journal of Experimental Social Psychology*, *19*(6), 560–576.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science*, *185*(4157), 1124–1131.

Uleman, J. S., & Moskowitz, G. B. (1994). Unintended effects of goals on unintended inferences. *Journal of Personality and Social Psychology*, *66*(3), 490–501. 10.1037/0022-3514.66.3.490

Uleman, J. S., Winborne, W. C., Winter, L., & Shechter, D. (1986). Personality differences in spontaneous personality inferences at encoding. *Journal of Personality and Social Psychology*, *51*(2), 396–403. 10.1037/0022-3514.51.2.396

Van Bavel, J. J., & Packer, D. J. (2021). *The power of us: Harnessing our shared identities to improve performance, increase cooperation, and promote social harmony*. Little, Brown Spark.

Van Dessel, P., Gawronski, B., & De Houwer, J. (2019). Does explaining social behavior require multiple memory systems? *Trends in Cognitive Science*, *23*, 368–369. 10.1016/j.tics.2019.02.001.

Vitriol, J., & Moskowitz, G. B. (2021). Reducing defensive responding to implicit bias feedback: On the role of perceived moral threat and efficacy to change. *Journal of Experimental Social Psychology*, *96*, 436–451. 10.1016/j.jesp.2021.104165

Webster, D. M., Richter, L., & Kruglanski, A. W. (1996). On leaping to conclusions when feeling tired: Mental fatigue effects on impressional primacy. *Journal of Experimental Social Psychology*, *32*(2), 181–195.

Wegener, D. T., & Petty, R. E. (1995). Flexible correction processes in social judgment: The role of naive theories in corrections for perceived bias. *Journal of Personality and Social Psychology*, 68, 36–51.

Wigboldus, D. H., Dijksterhuis, A., & van Knippenberg, A. (2003). When stereotypes get in the way: Stereotypes obstruct stereotype-inconsistent trait inferences.

Journal of Personality and Social Psychology, *84*(3), 470–484. 10.1037/0022-3514.84.3.470

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin*, *116*(1), 117–142. 10.1037/0033-2909.116.1.117

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review*, *107*(1), 101–126.

Winter, K., Scholl, A., & Sassenberg, K. (2021). A matter of flexibility: Changing outgroup attitudes through messages with negations. *Journal of Personality and Social Psychology*, *120*(4), 956–976. 10.1037/pspi0000305

Winter, L., & Uleman, J. S. (1984). When are social judgments made? Evidence for the spontaneousness of trait inferences. *Journal of Personality and Social Psychology*, *47*, 237–252. 10.1037/0022-3514.47.2.237

Wood, W., & Eagly, A. H. (1981). Stages in the analysis of persuasive messages: The role of causal attributions and message comprehension. *Journal of Personality and Social Psychology*, *40*(2), 246–259. 10.1037/0022-3514.40.2.246

Wood, W., Pool, G. J., Leck, K., & Purvis, D. (1996). Self definition, defensive processing, and influence: The normative impact of majority and minority groups. *Journal of Personality and Social Psychology*, *71*, 1181–1193.

Wyer, N. (2010). You never get a second chance to make a first (implicit) impression: The role of elaboration in the formation and revision of implicit impressions. *Social Cognition*, *28*, 1–19. 10.1521/soco.2010.28.1.1.

Wyer, N. A. (2016). Easier done than undone. by some of the people, some of the time: The role of elaboration in explicit and implicit group preferences. *Journal of Experimental Social Psychology*, *63*, 77–85.

Ybarra, O. (2001). When first impressions donat last: The role of isolation and adaptation processes in the revision of evaluative impressions. *Social Cognition*, *19*(5), 491–520.

Yzerbyt, V. Y., Rogier, A., & Fiske, S. T. (1998). Group entitativity and social attribution: On translating situational constraints into stereotypes. *Personality and Social Psychology Bulletin*, *24*(10), 1089–1103.

Zajonc, R. B. (1980) Feeling and thinking: Preferences need no inferences. *American Psychologist*, *35*, 151–175. 10.1037/0003-066X.35.2.151

19 Are We Stuck on the Face? New Evidence for When and How People Update Face-Based Implicit Impressions

Xi Shen[1] *and Melissa Ferguson*[2]
[1]*Cornell University*
[2]*Yale University*

People's faces are unique social stimuli that contain a lot of information. Some information is dynamic and transitory, such as emotional expressions, which reflect the person's momentary state. Some information is more static and remains relatively stable over time, such as facial structure. There is a long history of research studying how people perceive both types of facial information and how they play a role in impression formation. Without question, both types of facial information have a profound influence on forming impressions and social decisions. But in this chapter, we focus on the latter kind of information: information derived from the relatively static and stable features of faces. All the research we mention in this chapter uses static face pictures displaying neutral expressions. This kind of stimuli constrains the information included in the picture, and excludes many confounds such as clothes, makeup, and hairstyle. Unlike dynamic information that comes and goes as the context changes, the static features of faces, such as the shape of the face and the composition of features, are difficult to change and cannot be influenced by the context (we are not saying that the *inferences* made from these relatively stable facial structures cannot be influenced by context. Research has found that they can be; e.g., Brambilla et al., 2018).

What kinds of social information do people pick up from static faces? In addition to judgments such as attractiveness and babyfaceness, which are inherently judgments about facial appearances themselves, research has shown that people can infer many social attributes such as trustworthiness, competence, and aggressiveness from static faces with neutral expressions. Moreover, people can quickly form such first impressions and reach consensus based solely on static facial features. In the following sections, we will first describe the important role of faces in forming first impressions, followed by the difficulty of changing face-based implicit evaluations. Next, we move on to summarize the recent evidence for changing proposition-based implicit evaluations and how it can inform the research on changing face-based implicit impressions. We then summarize the current progress on changing face-based implicit evaluations and conclude with the next steps and future directions.

DOI: 10.4324/9781003045687-22

First Impressions from Faces and Why They Matter

Multiple Trait Inferences

People can infer a considerable amount of information, including various personality attributes, from faces with neutral expressions. The information can be gleaned with a brief exposure to a face (in as little as 40 ms; Bar et al., 2006; South Palomares, & Young, 2018; Todorov et al., 2005). People also reach a high consensus on the judgments they make (Todorov et al., 2009; Willis & Todorov, 2006), meaning people agree with each other on the information they infer based on facial appearances.

More importantly, not only do people form impressions based on facial appearances, these impressions influence social decisions and behaviors. Facial attractiveness is one of the main attributes that exert a significant influence. From earlier empirical studies to more recent meta-analyses, results consistently show that people attribute more socially desirable personality qualities to more versus less attractive people (Dion et al., 1972; Eagly et al., 1991). Attractive people are judged to be more trustworthy, sociable, competent, kind, interesting, and intelligent (Berscheid & Walster, 1974; Wilson & Eckel, 2006). Both women and men benefit from being attractive (Hosoda et al., 2003). Such halo effects based on attractiveness do not stop at the judgment level. Attractive people seem to also receive better treatment and experience better social interaction overall. For example, they can be given more attention, and have less negative and more positive interaction experiences (Langlois et al, 2000). In the specific yet consequential area of career prospects, attractive people enjoy advantages from the very start of the hiring process. This advantage holds regardless of the experience of the decision makers. Both amateurs (i.e., college students) and professionals (Hosoda et al., 2003) show bias in favor of attractive people. People seem to be inevitably influenced by the attractiveness glow.

Similar to attractiveness, people also associate certain traits with babyfaces. People who have a "babyface" are perceived to be warm, naive, submissive, and honest (Montepare & Zebrowitz-McArthur, 1989; Zebrowitz & Montepare, 1992). These associations with babyfaceness are reflected in real TV commercials where people with babyfaceness are more likely to be cast to deliver less expert but more trustworthy messages, the opposite pattern as their mature-faced counterparts (Brownlow & Zebrowitz, 1990). The impressions associated with babyfaceness also lead to social consequences. In job selection, babyfaceness influences the kind of job applicants get. People with babyfaceness are believed to fit better with a supporting position rather than a leadership position, which requires being dominant and exerting power when attractiveness and age are controlled (Zebrowitz et al., 1991). Even in making legal decisions, babyfaceness exerts an influence. Babyfaced people were less likely to be judged guilty if the case involved an intentional offense but more likely to be judged guilty if the case involved negligence

(Zebrowitz & McDonald, 1991). Interestingly, some research suggests that babyfaced people (in particular male adolescents and young adults) are at times more likely than their mature-faced peers to show aggression (Zebrowitz, Collins & Dutta, 1998) and exhibit high achievement motivation and behavior (Zebrowitz et al., 1998). One explanation for this effect is that babyfaced people might be trying to compensate for the negative stereotypes projected on them about leadership and competence (e.g., Andreoletti et al., 2001). Another potentially alternative (or, additional) explanation hinges on the testosterone-driven link between broader, rounder faces (a feature of babyfaceness and also the facial width to height ratio) and aggression (Zebrowitz et al., 2015).

Single Trait Inferences

Beyond facial attributes that are associated with multiple trait inferences, a lot of research has examined the role of single trait impressions formed from faces. Trustworthiness and competence are two of the most extensively studied social traits that people infer from faces instantly and that lead to significant downstream consequences.

Facial trustworthiness alone predicts a diverse range of social judgments. Studies show that people favor more trustworthy-looking financial service providers as their potential advisors to entrust their money with (Dean, 2017). Also, in economic games, people invest more in those who have more trustworthy-looking faces (Chang et al., 2010; Rezlescu et al., 2012; Stirrat & Perrett, 2010; Van't Wout & Sanfey, 2008). These results also show ecological validity. The same pattern of results was found in actual online financial websites where more trustworthy-looking borrowers are more likely to receive loans from lenders (Duarte et al., 2012). People with a more trustworthy face are even more likely to be appointed as CEO after firm misconduct. Such appointments are more likely to be received positively by stakeholders (Gomulya et al., 2017). Even in judicial judgments where facial appearances should play no role, people who look more trustworthy enjoy an advantage as they are less likely to be judged as guilty compared with those who do not look trustworthy (Dumas & Teste, 2006; Porter et al., 2010).

Looking competent seems to matter especially in the political arena in some countries, including the United States, France, Mexico, and Brazil (Olivola & Todorov, 2010). Judgments of facial competence predicted actual U.S. congressional elections (Todorov et al., 2005). Also, similar results have been found cross-culturally, replicating the role of faces in predicting election results in different countries, although warmth and trustworthiness seems to matter more than competence in some countries, such as Japan (Lawson & Lenz, 2007; Poutvaara et al., 2009). These results have been found even when the participants judging the candidates' faces are from another country than the candidates, thus ruling out the possibility of familiarity as an explanation for these findings (Antonakis & Dalgas, 2009). Strikingly,

judgments made with as little as 100 ms exposure to the face predicted election results and were not different from judgments made with 250 ms or with no time constraints (Ballew & Todorov, 2007), suggesting the robustness of the influence from competent-looking faces. Beyond political outcomes, facial competence also predicts how people judge scientists' work. People showed more interest in competent-looking scientist's work and considered their work to have higher quality (Gheorghiu et al., 2017). Last but not least, facial competence affects moral judgments. Specifically, the level of competence in the face predicted the perceived acceptability of social exclusion (Rudert et al., 2017).

Irresistible Influences of Faces

All of the previous research findings suggest that the influence of faces has permeated almost every aspect of people's lives. Faces not only invoke evaluative judgments and trait inferences, they also influence social judgments and consequences where faces are not supposed to play a role. For example, in legal settings, how someone looks should have no bearing on sentencing results. Conviction should be based on evidence and facts. In electing officials, voters are supposed to vote for candidates' abilities or political ideologies. In hiring decisions, the candidate's working experience and abilities should be the most relevant information for judging whether the person will excel in the job. Yet, all of the research we mentioned so far showed the unexpected role of faces in all of these decisions. Some may argue that it is because faces are the only (or the main) focus in this research. Participants are not given opportunities to consider other types of information or more relevant information. If people are provided with other information that is clearly more relevant and reliable to the decision, the face influence will disappear.

However, research findings have shown that this account is unlikely to be the primary explanation. Even when propositional information (e.g., behavioral information), which is considered to be more relevant and reliable than face information, is available, facial appearances still exert an impact on decision-making. For example, Rudoy and Paller (2009) asked participants to first memorize face-trait pairs before rating the people varying on facial trustworthiness. They found that judgments of novel targets' trustworthiness were influenced by facial cues even when trait information was provided and learned by the participants beforehand. The influence of facial trustworthiness showed an even bigger influence when the judgments were made under time pressure. Persistent influence from faces was also found in sexual orientation judgments (Rule et al., 2014). In the studies, participants were given the actual sexual orientation of each target face. However, knowledge about the novel targets did not override the influence of faces. Participants persistently relied on face information in judging sexual orientations, especially under time pressure. In addition, the inescapable influence of the face

affected how people choose trustworthy partners in trust games where the participants' goal is to maximize their own gains. From a rational standpoint, facial appearance does not help to achieve this goal. Nonetheless, how trustworthy a person looked not only influenced how much participants trusted them when no other information was available but also when potential candidates' past behavior histories were presented (Rezlescu et al., 2012). Similarly, participants were influenced by novel targets' facial appearances when payoff information, the most relevant information for maximizing reward in a trust game, was provided (Jaeger et al., 2019a; see also Pandey & Zayas, 2021). Some of the most compelling evidence comes from studying how people make decisions in mock jury cases. Even when people were given evidence for legal cases, faces still played a role by biasing how people used the evidence to reach a verdict and how confident people felt about their decision. Untrustworthy-looking defendants were in a disadvantaged position such that they were more likely to get a guilty verdict with less evidence. Meanwhile, participants felt subjectively more confident in their decision about the case when the defendant looked more untrustworthy (Porter et al., 2010).

Why Face-Based Impressions Are So Persistent and Difficult to Overcome

Why is it so difficult to eliminate the influence of faces, even for decisions with life-death consequences, and when more relevant propositional information is available and should be used from a rational perspective? Two factors might be at play. First, face processing is found to be intuitive and easy. With a simple glance, people can form first impressions from faces. It is striking how an extremely short amount of time is enough for people to make judgments from faces. Willis and Todorov (2006) found that as little as 100 ms is enough for people to make judgments, such as attractiveness, trustworthiness, and aggressiveness from static facial pictures. These rapid first impressions formed from faces were not different from judgments made with no time constraints. Other studies found even shorter amounts of exposure, as little as 40 ms, was sufficient for people to judge whether the person is threatening or not (Bar et al., 2006). Some research extends it further to show that face-based evaluations do not require intention or subjective awareness of the face stimulus (Todorov et al., 2009; Shen et al., 2020). These findings demonstrate the efficiency of face processing, implying that the face of an individual can exert an immediate influence on people, perhaps even before people realize it. This processing ease for face stimuli presumably leaves people with little ability to resist its impact. In particular, the intuitive judgments based on faces might contrast with the more effortful process associated with word processing. This difference in processing ease could lead to the overpowering influence of faces compared with language-based propositional information when both types of information are available. Some studies have examined this hypothesis by

manipulating participants' response time. In one line of work, facial trustworthiness and behavioral valence were manipulated. Participants were asked to learn face-behavior pairs before evaluating the faces. Although evaluations of the faces were influenced by the valence of the previously paired behavioral descriptions when participants had no time constraint in evaluating the face, this effect diminished when time pressure was imposed on making a response (Verosky et al., 2018). In another line of work, participants played trust games where they needed to decide whether to trust their partners. Participants' goal was to maximize their earnings. In these games, researchers provided potential partners' face pictures and payoff information for different decisions. Rationally, the payoff information should be more informative for making decisions. However, people consistently relied on the face information. More pertinent here, participants relied more on faces when they were asked to make intuitive decisions under time constraints than when they had ample time to make reflective decisions (Jaeger et al., 2019a). These results corroborate the hypothesis that reading from faces is intuitive and efficient, explaining why face information seems almost irresistible when making judgments.

Second, people might be unaware that seeing a face biases their evaluations and how they process subsequent propositional information. There is not enough research on this topic. Most research focuses on what people *can* get from faces when prompted. It is reasonable to speculate that people may not be fully aware of how much information they get from faces. In other words, when people are not prompted to intentionally evaluate others, it might not consciously occur to them that they are evaluating others based on faces. For this reason, faces might exert a greater influence relative to other forms of information particularly when those other forms of information are seen by people are obviously informative and diagnostic, such as information related to a defendant's behavior in a criminal case, or someone's actual trustworthy history in economic games. In such circumstances, faces might play an especially prominent role relative to the other information because people are relatively less aware of the face influence and so do nothing to combat it.

In addition, people probably do not know how exactly their judgments are influenced by the evaluations and inferences they make from faces (see Wilson & Brekke, 1994). Specifically, when both face and propositional information are present, given the prominence of the face and the ease of making inferences from faces, people may anchor on the face information, thus biasing the judgments of other information and the way they seek or use other information. Even if people can report how much they believe in the accuracy of face perceptions and know they judge people based on faces, it is likely that they do not know how they should map their beliefs onto actual decisions based on the strength of their beliefs. These possibilities might create difficulty for people in terms of correcting their judgments even when they are motivated to do so.

Influence of Faces in Implicit Impressions

In addition to explicit evaluations and judgments, faces have been argued to have an especially strong effect on implicit impressions. Traditional dual-system views have proposed that explicit and implicit evaluations are enabled by two separate systems. Explicit evaluations are thought to be generated by a fast-learning system where information can be learned and updated rapidly and intentionally. This system is sometimes assumed to be especially sensitive to propositional information, such as language. Implicit evaluations, in contrast, are assumed to result from a slow-learning system, where learning and updating happen much more slowly and unintentionally. This system is sometimes assumed to be more sensitive to visual cues such as faces. According to this view, explicit evaluations are sensitive to propositional information and change quickly in response to new and countervailing information. On the contrary, implicit evaluations, once formed, cannot be changed easily in response to new countervailing information. Even if such evaluations can be changed at all, they have been argued to change at a slower rate than explicit evaluations. Some studies show results that are consistent with this dual-process approach (e.g., Cao & Banaji, 2016; Gregg et al., 2006; Rydell & McConnell, 2006; Rydell et al., 2007). For example, Rydell and McConnell (2006) found that after learning counterattitudinal information, explicit evaluations toward a novel target changed dramatically and quickly. In contrast, implicit evaluations changed slowly. Only after learning a great number of pieces of counterattitudinal information (i.e., 100 pieces) did the (implicit) evaluations start to become inconsistent with initial evaluations. Notably, one limitation in this area of research is that almost all studies focused on propositional learning (i.e., behavioral information) as the way of forming and updating impressions. Facial appearances, as much as they have been studied with explicit measures about their role in impression formations, have not been explored much in implicit impressions. It is not clear whether people would form implicit impressions based on faces and how easy it would be to change initial implicit evaluations based on the face. One early and elegant paper that examined these topics was by McConnell et al. (2008). They studied implicit and explicit impressions formed based on facial cues such as attractiveness and obesity (vs. thinness). They found results consistent with dual-mode models such that explicit evaluations formed based on faces were easily changed by learning additional countervailing behavioral information. However, implicit evaluations formed based on facial cues were much harder to change (if at all). For example, in the study where facial obesity was examined, participants initially formed more negative impressions toward the novel target with the obese face than with the thin face. More importantly, extensive learning of countervailing behavioral information about the targets only changed implicit evaluations toward thin faces, but not obese faces, suggesting that the negativity from the cue of obesity was persistent in people's implicit responses to the target.

More recent research, however, has started to provide more evidence about when and how implicit evaluations might be updated. Studies have shown that implicit evaluations, like explicit evaluations, can be rapidly changed in some circumstances. Several conditions have been identified under which implicit evaluations can be quickly updated. Although most of the research is about impressions based on propositional learning, can the same principles be applied to changing face-based implicit impressions? Will face-based implicit evaluations show different characteristics than proposition-based implicit evaluations in terms of updating? As we mentioned previously, face-based first implicit impressions may be harder to change given the difficulty in erasing the effect of faces on social decisions, the intuitive availability of faces, and the unawareness of utilizing facial information in making decisions. In the following sections, we review the properties of new evidence that have been identified as critical for changing proposition-based implicit evaluations, and connect the research to the most recent research on changing face-based implicit evaluations.

Diagnosticity

Behavioral information has long been known as one of the main sources of evidence that people use to form impressions of others. People can quickly form impressions based on behaviors, spontaneously infer traits from behavioral descriptions, and readily associate any inferred traits with the respective actor (Newman & Uleman, 1993; Todorov & Olson, 2008; Todorov & Uleman, 2002, Todorov & Uleman, 2003; Uleman & Moskowitz, 1994). However, not all behaviors show the same effect on impressions. People show a negativity bias toward behaviors in the moral domain and a positivity bias toward the competence domain (Skowronski & Carlston, 1987,1989; Wojciszke et al., 1993) such that negative, immoral behaviors are weighted more than good, moral behaviors. In contrast, positive, competent behaviors are weighted more than incompetent behaviors. Although on the surface it may seem as though people can be therefore biased against both negative and positive information, explanations about these asymmetries all point to the critical role of evidence diagnosticity (Reeder & Brewer, 1979; Skowronski & Carlston, 1987; Skowronski & Carlston, 1989). Immoral behaviors and competent behaviors are assumed to be able to more strongly reveal the person's true character in their respective domain. This is what we mean by diagnosticity—behaviors that are assumed to be more likely to reveal a person's character and can be used to predict future behaviors. One common way to show diagnosticity is with extreme behaviors. By definition, behaviors that are extreme are less common and not shown by most people, and are therefore more likely to be attributed to the target's disposition or personality (e.g., see Fiske, 1980). This means that extreme behaviors are especially likely to be seen as demonstrating diagnosticity.

Consistent with this classic person perception work, more recent research on the updating of implicit evaluations with propositional learning has also

pointed to diagnosticity as one key factor for rapid change. In the work by Cone and Ferguson (2015), for example, participants first learned a novel positive target "Bob" through 100 behavioral statements. When a single negative diagnostic behavior was later provided about Bob ("Bob was convicted for mutilating a small, defenseless animal"), people quickly changed their initial implicit evaluations to negative. Notably, the updated implicit impressions about Bob predicted how people intended to interact with Bob (see also Mann et al., 2019). Other work has shown that one piece of diagnostic information can also quickly change implicit evaluations toward known targets with long-established evaluations. Van Dessel et al. (2019) found that well-established implicit positive impressions toward Gandhi quickly changed to negative by learning new negative diagnostic information. These studies established the effectiveness of diagnostic information in changing implicit first impressions based on propositional learning.

Is diagnostic propositional information also effective in changing implicit first impressions based on faces? If facial information is associative in nature as previously suggested, which has been argued to be processed by the implicit, slow-learning system, while propositional information will be processed by the explicit, fast-learning system, it is possible that implicit first impressions based on the face simply might not respond to new propositional information because the two types of information are processed by separate systems. However, this prediction was challenged in recent studies by Shen et al. (2020), in which they examined whether and when implicit first impressions based on facial trustworthiness are updated. These findings showed people's implicit evaluations of a target can change dramatically after learning new diagnostic, countervailing propositional information. Across studies, participants initially formed negative implicit impressions toward a novel target "Joe" by viewing his untrustworthy face paired with neutral information (e.g., "Joe has a lamp in his room"). After learning extremely trustworthy information about Joe (Joe was described as someone who kept his promise to take care of his neighbor's house while they were away. He even did extra work to keep the house nice and clean without complaining. When he heard there was a hurricane coming, he ended his own vacation and returned to stormproof his neighbor's house to make sure it wouldn't be damaged by the storm.), participants' showed reliably positive implicit impressions, indicating that they were now evaluating him in a much different manner. In the study where the strength of facial trustworthiness was manipulated, this reversal in how people implicitly evaluated the target occurred even for the target with the most untrustworthy looking face, demonstrating the effectiveness of changing face-based implicit evaluation with sufficiently diagnostic new information. Importantly, they also showed that information diagnosticity is key. In the study where they manipulated the diagnosticity of the new information, they found implicit evaluation did not reverse when the new

countervailing information was not extreme. It was only when the information was highly diagnostic did the flip of implicit evaluations occur. In another project where the changeability of implicit evaluations based on facial attractiveness was examined, similar results were found (Cone, Meagher, Mann, & Ferguson, 2021). Participants first formed strong positive implicit impressions toward the more attractive looking target than the unattractive looking one. However, the initial positivity toward the attractive target was quickly changed to strong negativity after participants learned that the target drowned her kids in the bathtub, a piece of highly diagnostic and negative information.

This work with facial trustworthiness and attractiveness suggests that impressions formed based on the face can be successfully flipped even implicitly. When the new information is diagnostic enough, people change how they implicitly evaluate the target in both directions (from positive to negative and vice versa). These results also suggest that the format of information does not seem to have a special influence on implicit evaluations as previous research suggests. Rather, the more important determinant of changing impressions is the strength of the evidence. When the new evidence is powerful enough, it can override initial impressions regardless of the format of that initial information.

Reliability

Does all diagnostic information lead to successful updating? When it comes to the effectiveness of changing people's minds, the reliability (or perceived truth) of the information has been found to be one critical factor in addition to diagnosticity. In the persuasion literature, more believable information is more effective in changing people's views (see Brinol & Petty for a review, 2009; Skowronski & Carlston, 1987, 1989; see exception: Tormala et al., 2006). However, all this research used explicit measures to capture people's opinions. Would information reliability matter for implicit evaluations as well? According to most dual-mode process theories, the implicit system is an associative system. This means that whether the information is true or false should not influence implicit evaluations. From this view, the explicit system instead operates on propositions, and thus it responds to the truth value of the information (e.g., see Peters & Gawronski, 2011). Based on most dual-mode perspectives, then, implicit evaluations should not be sensitive to information about reliability, which would predict that after people learn a piece of diagnostic information, implicit evaluations will change even if the information is not reliable at all. Only explicit evaluations will reflect the reliability of the information. Although some work shows that the formation of implicit impressions in fact seems reliant on the perceived reliability of the information (Moran et al., 2017; Smith et al., 2013), other work suggests that this only happens when there is no other information available about the stimulus. After an impression about someone or something has been formed, the reliability of

new (counterattitudinal) information does not seem to influence the implicit impressions (Cao & Banaji, 2016; Hu et al., 2017; Peters & Gawronski, 2011), which is generally in line with the dual-mode perspective.

However, given the work we just reviewed on how implicit impressions are sensitive to the diagnosticity of propositional information, from our perspective, implicit impressions should be sensitive to the believability of diagnostic information. This is consistent with recent work suggesting that instead of implicit impressions being driven by simple co-occurrences between stimuli in the environment, they seem sensitive to the relations between stimuli, such as those indicating causality (see De Houwer, 2018; De Houwer & Hughes, 2016; Kurdi & Dunham, 2020; Mandelbaum, 2016; Van Dessel, Hughes, & De Houwer, 2019). The prediction that implicit impressions also respond to information reliability was more directly tested in a recent paper by Cone and colleagues (Cone et al., 2019). They found that the perceived reliability of the new information matters for how much people changed their initial implicit evaluations toward a novel target. Initial positive implicit evaluations successfully updated when people learned a highly reliable piece of information (from police reports) but not when people learned a questionable piece of information (rumors). This research also showed that the reliability of the information tracks the extent to which people correct their initial implicit impressions. However, these findings were all based on propositional learning. Would updating face-based implicit impressions also be subjected to the same principle?

Information reliability may matter much less or not matter at all for changing face-based implicit evaluations. In addition to the two reasons we raised in the earlier sections concerning the intuitiveness of facial inferences and the potential lack of awareness of using face information for subsequent judgments, there is a third reason that might point to faces being immune to propositional information's reliability. This third reason is uniquely relevant to face-based implicit impressions: the subjectivity of the face reliability. This could make it much harder to predict the influence of the propositional information's reliability. When impressions are entirely formed based on propositional information, the reliability of the information is clearly cued in the description and can be easily compared between prior and new learnings. But how might people compare the reliability of the face information with newer propositional information? First, it is unclear how people perceive the reliability of the inferences made from faces. Although there is research examining people's beliefs about the reliability of information read from faces (Jaeger et al., 2019b; Suzuki et al., 2019), these reports are collected via explicit measures, which are subjected to social norms and self-presentation pressures. It is possible that self-reported explicit beliefs in facial reliability may not accurately reflect how reliable people think face-inferred information is, let alone reflect the relationship between such beliefs and implicit evaluations based on the face. Our own findings showed that people's explicit beliefs do not predict face-based implicit impressions (Shen & Ferguson, in prep).

Besides the methodological issues with explicit measures, another issue also complicates the interpretation of the results from explicit scales. For explicit reports to be valid and accurate, people need to have access to their insights. For example, they need to know how much information they read from faces and to what extent they are using the information to make evaluations (Wilson & Brekke, 1994). However, it is likely that people have limited capacity to know and report this information, or even more, people do not realize their evaluations are influenced by face information and to what extent. Just as with many types of sources of information (e.g., group membership, situational context, etc.), people may remain totally unaware of the inferences they are making from faces. Thus, how can people adjust evaluations based on information reliability? Third, even if perceived face reliability can be accurately measured by explicit questions, it is an open question of how people compare the reliability of the face with the reliability of propositional information. Can people intentionally calculate and compare the two, or is there an intuitive rule governing this comparison process?

Despite all the potential reasons why face-based implicit evaluations might not be sensitive to the reliability of propositional information for updating, it might still be the case that people adjust the way they implicitly evaluate a target based on the reliability of new counterattitudinal evidence. Even if they cannot manage the calibrations and comparisons we mentioned previously, in clear-cut cases where new counter-evidence is completely believable versus patently false, people may be able to shift their implicit impression accordingly, whether it is incorporating reliable information or dismissing unreliable information. Because of the prior work showing that implicit impressions based on the face are sensitive to the perceived diagnosticity of new evidence (and that people seem to disregard new evidence that does not seem to reflect something useful and predictive about the target), we predicted that people would similarly incorporate new evidence into their implicit impressions based on its reliability. In Shen and Ferguson (2021), they manipulated facial trustworthiness and had participants first form implicit first impressions solely on the face. Not surprisingly, people initially formed positive impressions toward a trustworthy-looking target ("Joe"). Then the experimenters gave participants one piece of new diagnostic negative information about the target (e.g., Joe molested a child) but varied its reliability. It was only when the information was highly reliable (Joe admitted in court that he molested a child) did initial implicit evaluations reverse from positive to negative, but not when the new information was low on reliability (the same diagnostic information was spread by Joe's competitor who lost a competition to him and benefited after spreading this information about Joe) or moderate on reliability (the information was spread by a person who heard it from a female co-worker who was mad at Joe dumping her but said she was sure about the information about Joe). These findings suggest that diagnosticity is not the only evidence dimension that matters. Reliability is also influential in flipping how people implicitly

evaluate targets when both face information and propositional evidence are available. In addition, the findings that the initial impressions became slightly more negative (although did not reverse) when the new diagnostic negative information was moderately reliable but did not change at all when the new information was low on reliability illustrated that reliability can be nicely tracked by implicit measures, at least when the information was highly diagnostic. This line of research again demonstrates that faces do not hold a special "invulnerable" status in implicit evaluations. Face-based implicit impressions are responsive to information reliability, a property that is important for proposition-based implicit impressions.

Reinterpretation

In the previous two sections, we described work showing that the properties of diagnosticity and reliability of new evidence matter for whether people change their implicit mind about someone. In this work, however, the new information had no direct (referential) connection to the prior information. Here we consider a different type of learning in which the new counter-attitudinal information about someone is closely related to the initial learning such that the new information can reinterpret the information that was learned earlier.

In research by Mann and Ferguson (2015; see also Mann & Ferguson, 2017), participants learned about a novel target "Francis West" who allegedly broke into a neighbor's house, damaging their homes and taking "precious things" from the homes. After learning about Francis, participants first formed negative implicit impressions about him. Then, the researchers introduced information explaining that the reason Francis broke into the house was because he saw the house was on fire, and the "precious things" he took from the house were the kids trapped inside. This new information reinterprets the prior ostensibly rampaging behaviors. Those behaviors were actually carried out due to the motivation to save the kids from the fire. After participants learned the new information, they implicitly evaluated Francis in a strongly positive manner.

Can we borrow this tool for changing proposition-based implicit impressions and apply it to changing face-based implicit impressions? Mann and Ferguson (in prep) used a similar story from the Francis West studies and used it to examine the changeability of implicit impressions from facial scars. Not surprisingly, when faces were the only information available, people formed more negative initial impressions toward a novel target with prominent facial scars than the target with no facial scars. However, after people learned that the scar was caused by the target's heroic act trying to save the kids from fire, they showed strong positive implicit impressions toward the scarred target. This new and ongoing line of work examines whether and how the process of reinterpretation works for implicit impressions when initial information is visually based (faces) rather than propositional.

The Persistence of Updated Face-Based Implicit Impressions

Another claim from dual-mode theories that argue for an implicit, associative system is that implicit evaluations are rigid. It suggests that implicit evaluations are not just difficult to change; even if they are changed, the change will not last but quickly revert back to their baseline. In general, there is limited research on this topic. Among the papers that have examined the stability of implicit evaluations based on propositional information, some results are consistent with the dual-mode models such that interventions only changed implicit evaluations temporarily and returned to their initial level after hours or days (Lai et al., 2016). But others found different results. For example, using the same reinterpretation method we introduced before, updated implicit evaluations persisted after a two-day delay (Cone et al., 2021; Mann & Ferguson, 2017). There may be many procedural differences between these lines of work that explain why updating evaporated in some studies but sustained in others. For example, whereas the counterattitudinal evidence provided in the studies showing durability was believable and diagnostic, and was often tied to one individual, it often was not so in the other studies showing evaporation, and this might explain some of the divergence.

But would updated face-based implicit evaluations sustain over time? Compared with proposition-based implicit impressions, face information is more intuitively available than propositional information, which needs to be retrieved from memory. So would face information dominate the implicit evaluations examined after a delay? In a recent study, Shen and Ferguson (2021) manipulated facial trustworthiness and the reliability of additional, new negative information. Immediately after learning the diagnostic and negative information, implicit evaluations changed, tracking the level of reliability as we described before. Critically, implicit evaluations toward the same target sustained after a three-day delay, even with trustworthy-looking targets. This is initial, suggestive evidence that updating of face-based implicit impressions might be durable at least in some circumstances.

Lessons from Research on Changing Face-Based Implicit Evaluations

Faces Are Not a "Special" Type of Information

Although some models theorize faces as a type of associative information that has an especially strong influence on implicit evaluations, relative to propositional information (e.g., McConnell et al., 2008), more recent research finds that face-based implicit evaluations can be quickly reversed just like proposition-based implicit evaluations. The same evidence properties that matter for changing proposition-based implicit evaluations—diagnosticity, reliability, and reinterpretability—also apply to face-based implicit evaluations,

suggesting the format of information is not key in determining whether and to what extent implicit evaluations can be changed. What does this new evidence show? We think it shows that faces do not constitute a special "kind" of information that is encapsulated from propositional reasoning. This is important to know because it has implications for the structure of the mind, and directly addresses various kinds of dual-mode theories. And yet, it still might be true that faces (vs. propositional information) on average exert a stronger influence on implicit and explicit judgments, given all of the reasons we have outlined previously. But by our account, though, we can examine (and measure and manipulate) how the properties (perceived and actual) of evidence from faces versus other sources influence impression formation and updating.

Despite the research suggesting the irresistible influence of faces, evaluations anchored on the face can actually be reversed completely by other information. The initial positivity or negativity associated with certain facial features against others can (at least in some cases) be overridden by new information. These results demonstrate that biases based on facial cues are not insurmountable. This work raises the possibility for potentially identifying ways to reduce the biasing influence of faces in decisions.

However, if faces are not a special type of information, and implicit evaluations based on faces can be quickly changed (in some cases), then what might explain the previous research findings that explicit evaluations can be much more easily and quickly changed while implicit evaluations cannot? Although this question is beyond the scope of this chapter, some work shows that different implicit measures show differential sensitivity to change (Van Dessel et al., 2019). And future work might examine whether at least some of the differences lie in the context of implicit and explicit measures and their different measurement features (see Ferguson et al., 2019; Gawronski, Chapter 11, this volume).

How Can One Escape Face Bias?

Although we have reviewed evidence suggesting that implicit impressions are not inevitably tied to the face, it is important to note that the kinds of counterattitudinal information needed to change those impressions were extreme. That is, yes, it is true that the initial negative implicit impression people form of a person with an untrustworthy-looking face can be overturned, these new findings suggest that the person needs to show near saintly behavior (that is also believable!) in order for that to happen. This is troubling, especially considering that the accuracy of face-based impressions (including trustworthiness and competence) is highly controversial (see Todorov et al., 2015). Do people have to go above and beyond in order to escape the initial biases their faces introduce?

One necessary area of future work, on this point, will be to expand beyond these one-shot scenarios where people are learning about hypothetical others usually in single experimental sessions. In these kinds of scenarios, at least so

far, it does seem as though a target must truly be heroic in order to escape an untrustworthy-looking face, for instance, but it remains to be seen whether this is only for those we do not know or care much about. The findings on face judgments and electoral success suggest that even considerably more information does not erase the (dis)advantages from faces, but even this work deals with situations in which we may not personally know or care about the target (i.e., a politician). It could be that a real person who we meet, with whom we have some kind of shared circumstances (a co-worker, a neighbor, a friend) who happens to have an untrustworthy or incompetent-looking face, is freed from that face bias with only minimal counterattitudinal evidence and it will be important to examine and better understand the types of evidence needed for escaping face bias.

The Properties of Evidence Matter

Not all types of information effectively lead to the updating of face-based implicit evaluations. Recent work found that the same information qualities that enable rapid change of proposition-based implicit evaluation—diagnostic, reliable, or reinterprets prior information—are also effective for changing face-based implicit first impressions. These findings point to the critical underlying mechanism for changing impressions: the new information needs to be more revealing of the person's character and can be used to predict their behaviors, regardless of the type of information, whether it is visual or language-based. On some level, the new information needs to let people disregard the old information and believe the new information is the new basis for getting to know the "true self" of the other person.

However, there are remaining questions unique to face-based implicit impressions. As we mentioned before, when impressions are all formed based on propositions, the diagnosticity and reliability of the old and new information can be easily compared because they are clearly cued in the (linguistic) stimuli. When the first impression is based on a face, however, the comparison of evidence strength and reliability is much less clear. Given that first implicit impressions based on faces can be updated by extreme but not moderate new information, we can probably infer that face inferences are not extremely strong evidence, but also not weak. No research has directly measured the implicit beliefs of the diagnosticity or the reliability of the information inferred from faces or examined how they are comparable to propositional information. On the one hand, the diagnosticity of information inferred from faces can be objectively manipulated by the strength of facial cues. The stronger the facial cues, the stronger the inferences made from faces. For example, the more trustworthy the faces are manipulated to be on the trustworthiness continuum, the more people infer a higher level of trustworthiness from the person.

On the other hand, it is possible that the diagnosticity of the face information relies heavily on the perceiver and interacts with people's

subjective value about the reliability of face information. When judging the same face, those who strongly believe faces reveal true characters might infer a stronger attribute than those who believe less in the face's kernel of truth. Also, those who do not (or weakly) believe in the kernel of truth may perceive the objectively same amount of difference in diagnosticity between faces manipulated on a dimension to a lesser degree compared with those who believe in the kernel of truth. Given that the reliability of face information is another subjective judgment, determining the diagnosticity of face inferences and being able to compare that with propositional information quickly becomes more complicated than when dealing with propositional information alone. However, these are important questions to examine. They cannot only teach us about the nature of face inferences but also more directly inform us about the possible mechanisms that drive updating implicit impressions.

Reinterpretation is slightly different from diagnosticity and reliability in that it emphasizes the relationship between the new and the old information, and more specifically, denotes the relevance of the new information to prior knowledge. We mentioned studies that used reinterpretation to change first impressions based on facial scars, a facial feature acquired later in life. However, to what extent can reinterpretation change evaluations based on facial features that people are born with? Many facial features are innate, like attractiveness and trustworthiness, and are the result of people's natural facial structure. Everyone probably agrees that how someone's facial structure looks is uncontrollable and unintentional (at least for most people). Unlike behaviors, many of which are intended by the person given various motivations, there would seem to be simply less room for reinterpretation of face structures. Thus, it remains unknown whether and to what extent reinterpretation can be applied to different types of facial attributes.

Changing Impressions about Individuals to Groups

All of the research we mentioned so far is about updating impressions about individuals, but can these lessons about updating toward individuals generalize to groups? For example, when participants' positive implicit evaluations toward a trustworthy looking person are updated to be negative after learning new information, does this influence how participants evaluate other people with trustworthy faces? On the one hand, some work found that learning about faces can be generalized to perceptually similar faces (Verosky & Todorov, 2010). Across several studies, researchers paired faces with three pieces of negative, positive, or neutral information. Next, researchers morphed the learned faces with novel faces to create new ones. They found although people did not recognize the old face components and categorized the newly created hybrid faces as novel, their evaluations of the new faces were influenced by the evaluations associated with the old face, so if the new face contains the face component of an old face that was initially paired with

negative information, the new face was evaluated as more negative (and vice versa for old faces paired with positive information). Although this work used only explicit evaluations, it suggests the potential of evaluative generalization between faces sharing similar perceptual features.

On the other hand, classic research on stereotyping has found that when people encounter members that violate the expectations for a group, instead of changing stereotypes about the group, people may subtype the specific individuals as a subordinate category, thus leaving the original representations about the group intact (Weber & Crocker, 1983). However, there are situations in which impressions about the group as a whole can generalize to individual group members, even implicitly (Hamilton, 2015), and future work can examine how these lessons extend to the updating of implicit impressions based on faces.

In addition, compared with impressions based on other types of groups, such as gender and race, which are formed through long and rich learning, impressions based on faces might be less well established with regard to the amount of learning. Some theories argue that inferences made from faces are due to the generalization of emotional expressions (Montepare & Dobish, 2003; Oosterhof & Todorov, 2009), suggesting the automatic and natural tendency with this phenomenon. The implication of these theories for changing face-based group impressions is twofold. First, it might be especially hard to overcome group facial cues if they are developed based on emotion recognition, which has been argued to be an innate ability (Izard, 1994; Matsumoto & Willingham, 2009). Second, it might be easier to challenge the foundation of inferring information from categorizing facial cues in the first place if people are made aware that it is a cognitive bias that is not completely about facial structures.

Considering all the possibilities mentioned previously, it is up to empirical research to find out whether and how changes made about an individual's face impressions generalize to the group who share similar facial features.

Conclusion

We reviewed extensive work suggesting that first impressions based on faces are hard to overcome, including implicitly. However, we summarized three lines of work that have identified key properties of evidence—diagnosticity, reliability, and reinterpretation—that lead to rapid and durable updating of implicit evaluations. Although most of this work has focused on learning based on propositional information, we highlighted emerging work that applies these lessons to implicit impressions based on faces. We believe the work on changing proposition-based implicit evaluations can shed light on changing face-based implicit evaluations. We identified key areas for future research on this topic, including identifying the circumstances in which people can move beyond face-based impressions to incorporate other, more relevant, kinds of evidence in their impressions of others.

References

Andreoletti, C., Zebrowitz, L. A., & Lachman, M. E. (2001). Physical appearance and control beliefs in young, middle-aged, and older adults. *Personality and Social Psychology Bulletin*, *27*(8), 969–981.

Antonakis, J., & Dalgas, O. (2009). Predicting elections: Child's play!. *Science*, *323*(5918), 1183.

Ballew, C. C., & Todorov, A. (2007). Predicting political elections from rapid and unreflective face judgments. *Proceedings of the National Academy of Sciences*, *104*(46), 17948–17953.

Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion*, *6*(2), 269–278.

Berscheid, E., & Walster, E. (1974). Physical attractiveness. In *Advances in experimental social psychology* (Vol. 7, pp. 157–215). Academic Press.

Brambilla, M., Biella, M., & Freeman, J. B. (2018). The influence of visual context on the evaluation of facial trustworthiness. *Journal of Experimental Social Psychology*, *78*, 34–42.

Brinol, P., & Petty, R. E. (2009). Source factors in persuasion: A self-validation approach. *European Review of Social Psychology*, *20*(1), 49–96.

Brownlow, S., & Zebrowitz, L. A. (1990). Facial appearance, gender, and credibility in television commercials. *Journal of Nonverbal Behavior*, *14*(1), 51–60.

Cao, J., & Banaji, M. R. (2016). The base rate principle and the fairness principle in social judgment. *Proceedings of the National Academy of Sciences*, *113*(27), 7475–7480.

Chang, L. J., Doll, B. B., van't Wout, M., Frank, M. J., & Sanfey, A. G. (2010). Seeing is believing: Trustworthiness as a dynamic belief. *Cognitive Psychology*, *61*(2), 87–105.

Cone, J., & Ferguson, M. J. (2015). He did what? The role of diagnosticity in revising implicit evaluations. *Journal of Personality and Social Psychology*, *108*(1), 37–57.

Cone, J., Flaharty, K., & Ferguson, M. J. (2019). Believability of evidence matters for correcting social impressions. *Proceedings of the National Academy of Sciences*, *116*(20), 9802–9807.

Cone, J., Flaharty, K., & Ferguson, M. J. (2021). The long-term effects of new evidence on implicit impressions of other people. *Psychological Science*, *32*(2), 173–188.

Dean, D. H. (2017). The benefit of a trustworthy face to a financial services provider. *Journal of Services Marketing*, *31*(7), 771–783.

De Houwer, J. (2018). Propositional models of evaluative conditioning. *Social Psychological Bulletin*, *13*(3), 1–21.

De Houwer, J., & Hughes, S. (2016). Evaluative conditioning as a symbolic phenomenon: On the relation between evaluative conditioning, evaluative conditioning via instructions, and persuasion. *Social Cognition*, *34*(5),480–494.

Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, *24*(3), 285–290.

Duarte, J., Siegel, S., & Young, L. (2012). Trust and credit: The role of appearance in peer-to-peer lending. *The Review of Financial Studies*, *25*(8), 2455–2484.

Dumas, R., & Teste, B. (2006). The influence of criminal facial stereotypes on juridic judgments. *Swiss Journal of Psychology*, *65*, 237–244.

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good, but…: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin*, *110*(1), 109–128.

Ferguson, M. J., Mann, T. C., Cone, J., & Shen, X. (2019). When and how implicit first impressions can be updated. *Current Directions in Psychological Science*, *28*(4), 331–336.

Fiske, S. T. (1980). Attention and weight in person perception: The impact of negative and extreme behavior. *Journal of Personality and Social Psychology*, *38*(6), 889–906.

Gheorghiu, A. I., Callan, M. J., & Skylark, W. J. (2017). Facial appearance affects science communication. *Proceedings of the National Academy of Sciences*, *114*(23), 5970–5975.

Gomulya, D., Wong, E. M., Ormiston, M. E., & Boeker, W. (2017). The role of facial appearance on CEO selection after firm misconduct. *Journal of Applied Psychology*, *102*(4), 617–635.

Gregg, A. P., Seibt, B., & Banaji, M. R. (2006). Easier done than undone: Asymmetry in the malleability of implicit preferences. *Journal of Personality and Social Psychology*, *90*(1), 1–20.

Hosoda, M., Stone-Romero, E. F., & Coats, G. (2003). The effects of physical attractiveness on job-related outcomes: A meta-analysis of experimental studies. *Personnel Psychology*, *56*(2), 431–462.

Hamilton, D. L. (2015). *Cognitive processes in stereotyping and intergroup behavior*. Psychology Press.

Hu, X., Gawronski, B., & Balas, R. (2017). Propositional versus dual-process accounts of evaluative conditioning: I. The effects of co-occurrence and relational information on implicit and explicit evaluations. *Personality and Social Psychology Bulletin* , *43*(1), 17–32.

Izard, C. E. (1994). Innate and universal facial expressions: Evidence from developmental and cross-cultural research. *Psychological Bulletin*, *115*(2), 288–299.

Jaeger, B., Evans, A. M., Stel, M., & van Beest, I. (2019a). Explaining the persistent influence of facial cues in social decision-making. *Journal of Experimental Psychology: General*, *148*(6), 1008.

Jaeger, B., Evans, A., Stel, M., & van Beest, I. (2019b). Lay beliefs in physiognomy explain overreliance on facial impressions. *PsyArXiv*. 10.31234/osf.io/8dq4x

Lai, C. K., Skinner, A. L., Cooley, E., Murrar, S., Brauer, M., Devos, T.,… & Nosek, B. A. (2016). Reducing implicit racial preferences: II. Intervention effectiveness across time. *Journal of Experimental Psychology: General*, *145*(8), 1001–1016.

Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin*, *126*(3), 390.

Lawson, C., & Lenz, G. S. (2007). Looking like a presidente: Appearance and electability among Mexican candidates. *Unpublished manuscript, Department of Political Science, Massachusetts Institute of Technology*.

Mann, T. C., Cone, J., Heggeseth, B., & Ferguson, M. J. (2019). Updating implicit impressions: New evidence on intentionality and the affect misattribution procedure. *Journal of Personality and Social Psychology*, *116*(3), 349–374.

Mann, T. C., & Ferguson, M. J. (2015). Can we undo our first impressions? The role of reinterpretation in reversing implicit evaluations. *Journal of Personality and Social Psychology*, *108*(6), 823–849.

Mann, T. C., & Ferguson, M. J. (2017). Reversing implicit first impressions through reinterpretation after a two-day delay. *Journal of Experimental Social Psychology*, *68*, 122–127.

Matsumoto, D., & Willingham, B. (2009). Spontaneous facial expressions of emotion of congenitally and noncongenitally blind individuals. *Journal of Personality and Social Psychology*, *96*(1), 1–10.

McConnell, A. R., Rydell, R. J., Strain, L. M., & Mackie, D. M. (2008). Forming implicit and explicit attitudes toward individuals: Social group association cues. *Journal of Personality and Social Psychology*, *94*(5), 792–807.

Montepare, J. M., & Dobish, H. (2003). The contribution of emotion perceptions and their overgeneralizations to trait impressions. *Journal of Nonverbal behavior*, *27*(4), 237–254.

Montepare, J. M., & Zebrowitz-McArthur, L. (1989). Children's perceptions of babyfaced adults. *Perceptual and Motor Skills*, *69*(2), 467–472.

Moran, T., Bar-Anan, Y., & Nosek, B. A. (2017). The effect of the validity of co-occurrence on automatic and deliberate evaluations. *European Journal of Social Psychology*, *47*(6), 708–723.

Newman, L. S., & Uleman, J. S. (1993). When are you what you did? Behavior identification and dispositional inference in person memory, attribution, and social judgment. *Personality and Social Psychology Bulletin*, *19*(5), 513–525.

Olivola, C. Y., & Todorov, A. (2010). Elected in 100 milliseconds: Appearance-based trait inferences and voting. *Journal of Nonverbal Behavior*, *34*(2), 83–110.

Oosterhof, N. N., & Todorov, A. (2009). Shared perceptual basis of emotional expressions and trustworthiness impressions from faces. *Emotion*, 9 (1), 128–133.

Peters, K. R., & Gawronski, B. (2011). Are we puppets on a string? Comparing the impact of contingency and validity on implicit and explicit evaluations. *Personality and Social Psychology Bulletin*, *37*(4), 557–569.

Pandey, G., & Zayas, V. (2021). What is a face worth? Facial attractiveness biases experience-based monetary decision-making. *British Journal of Psychology*, *112*(4), 934–963.

Porter, S., ten Brinke, L., & Gustaw, C. (2010). Dangerous decisions: The impact of first impressions of trustworthiness on the evaluation of legal evidence and defendant culpability. *Psychology, Crime & Law*, *16*(6), 477–491.

Poutvaara, P., Jordahl, H., & Berggren, N. (2009). Faces of politicians: Babyfaceness predicts inferred competence but not electoral success. *Journal of Experimental Social Psychology*, *45*(5), 1132–1135.

Reeder, G. D., & Brewer, M. B. (1979). A schematic model of dispositional attribution in interpersonal perception. *Psychological Review*, *86*(1), 61–79.

Rezlescu, C., Duchaine, B., Olivola, C. Y., & Chater, N. (2012). Unfakeable facial configurations affect strategic choices in trust games with or without information about past behavior. *PloS One*, *7*(3), e34293.

Rudert, S. C., Reutner, L., Greifeneder, R., & Walker, M. (2017). Faced with exclusion: Perceived facial warmth and competence influence moral judgments of social exclusion. *Journal of Experimental Social Psychology*, *68*, 101–112.

Rudoy, J. D., & Paller, K. A. (2009). Who can you trust? Behavioral and neural differences between perceptual and memory-based influences. *Frontiers in Human Neuroscience*, *3*, 16.

Rule, N. O., Tskhay, K. O., Freeman, J. B., & Ambady, N. (2014). On the interactive influence of facial appearance and explicit knowledge in social categorization. *European Journal of Social Psychology*, *44*(6), 529–535.

Rydell, R. J., & McConnell, A. R. (2006). Understanding implicit and explicit attitude change: a systems of reasoning analysis. *Journal of Personality and Social Psychology*, *91*(6), 995–1008.

Rydell, R. J., McConnell, A. R., Strain, L. M., Claypool, H. M., & Hugenberg, K. (2007). Implicit and explicit attitudes respond differently to increasing amounts of counterattitudinal information. *European Journal of Social Psychology*, *37*(5), 867–878.

Shen, X., Mann, T. C., & Ferguson, M. J. (2020). Beware a dishonest face? Updating face-based implicit impressions using diagnostic behavioral information. *Journal of Experimental Social Psychology*, *86*, 103888.

Shen, X., & Ferguson, M. J. (2021). How resistant are implicit impressions of facial trustworthiness? When new evidence leads to durable updating. *Journal of Experimental Social Psychology*, *97*, 104219.

Shen, X., & Ferguson, M. J. (in prep). What does face-based implicit impressions reflect?

Skowronski, J. J., & Carlston, D. E. (1987). Social judgment and social memory: The role of cue diagnosticity in negativity, positivity, and extremity biases. *Journal of Personality and Social Psychology*, *52*(4), 689–699.

Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin*, *105*(1), 131.

Smith, C. T., De Houwer, J., & Nosek, B. A. (2013). Consider the source: Persuasion of implicit evaluations is moderated by source credibility. *Personality and Social Psychology Bulletin*, *39*(2), 193–205.

South Palomares, J. K., & Young, A. W. (2018). Facial first impressions of partner preference traits: Trustworthiness, status, and attractiveness. *Social Psychological and Personality Science*, *9*(8), 990–1000.

Stirrat, M., & Perrett, D. I. (2010). Valid facial cues to cooperation and trust: Male facial width and trustworthiness. *Psychological Science*, *21*(3), 349–354.

Suzuki, A., Tsukamoto, S., & Takahashi, Y. (2019). Faces tell everything in a just and biologically determined World: Lay theories behind face reading. *Social Psychological and Personality Science*, *10*(1), 62–72.

Todorov, A., Funk, F., Olivola, C.Y. (2015). Response to Bonnefon et al.: Limited 'kernels of truth' in facial inferences. *Trends in Cognitive Sciences*, *19*(8), 422–423.

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science*, *308*(5728), 1623–1626.

Todorov, A., & Olson, I. R. (2008). Robust learning of affective trait associations with faces when the hippocampus is damaged, but not when the amygdala and temporal pole are damaged. *Social Cognitive and Affective Neuroscience*, *3*(3), 195–203.

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition*, *27*(6), 813–833.

Todorov, A., & Uleman, J. S. (2002). Spontaneous trait inferences are bound to actors' faces: Evidence from a false recognition paradigm. *Journal of Personality and Social Psychology*, *83*(5), 1051–1065.

Todorov, A., & Uleman, J. S. (2003). The efficiency of binding spontaneous trait inferences to actors' faces. *Journal of Experimental Social Psychology*, *39*(6), 549–562.

Tormala, Z. L., Briñol, P., & Petty, R. E. (2006). When credibility attacks: The reverse impact of source credibility on persuasion. *Journal of Experimental Social Psychology*, *42*(5), 684–691.

Uleman, J. S., & Moskowitz, G. B. (1994). Unintended effects of goals on unintended inferences. *Journal of Personality and Social Psychology*, *66*(3), 490–501.

Van Dessel, P., Ye, Y., & De Houwer, J. (2019). Changing deep-rooted implicit evaluation in the blink of an eye: Negative verbal information shifts automatic liking of Gandhi. *Social Psychological and Personality Science*, *10*(2), 266–273.

Van Dessel, P., Hughes, S., & De Houwer, J. (2019). How do actions influence attitudes? An inferential account of the impact of action performance on stimulus evaluation. *Personality and Social Psychology Review*, *23*(3), 267–284.

Van't Wout, M., & Sanfey, A. G. (2008). Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. *Cognition*, *108*(3), 796–803.

Verosky, S. C., Porter, J., Martinez, J. E., & Todorov, A. (2018). Robust effects of affective person learning on evaluation of faces. *Journal of Personality and Social Psychology*, *114*(4), 516–528.

Verosky, S. C., & Todorov, A. (2010). Generalization of affective learning about faces to perceptually similar faces. *Psychological Science*, *21*(6), 779–785.

Weber, R., & Crocker, J. (1983). Cognitive processes in the revision of stereotypic beliefs. *Journal of Personality and Social Psychology*, *45*(5), 961–977.

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, *17*(7), 592–598.

Wilson, R. K., & Eckel, C. C. (2006). Judging a book by its cover: Beauty and expectations in the trust game. *Political Research Quarterly*, *59*(2), 189–202.

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: unwanted influences on judgments and evaluations. *Psychological Bulletin*, *116*(1), 117–142.

Wojciszke, B., Brycz, H., & Borkenau, P. (1993). Effects of information content and evaluative extremity on positivity and negativity biases. *Journal of Personality and Social Psychology*, *64*(3), 327–335.

Zebrowitz, L. A., & McDonald, S. M. (1991). The impact of litigants' baby-facedness and attractiveness on adjudications in small claims courts. *Law and Human Behavior*, *15*(6), 603–623.

Zebrowitz, L. A., & Montepare, J. M. (1992). Impressions of baby-faced individuals across the life span. *Developmental Psychology*, *28*(6), 1143.

Zebrowitz, L. A., Tenenbaum, D. R., & Goldstein, L. H. (1991). The impact of job applicants' facial maturity, gender, and academic achievement on hiring recommendations. *Journal of Applied Social Psychology*, *21*(7), 525–548.

Zebrowitz, L. A., Collins, M. A., & Dutta, R. (1998). The relationship between appearance and personality across the life span. *Personality and Social Psychology Bulletin*, *24*(7), 736–749.

Zebrowitz, L. A., Andreoletti, C., Collins, M. A., Lee, S. Y., & Blumenthal, J. (1998). Bright, bad, babyfaced boys: appearance stereotypes do not always yield self-fulfilling prophecy effects. *Journal of Personality and Social Psychology*, *75*(5), 1300–1320.

Zebrowitz, L. A., Franklin, R. G., & Boshyan, J. (2015). Face shape and behavior: Implications of similarities in infants and adults. *Personality and Individual Differences*, *86*, 312–317.

20 Memory Consolidation: The Cornerstone for Gauging Spontaneous Impression Longevity

Jessica R. Bray, Angel D. Armenta, and Michael A. Zárate
The University of Texas at El Paso

Imagine you are a student walking into the first day of class. There is a free seat next to someone who sits at their desk quietly with a notebook and pen ready; and a free seat next to someone who is sleeping with their headphones in. Based on what you see, who do you choose to sit next to? Choosing where to sit relies on many factors such as your preference for how close you want to be to the front, how well you can see the screen from the prospective seat, and your impression of those you will be sitting next to. Initial impressions are critical in making quick decisions about others. These impressions come in many forms and are based on heuristic cues (e.g., stereotypes) and behaviors. With an array of social information around us, one might ask "How do first impressions form?" This chapter focuses on the formation of first impressions, whether first impressions can be altered, and the longevity of first impressions.

Forming and Categorizing First Impressions

When we first meet someone, we extract social information about them. Going back to choosing a seat on the first day of class, we can extract social information about each student by making assumptions about their respective character traits. One is ready for lecture and assumed to be studious, while the other is assumed to be lazy for sleeping on the first day of class. With a decision as crucial as picking your "assigned" seat for the remainder of the semester, it is important to make sure you extract enough social information to make an informed decision on who to sit next to. Many social factors are important in making that first impression (Zárate & Smith, 1990), all of which allow perceivers to extract valuable information that is used to form a holistic impression of a person.

Basic level categorical characteristics such as age and gender are spontaneously and automatically used to categorize and make inferences about others (Zárate & Smith, 1990). For example, Zárate and Smith (1990) used a social category verification task to examine whether ethnic and gender

DOI: 10.4324/9781003045687-23

stereotypes influenced how quickly participants categorized photos of different target individuals. For the task, participants were first shown a headshot of either a Black or White person. They were then shown a category (male/female or Black/White) and were asked to determine whether the person belonged to the displayed category. Results indicated that the faster one categorized other people by particular characteristics (sex or race), the more they stereotyped others based on those categorical inferences. So, the participants who more quickly categorized target individuals by race were more likely to activate racial stereotypes rather than gender stereotypes. Thus, basic-level categorization processes are one of the first in which people automatically encode and remember others.

Those processes can be influenced by a number of factors. Context matters. The lone male in a group of five women will more quickly be categorized by gender. Similarly, one might scan memory differently if one is thinking of a potential date for an upcoming event versus considering who might make a good study partner (ideally, the same person might come to mind, but that is another issue). Target characteristics matter as well. Particularly prototypic exemplars might be categorized along a dimension faster than might others. Zárate and Smith (1990) identified these exemplars (Smith & Zárate, 1992) and individual-level factors, such as one's gender or race, that influence first categorization processes.

The target person's behavior might also spontaneously influence those processes, and those trait inferences seem to occur quickly as well. When we observe behaviors, numerous spontaneous inferences form based on the perceiver's goal (see Uleman et al., 2008 for a review). Spontaneous inferences generally occur without conscious intention and arise from interpreting some form of social behavior (Hassin et al., 2002; Uleman, 1999; Uleman, Hon, et al., 1996; Uleman, Newman, & Moskowitz, 1996; Uleman et al., 2008). Since Winter and Uleman's (1984) seminal work on spontaneous trait inferences, research has primarily focused on how traits are inferred from some presented behaviors. However, in the early 2000s, researchers began investigating whether behaviors also imply other information such as goals, situations, and positive or negative evaluations of people (see Schneid et al., this volume). Each spontaneous inference has its' own nuances and most work in conjunction with spontaneous trait inferences.

Character Trait Impressions

Spontaneous trait inferences occur when trait information about a person is inferred based on their behavior (Carlston & Skowronski, 1994; Todorov & Uleman, 2002; Uleman, Hon, et al., 1996; Winter & Uleman, 1984). For example, you might infer Joe is helpful if you see him hailing a cab for an elderly person. This process helps form impressions because the inferred traits are encoded into memory and are available to be retrieved and later used to shape explicit impressions (Carlston & Skowronski, 1994; Moskowitz &

Roman, 1992; Winter & Uleman, 1984; Uleman, Newman, & Moskowitz, 1996). They can also be used to predict people's future behavior (e.g., McCarthy & Skowronski, 2011; Schneid, Carlston, & Skowronski, 2015). For example, because Joe has been helpful in the past, you may be likely to infer that he will be helpful in the future.

Spontaneous trait inferences can also lead to changes in perception and behavior. For example, research has shown that in as little as 100 ms, people infer other's perceived level of competence (Olivola & Todorov, 2010) and that there is a strong positive correlation between competence and voting patterns (Todorov et al., 2005). Todorov et al. (2005) demonstrated how competence inferences impact who people vote for. In their study, participants rated photographs of politicians on their level of competence. Importantly, participants had no information regarding the candidates other than their photos. Results showed that politicians who were rated as more competent were more likely voted into office than those who were rated as less competent. Thus, automatic inferences may influence later decision making.

Additionally, some research suggests that stereotypes may inhibit the formation of spontaneous trait inferences (Wang & Yang, 2017; Wigboldus et al., 2003; Wigboldus et al., 2004). This research proposes that the formation of spontaneous trait inferences is obstructed when people perform behaviors that are inconsistent with the stereotypes associated with their social group. For example, priests-compared to junkies-are stereotyped as being honest. So, if we observe both a priest and a junkie returning a lost purse, we infer the priest as being honest but hesitate to make the same inference about the junkie. Although they may perform the same behavior (i.e., returning a lost purse), trait inferences are weakened when the behavior is inconsistent with stereotypes about the person (i.e., junkies are not honest) rather than when the behavior is consistent with the person's stereotypes (i.e., priests are honest). Therefore, activated stereotypes hinder the trait inference process when behaviors do not align with those stereotypes.

Perceiving Others' Goals

Spontaneous goal inferences occur when people infer others' goals based on their behaviors (Aarts et al., 2008; Hassin et al., 2005; Van der Cruyssen et al., 2009). Goal inferences appear less likely to be influenced by stereotypes and other impressionable information that has been shown to impact person perception (Moskowitz & Olcaysoy Okten, 2016). For example, let's suppose you observe an athlete tackling another person during a game. Multiple inferences can be made about the athlete. They might be perceived as aggressive for tackling someone else or their behavior might be ascribed to the goal of winning. Thus, this one behavior leads to a trait inference (the athlete is aggressive) and a goal inference (the athlete wants to win).

Van Overwalle et al. (2012) examined multiple inference formation by comparing the speed of activation for spontaneous goal inference formation

to spontaneous trait inference formation. Across four experiments, participants were presented with paragraphs that either implied certain goals (e.g., "The boy took the ball and ran outside" implies that his goal is to play), traits (e.g., "Little Oscar never says thank you" implies that Oscar is impolite), or both goals and traits (e.g., "After paying the bill, she left 5 euros on the table." implies the woman has a goal to tip the waitress and is generous) about target people. After reading the paragraph, participants completed a false recognition task where they had to categorize words as being present or absent from the paragraph they just read. The critical trials included goals and traits that were implied (but not explicitly stated) from the paragraphs. After comparing speeds to goal and trait inferences, results showed that goal inferences formed within 350 ms. However, trait inferences only occurred after 350 ms and when the paragraphs described behaviors that implied both traits and goals. Their results suggest that spontaneous goal inferences are activated faster than spontaneous trait inferences, indicating that behaviors are interpreted with distinct types of inferences, and those inferences are activated at different times.

Making Judgments on Whether Someone Is Good or Bad

Spontaneous evaluative inferences involve making evaluative judgments about others based on their behaviors. Specifically, spontaneous evaluative inferences occur when one infers that another person is either good or bad based solely on their behavior (Olcaysoy Okten et al., 2019; Schneid et al., 2015). Work on spontaneous evaluative inferences suggests that affective judgments are susceptible to change and take place before spontaneous trait inferences due to the longer processing time associated with encoding trait information (Olcaysoy Okten et al., 2019; Schneid, Carlston, & Skowronski, 2015; Schneid, Crawford, et al., 2015). Additionally, research has shown that relative to trait inferences, evaluative inferences can be updated more readily (Olcaysoy Okten et al., 2019). For example, Jack might infer Jill is both mean and a bad person for yelling at a child. However, if Jack found out that Jill yelled at the child to protect them from a hot stove then Jack's new evaluative impression of Jill is that she's a good person. Thus, while the trait inference remains intact, Jack's overall evaluative judgment of Jill is updated (see Moskowitz et al., Chapter 18, this volume).

Opportunities for Automatic Inferences Are Unbounded

Research shows, in a well replicable fashion, that individuals automatically encode others by race and gender, make spontaneous inferences about their goals and traits, and they do this seemingly effortlessly. It is also the case, however, that the social environment is rich. Multiple motives and processes appear to be at work in any social interaction beyond simply their social

category membership and inferences from their actual behavior. First impressions often provide multiple types of information about a person.

One clear factor people seem to "automatically attend to" is physical attractiveness. Impressions based on facial cues are not only formed relatively quickly (Willis & Todorov, 2006), but people with higher levels of physical attractiveness are thought of as being more intelligent, socially competent, and outgoing (Zebrowitz et al., 2002; Zebrowitz & Montepare, 2008). This positivity bias ascribed to highly attractive individuals has been described as the "what's beautiful is good" stereotype which has been noted as one of the strongest measures of social competence (Dion et al., 1972). Additionally, the positive impressions we form about physically attractive people have been shown to influence behaviors. For example, physically attractive individuals receive higher grades, intelligence ratings, academic potentials, and more lenient punishments in comparison to people who are less physically attractive (Ritts et al., 1992). Clearly, physical attractiveness is one of the primary heuristics that aids in impression formation; however, people use other cues as well.

Stimulus valence—whether a stimulus is positive or negative—is another type of automatic inference process. Research shows that positive evaluations are tied to white stimuli while negative evaluations are associated with black stimuli (Eder & Rothermund, 2010; Meier et al., 2004). For example, Meier et al. (2004) designed a series of five experiments where participants were asked to categorize words as either positive or negative in meaning. The words were presented on grey backgrounds and were either typed in white or black font. Results across all experiments showed that participants were faster and more accurate at categorizing positive words written in white and negative words written in black relative to valenced words written with a mismatched color (e.g., positive word written in black font). These results suggest that there is an automatic association and activation of positivity being light and negativity being dark. However, automatic inferences are not limited to blatant cues (e.g., color); They also occur with subtle environmental cues.

Aggressive stimuli have also been shown to influence people's perceptions. Todorov and Bargh (2002) reviewed a series of experiments that included priming aggression. Their review suggests that when an environment has aggressive stimuli, perceptions of ambiguous behaviors are automatically interpreted as more aggressive. This concept translates into the sinister attribution error found in consumer psychology. The sinister attribution error is the tendency for consumers to feel salespeople's behaviors are untrustworthy. For example, in a series of experiments, Main et al. (2007) had participants purchase a pair of sunglasses from a store on campus. The participants were flattered by the salesperson either before purchasing the sunglasses or after. Results showed that participants who were flattered before purchasing thought the salesperson was less trustworthy than when flattered after purchasing or not flattered at all. Overall, these studies suggest that environmental (e.g., aggressive priming) and situational cues (e.g., buying from a salesperson) influence the type of automatic impression that is made.

Our review so far suggests that social perceivers are influenced by a variety of factors at encoding to perceive the social world. Our research also shows cultural differences in these encoding operations as well. Specifically, individualistic people are more likely to form and encode trait inferences relative to collectivistic people. Zárate et al. (2001) first tested the influence of culture on spontaneous trait inferences by using a modified lexical decision task. In critical trials, participants were shown sentences that imply traits (e.g., "He checked everyone's seat belts before starting off.") followed by a letter string that was either a trait implied by the sentence (e.g., "cautious") or an unrelated trait word. Results indicated that White participants responded significantly faster to implied traits relative to unrelated traits while Latinx participants responded equally fast to both trait types. These results were the first to suggest that culture plays a role in the trait inference process such that individualistic people take primacy in forming trait inferences while collectivistic people may prefer forming impressions based on situational cues or other group-oriented cues. Recent work by Lee et al. (2017) extends these findings by showing that European Canadians formed spontaneous trait inferences more often than spontaneous situational inferences while Japanese participants showed no difference in trait and situational inference formation. Shimizu et al. (2017) replicated these results by showing that American participants formed spontaneous trait inferences more often than Japanese participants. Thus, the type of automatic impressions that are formed are culture-dependent.

We have outlined only a few of the primary influences (aside from the perceived behavior) on how social perceivers seem to "automatically" or spontaneously encode others. Uleman and Saribay (2012) further outline a number of ways in which social information is initially encoded. Race, sex, and age are automatically perceived. People make spontaneous trait and evaluative inferences. Attractiveness, competence, sexual orientation, social class, and a host of other factors influence initial judgments. Compound that with the realization that all of these effects might be influenced by various cultural factors, and it suggests a much more complicated process than suggested by any particular research finding. Because of the many different types of automatic inferences that researchers have identified, complicated by the further cultural difference in those processes, our more recent work is investigating for long-term effects of these initial impressions. With the implosion of inferences that social perceivers might make, our research has begun to investigate some long-term consequences of the many possible first inferences. We use the term "implosion" to suggest that the social world imposes many possible social inference outcomes, and if in fact, individuals were doing all the things outlined above, it would suggest an endless number of inferences. So how does the mind make long-term use of this information? Knowing the long-term consequences has many benefits. It might suggest what people do, without prompting, in the first interaction. Thus, in the typical social psychological experiment, one might make multiple "attractiveness" ratings in a row. That attention to

attractiveness in the first few trials might bias the entire process and make that particular variable more salient than it normally might be. An investigation of the long-term effects might also suggest what types of information get the most initial attention, or at least what might be the most important (to the perceiver) attribute. Finally, we contend that social cognitive researchers rarely actually study long-term memory processes. Most studies are completed in one experimental session, and that tells us little about any long-term consequences of those initial interactions. Experimental procedures often include five-minute distractor tasks, but that hardly reflects long-term memory. Here, we outline a research program that investigates how person perception experiences on one day influence reactions two days later. The underlying idea is that over time, the most salient and impactful first judgments will be more evident than the mundane.

Memory Consolidation

Given the many possible inferences people make, our question becomes, what is stored and used later? How does one make long-term sense of the daily barrage of information and inferences one receives and makes during the day? To that end, we have developed a model of memory consolidation in impression formation. Outlined below is a summary of the memory consolidation literature, and a short introduction to our model.

Memory consolidation is the physiological process of moving newly acquired information into long-term, stable memories. In essence, memory consolidation describes how information withstands the passage of time (Stickgold & Walker, 2007). When we encounter some form of new information, such as reading a new article, the information we gather is "weak". During sleep, neural connections that form memories are strengthened, subsequently allowing for the "weak" information to also be strengthened which leads to more stable memories after sleep. Specifically, information that has been consolidated is not easily influenced by dual memory tasks which is evidence that it has become more concrete in long-term memory structures (Diekelmann et al., 2009; Payne et al., 2008; Walker & Stickgold, 2004). To date, most research on memory consolidation focuses on the neural components involved with memory or how physiological factors influence memory (see Nadel et al., 2012 for a review). However, little research in social cognition has integrated memory consolidation concepts with social psychological theory.

Our integration of the aforementioned memory consolidation literature led to the development of a social model of memory consolidation. The model merges social psychological constructs with memory consolidation findings to suggest new research directions in social cognition. The model starts with the assumption that at some level, social perceivers tend to develop whole or gestalt impressions of others (Asch, 1946). Over time, features become integrated perceptions rather than remain single features. The research also

focuses on evaluative judgments, as evaluative judgments seem critical in social perception. Finally, the model recognizes that the development of most automatic associations takes practice. All models of automaticity share one common feature—practice. Thus, it may be unreasonable to think that some ten minute training method will have a great influence today on well-practiced learned associations. It may take time to influence and change any learned memory associations.

Our model of memory consolidation has three distinct parts: integration, accessibility, and generalizability (Enge et al., 2015; Lupo & Zárate, 2019; Zárate & Enge, 2013). Integration involves overlapping newly acquired information with information stored in existing long-term memory structures (Huguet et al., 2019; Lupo & Zárate, 2019). Specifically, through consolidation processes, new information activates and subsequently melds with preexisting cognitive structures that hold old information, which then leads to learning. For example, let's pretend that you are attending your first college football game. When you get there, you learn that other students from your school are wearing spirit gear (e.g., foam fingers, football jerseys, college t-shirts, etc.). Once learned, this information is theorized to be held in neocortical systems. Later, when you attend your first basketball game, you realize that students from your school are also wearing spirit gear. This information is held in the hippocampus. During integration, the new information (i.e., spirit gear is worn during basketball games) activates related old information (i.e., spirit gear is worn during football games) which, in turn, makes the new information easier to access and, as a result, easier to transfer to long-term memory structures.

This integration process strengthens the memory trace for the newly acquired information, leading to more stable and less malleable memories over time. As illustrated by the college sports games example, integration is theorized to facilitate stable memories over time due to the coactivation of related constructs. Thus, learning that students show their school spirit at one sports game helps you draw the same conclusion faster at a different sports game. Another example may be learning about personality traits among ingroup members. For instance, if you learn that a racial ingroup member is kind, the two constructs (i.e., your racial ingroup and the personality trait kind) become associated during integration. Thus, when the representation of one construct (e.g., an ingroup member) is activated, the other representation (e.g., kind) also becomes activated, leading to more opportunities for learned information to become activated and therefore more likely to be remembered across time. However, not all events are integrated equally well with existing memory structures. Some types of events, like more emotional events, are more likely to become integrated and remembered later. Thus, the integration of new information does not occur instantaneously. Integration occurs at multiple levels (from the synaptic level to the systems level) and takes time. Our research most often tests for the effects of consolidation two days after the learning episode. We note that two nights of sleep should not differ theoretically from one night

of sleep. The two-day delay simply reflects student schedules and helps with participant recruitment. So far, all participants have slept during that two day period. Consequently, studies that test for the effects of prior experiences within the single-session study are testing non-consolidated memories. For example, in one particular study (Enge et al., 2015), participants learned about various ingroup and outgroup members. They were shown target photos paired with news articles that presented either positive or negative information. In a later (four hours or two days) lexical decision task, participants demonstrated paired-associate learning to the outgroup information with the negative trait information—but only after two days. That learning took time and demonstrated how specific negative information takes time for it to be automatically associated with particular targets.

The next step in the consolidation process is accessibility. Accessibility is a byproduct of integration. Once memories are integrated into long-term memory structures, they become more readily accessible. Accessibility refers to the increased speed and ease that the stored information can be retrieved. Information that is integrated with other cognitive structures tends to be accessed quicker and is important for fast decision making. This is also the least studied component of the model. The model suggests, for instance, that the retrieval of consolidated information should be easier and completed more automatically than non-consolidated information. We hypothesize, for instance, that consolidated information (information learned days earlier) should be retrieved more automatically and in a more "know" fashion than non-consolidated information (information learned two hours earlier). We are currently testing this hypothesis within an intergroup context and cannot yet say whether data support our predictions. However, this work parallels attitude accessibility theory. When people have strong attitudes about a given topic, they readily express their opinions and feelings. However, when people are unsure of how they feel about a topic, it takes them longer to respond (Fazio et al., 1982; Krosnick, 1989). Thus, compared to weaker attitudes, stronger attitudes are more integrated, accessible, and utilized much like consolidated memories.

In contrast to consolidated information, we hypothesize that non-consolidated information should be retrieved in a more conscious, active retrieval type process. Because non-consolidated information has not been integrated into long-term memory structures, it would not be as easily cued and subsequently less likely to be activated. Thus, one would have to try to deliberately and intentionally access non-consolidated information whereas consolidated information should be activated with minimal effort. Within a "remember-know" type dichotomy, non-consolidated information should be retrieved in a more "remember"-type manner than in a "know"-type manner (Jacoby, 1991).

The final step in the consolidation process is generalizability. Once the information has been integrated and can be easily accessed, it can be used to make generalizations about new information. Generalization is the process of

extending integrated, stable, memory traces to newly acquired, relevant information. For example, Lupo and Zárate (2019) illustrated how being a part of a social group led to generalizing group characteristics to self-evaluations. Participants completed a "figure/ground" task where they had to indicate whether they recognized the figure or ground of an image first. Participants were then randomly assigned to one of two groups, purportedly on their performance on this task. Those groups would later compete in a trivia competition. In reality, participants were randomly assigned to be part of either Group A or Group B. They learned that members of Group A were generally anxious and warm, and members of Group B were generally arrogant and competent. It is noted that each group was assigned a relatively negative trait (anxious or arrogant). Participants then formed impressions about four members from Group A and four members of Group B by rating each individual on how much they liked the person, how friendly the person appeared, and if the person was a part of their assigned group. Later on (either the same day or after sleep), participants rated each group member and themselves on their respective traits (i.e., warm, anxious, arrogant, competent) and indicated their similarity to ingroup members. Results showed that after sleep, participants rated themselves higher on the negative traits, compared to before sleep. Hence, that negative information became integrated into their self-concepts and led to elevated ratings on even negative traits. Thus, if the participant was a Group A member, they identified themselves as being warm. Additionally, after sleep, participants in Group A also rated themselves as being more anxious. Participants generalized from four group members to the entire group—but only after time. This line of research shows that group information was learned, and some associated trait information was associated with that group structure. Only after consolidation into long-term memory (i.e., after time elapsed), group information became an integrated part of the self-concept, and participants evaluated themselves higher on those negative traits.

Similar research has been found for spontaneous trait inferences. Recent studies demonstrate that once spontaneous trait inferences are consolidated and integrated into long-term memory structures, those inferences are not affected by new trait-inconsistent information (Olcaysoy Okten & Moskowitz, 2020). In line with our memory consolidation model, this suggests that once information is consolidated, the information becomes more stable and less malleable across time.

Why Do Some Impressions Stick While Others Do Not?

This review does not question the automaticity of first impressions. We have learned in this chapter that seemingly irrelevant social information, such as one's gender or attractiveness can lead to a myriad of inferred impressions and those impressions can be altered by other internal factors (e.g., by prejudices). However, the literature thus far ignores one key aspect: the longevity of these

impressions is seldom tested. Studying spontaneous impressions in cross-sectional methodologies only tells us about the formation of these impressions. As you can imagine, remembering impressions, building on those impressions, and perhaps altering impressions can be valuable for successful social interactions and close relationships. A memory consolidation framework can serve as a steppingstone into investigating the longevity of automatic impressions. For example, Olcaysoy Okten and Moskowitz (2020) show spontaneous trait inferences automatically occur, persist after 48 hours, and are difficult to update. Their work is perhaps one of the first that directly tests, and shows, that automatic inferences are consolidated. However, work using a memory consolidation framework suggests that not all information is consolidated equally. Situational and intrinsic characteristics influence the likelihood of information being consolidated. Information is better recalled if it is negative, important to one's beliefs or interests, or highly emotional.

Information valence—whether the information is negative or positive—heavily influences the likelihood that information will be encoded and later retrieved. Generally, negative information is encoded and retrieved more quickly than positive or neutral information (Baumeister et al., 2001; Bebbington et al., 2017; Dijksterhuis & Aarts, 2003; Enge et al., 2015; Lupo & Zárate, 2019; Norris, 2019). One explanation for this negativity bias is that negative information is more diagnostic than positive or neutral information by providing us with vital information needed to stray away from danger (Baumeister et al., 2001; Norris, 2019; Rozin & Royzman, 2001). This negativity bias is evident, even at the automatic inference level (Carlston & Skowronski, 2005; Schneid et al., 2015; Shimizu, 2017). For example, Shimizu (2017) investigated the spontaneity of the negativity bias by testing whether spontaneous trait inference activation differed depending on the valence and frequency of behaviors. Participants read about different positive and negative behaviors performed by targets and were later asked to learn target-trait pairings. Critical trials included traits that were implied by the behaviors participants initially read. Results for this study showed that regardless of behavioral frequency, negative traits were spontaneously inferred more compared to positive traits. These results suggest that negative information provides some sort of hierarchy in impression formation and person perception. Negative information, therefore, also has a heavy influence on both the type of impressions we make about others and how pervasive those impressions are. Additional research on personality assessment and moral character assessment shows similar patterns. The availability of negative information leads to less favorable character judgments, whereas positive information has little effect on impression formation (Brannon et al., 2017; Lupfer et al., 2000; Stewart, 1998).

Our work extends these findings by showing that the negativity bias persists days after the learning episode. Regardless of one's group membership (i.e., ingroup/outgroup), learned negative traits generalize to novel others who are assumed to be from the same group. For example, if I learn that members of some arbitrary group are cold, then by extension, any new

member of that group is also assumed to be cold. Not only does this negativity bias extend to group members, after a delay of a day (allowing for consolidation) participants also evaluate themselves higher on negative traits associated with their ingroup (Lupo & Zárate, 2019). Participants learned about two groups, and also learned that two members of one of the groups were "cold" or "cruel." Participants learned this individuated information about only two of the group members and did not learn any individuating information about the other group members. The next day (but not four hours later) they rated all the relevant group members as cold or cruel. We had hypothesized that during sleep (or time more generally), participants would form a cohesive impression of the group, and that trait evaluations of two members would extend to the rest of the group. That is exactly what occurred. The same did not occur for the positive traits. Over time, participants did not ascribe more "warm" and "considerate" trait ratings to the relevant group members. Thus, only the negative information generalized to the group over time. Together, the two studies in Lupo and Zárate show that if I join the "cold" group, I also rate myself as cold. It also shows that people are more likely to generalize negative information about some group members to the entire group. Thus, negative information is diagnostic and may have a preference in how social impressions of others are formed and maintained across time.

Another factor that influences the likelihood of remembering an impression is how salient, or relevant, the information is to the interpreter. Information that is highly relevant to one's goals, ideals, and social group is better consolidated and retrieved faster relative to information that is not deemed as important. This effect is consistently demonstrated within attitudes literature: Stronger attitudes are more accessible and influence subsequent behavior more so than weaker attitudes (Fazio et al., 1982; Fazio et al., 1989; Fazio & Williams, 1986; see Kraus, 1995 for a meta-analysis). One type of information that people find relevant is their group membership. Typically, information about ingroup members (those similar to one's social group) is better remembered when the information is positive. On the other hand, negative information is better ascribed and remembered when it is associated with outgroup members (those not in one's social group). For example, Otten and Moskowitz (2000) examined whether spontaneous trait inference activation differed depending on the group membership of the target by using a minimal group paradigm. The minimal group paradigm randomly assigns participants to one of two arbitrary groups. Participants then categorized targets into each of the two groups based on ostensibly irrelevant judgments. Results showed that participants inferred positive traits about ingroup members more so than outgroup members. These results suggest that group membership impacts the type of information used to form impressions of others. Specifically, positive information is attributed and inferred more when impressions of similar others are made.

Our own work on memory consolidation has also shown how group membership influences impression formation. For example, Enge et al. (2015) examined the formation of prejudice within a memory consolidation

framework. Participants were asked to form impressions of racial ingroup and outgroup members by reading positive or negative articles about each target. Their knowledge of the targets was then tested after a four-hour delay and after two nights of sleep. Results showed that participants were faster to retrieve negative information about outgroup members relative to ingroup members and faster to retrieve positive information about ingroup members relative to outgroup members—but primarily after consolidation. These findings were extended by Lupo and Zárate (2019) who showed that group membership also influences evaluations of the self; after sleep, negative and positive traits about ingroup members are adopted. Collectively, these findings highlight the importance of group membership on person perception and sense of self; there is a proclivity to ascribing oneself with shared ingroup attributes (whether those attributes are positive or negative) and dissimilar others with negative attributes. Thus, the memory consolidation research thus far suggests that negative stereotypes develop over time, yet the social psychological research has not fully identified that process.

Our work on memory consolidation and impression formation is some of the first that gauges how impressions develop over time. Studying impression formation before and after sleep provides a more holistic account of how social memories are integrated and retrieved. Given the sometimes inconsistent effects of these processes with stereotypes (which are well-formed memory structures), a memory consolidation framework provides a great avenue for future research. This work highlights the importance of promoting time as a vital variable in social impressions. Time, and we argue sleep, is often overlooked in social psychological research. Not taking time and sleep into consideration for social psychological research is hides the true magnitude of psychological effects. For example, in our own research, we generally find that before sleep, recollection of positive and negative social information does not vary. However, after sleep, there is a consistent negativity bias where negative information is better recalled relative to positive information. Without including time as a critical variable in our experimental designs, our conclusions about impression formation completely change. Incorporating memory consolidation frameworks can help answer more nuanced questions about social interactions. For example, we are currently investigating whether different types of memory processes occur when people form and recollect impressions about ingroup and outgroup members. We are also looking into whether threat may alter how social information is consolidated. However, these questions are just the tip of the iceberg.

What's Next for First Impression Research?

Thus far, we have mentioned how first impressions form and how these impressions last across time. While research has elaborated on the neurological components associated with memory consolidation (see Nadel et al., 2012 for a review), little research has examined how memory consolidation

may be influenced by social interactions. Future research should examine how social factors (e.g., group membership, stereotypes, etc.) influence the different stages of memory consolidation. To our knowledge, only two studies have examined the influence of group membership on memory consolidation (Enge et al., 2015; Lupo & Zárate, 2019).

Future research should also focus on defining characteristics that can influence the stability of first impressions. Recent work by Mann and Ferguson (2015; 2017) has touched on this issue by examining whether first impressions can be reversed over time. Their work suggests that adding information that explains behaviors can "reverse" first impressions. To test this theory, Mann and Ferguson (2015; 2017) had participants form an initial impression of a person who ransacked their neighbor's house. Participants then learned that the person ransacked their neighbor's house to save children because the home was on fire. Results indicated that participants initially formed a negative impression of the person, however, that impression turned positive when their behavior was explained. The reversed impression was still evident three days after the initial learning session. These findings suggest that the updated impression was consolidated between testing sessions. This research can be extended to examine how other factors (e.g., group membership) influence the likelihood of updating impressions and whether impressions change or remain stable across the passage of time.

Another avenue for future research may be to examine the longevity of automatic impressions. Throughout this chapter, we've described a subset of the innumerable impressions that form automatically. We have also touched base on novel research that suggests that these processes work in tandem to form holistic impressions of others (e.g., Ham & Vonk, 2003). However, researchers have just started investigating whether these automatic impressions withstand the passage of time. Are some types of automatic impressions more diagnostic than others? What automatic impressions lead to recognizing someone faster or more accurately in the future? Are some automatic impressions better for certain social situations compared to others? Using a memory consolidation framework can elucidate these questions. Researchers have established the formation and influence of automatic impressions in the moment, but we cannot truly understand their impact on social interaction without investigating how these impressions are encoded into our stable memory system.

Concluding Remarks

The social world is a rich environment that provides people with unlimited opportunities to form impressions of others. Impressions are not all made equally though. Memory is selective. Our goals, situations, group affiliations, and valence of our environment all impact the type of impression we form and its longevity. We are more likely to ascribe and encode negative impressions to outgroup members and positive impressions to ingroup members.

The process of memory consolidation can explain why some automatic impressions are longer lasting than others. Information that is salient to oneself is transferred from short-term memory into long-term, stable memory after sleep. Thus, while we can automatically form an impression of someone, sleep is an integral component in making that first impression long-lasting. Falling fast asleep then is not only vital for our general health, but it also plays a critical role in stabilizing the many impressions we make of others.

References

Aarts, H., Dijksterhuis, A., & Dik, G. (2008). Goal contagion: Inferring goals from others' actions – and what it leads to. In J. Y. Shah & W. Gardner (Eds.), *Handbook of motivation science*. (pp. 265–280). Guildford.

Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology*, *41*(3), 258–290. 10.1037/h0055756

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohns, K. D., (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323–370.

Bebbington, K., MacLeod, C., Ellison, T. M., & Fay, N. (2017). The sky is falling: Evidence of a negativity bias in the social transmission of information. *Evolution and Human Behavior*, *38*(1), 92–101. 10.1016/j.evolhumbehav.2016.07.004

Brannon, S. M., Sacchi, D. L. M., & Gawronski, B. (2017). (In)consistency in the eye of the beholder: The roles of warmth, competence, and valence in lay perceptions of inconsistency. *Journal of Experimental Social Psychology*, *70*, 80–94. 10.1016/j.jesp.2016.12.011

Carlston, D. E., & Skowronski, J. J. (1994). Savings in the Relearning of Trait Information as Evidence for Spontaneous Inference Generation. *Journal of Personality and Social Psychology*, 66(5), 840–856.

Carlston, D. E., & Skowronski, J. J. (2005). Linking versus thinking: Evidence for the different associative and attributional bases of spontaneous trait transference and spontaneous trait inference. *Journal of Personality and Social Psychology*, *89*(6), 884–898. 10.1037/0022-3514.89.6.884

Diekelmann, S., Wilhelm, I., & Born, J. (2009). The whats and whens of sleep-dependent memory consolidation. *Sleep Medicine Reviews*, *13*(5), 309–321. 10.1016/j.smrv.2008.08.002

Dijksterhuis, A., & Aarts, H. (2003). On wildebeests and humans: The preferential detection of negative stimuli. *Psychological Science*, *14*(1), 14–18.

Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, *24*(3), 285–290.

Eder, A. B., & Rothermund, K. (2010). Automatic influence of arousal information on evaluative processing: Valence-arousal interactions in an affective Simon task. *Cognition and Emotion*, *24*(6), 1053–1061.

Enge, L. R., Lupo, A. K., & Zárate, M. A. (2015). Neurocognitive mechanisms of prejudice formation: The role of time-dependent memory consolidation. *Psychological Science*, *26*(7), 964–971. 10.1177/0956797615572903

Fazio, R. H., Chen, J., McDonel, E. C., & Sherman, S. J. (1982). Attitude accessibility, attitude-behavior consistency, and the strength of the object-evaluation association. *Journal of Experimental Social Psychology*, *18*(4), 339–357. 10.1016/0022-1031(82)90058-0

Fazio, R. H., Powell, M. C., & Williams, C. J. (1989). The role of attitude accessibility in the attitude-to-behavior process. *Journal of Consumer Research*, *16*(3), 280–288. 10.1086/209214

Fazio, R. H., & Williams, C. J. (1986). Attitude accessibility as a moderator of the attitude–perception and attitude–behavior relations: An investigation of the 1984 presidential election. *Journal of Personality and Social Psychology*, *51*(3), 505–514. 10.1037/0022-3514.51.3.505

Ham, J. & Vonk, R. (2003). Smart and easy: Co-occurring activation of spontaneous trait inferences and spontaneous situational inferences. *Journal of Experimental Social Psychology*, *39*, 434–447.

Hassin, R. R., Aarts, H., & Ferguson, M. J. (2005). Automatic goal inferences. *Journal of Experimental Social Psychology*, *41*(2), 129–140. 10.1016/j.jesp.2004.06.008

Hassin, R.R., Bargh, J. A., & Uleman, J. S. (2002). Spontaneous causal inferences. *Journal of Experimental Social Psychology*, *38*, 515–522.

Huguet, M., Payne, J. D., Kim, S. Y., & Alger, S. E. (2019). Overnight sleep benefits both neutral and negative direct associative and relational memory. *Cognitive, Affective, & Behavioral Neuroscience*, *19*(6), 1391–1403. 10.3758/s13415-019-00746-8

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513–541.

Kraus, S. J. (1995). Attitudes and the prediction of behavior: A meta-analysis of the empirical literature. *Personality and Social Psychology Bulletin*, *21*(1), 58–75. 10.1177/0146167295211007

Krosnick, J. A. (1989). Attitude Importance and Attitude Accessibility. *Personality and Social Psychology Bulletin*, *15*(3), 297–308. 10.1177/0146167289153002

Lee, H., Shimizu, Y., Masuda, T., & Uleman, J. S. (2017). Cultural differenecs in spontaneous trait and situation inferences. *Journal of Cross-Cultural Psychology*, *48*(5), 1–17.

Lupfer, M. B., Weeks, M., & Dupuis, S. (2000). How pervasive is the negativity bias in judgements based on character appraisal? *Personality and Social Psychology Bulletin*, *26*(11), 1353–1366.

Lupo, A. K., & Zárate, M. A. (2019). Guilty by association: Time-dependent memory consolidation facilitates the generalization of negative – but not positive – person memories to group and self-judgments. *Journal of Experimental Social Psychology*, *83*, 78–87. 10.1016/j.jesp.2019.02.008

Main, K. J., Dahl, D. W., & Darke, P. R. (2007). Deliberative and automatic bases of suspicion: Empirical evidence of the sinister attribution error. *Journal of Consumer Psychology*, *17*(1), 59–69.

Mann, T. C., & Ferguson, M. J. (2015). Can we undo our first impressions? The role of reinterpretation in reversing implicit evaluations. *Journal of Personality and Social Psychology*, *108*(6), 823–849. 10.1037/pspa0000021

Mann, T. C., & Ferguson, M. J. (2017). Reversing implicit first impressions through reinterpretation after a two-day delay. *Journal of Experimental Social Psychology*, *68*, 122–127. 10.1016/j.jesp.2016.06.004

Meier, B. P., Robninson, M. D., & Clore, G. L. (2004). Why good guys wear white. Automatic Inferences about stimulus valence based on brightness. *Psychological Science*, *15*(2), 82–87.

McCarthy, R. J., & Skowronski, J. J. (2011). What will Phil do next?: Spontaneously inferre traits influence predictions of behavior. *Journal of Experimental Social Psychology*, *47*(2), 321–332. 10.1016/j.jesp.2010.10.015

Moskowitz, G. B., & Olcaysoy Okten, I. (2016). Spontaneous Goal Inference (SGI): Goal Inferences. *Social and Personality Psychology Compass*, *10*(1), 64–80. 10.1111/spc3.12232

Moskowitz, G. B., & Roman, R. J. (1992). Spontaneous trait inferences as self-generated primes: Implications for conscious social judgment. *Journal of Personality and Social Psychology*, *62*(5), 728–738. 10.1037/0022-3514.62.5.728

Nadel, L., Hupbach, A., Gomez, R., & Newman-Smith, K. (2012). Memory formation, consolidation and transformation. *Neuroscience and Behavioral Reviews*, *36*, 1640–1645.

Norris, C. J. (2019). The negativity bias, revisited: Evidence from neuroscience measures and an individual differences approach. *Social Neuroscience*, *16*(1), 1–15. 10.1080/17470919.2019.1696225

Olcaysoy Okten, I., & Moskowitz, G. B. (2020). Easy to make, hard to revise: Updating spontaneous trait inferences in the presence of trait-inconsistent information. *Social Cognition*, *38*(6), 571–624. 10.1521/soco.2020.38.6.571

Olcaysoy Okten, I. O., Schneid, E. D., & Moskowitz, G. B. (2019). On the updating of spontaneous impressions. *Journal of Personality and Social Psychology: Attitudes and Social Cognition*, *117*(1), 1–25.

Olivola, C. Y., & Todorov, A. (2010). Elected in 100 milliseconds: Appearance-based trait inferences and voting. *Journal of Nonverbal Behavior*, *34*, 83–110.

Otten, S., & Moskowitz, G. B. (2000). Evidence for implicit evaluative in-group bias: Affect-biased spontaneous trait inference in a minimal group paradigm. *Journal of Experimental Social Psychology*, *36*, 77–89.

Payne, J. D., Ellenbogen, J. M., Walker, M. P., & Stickgold, R. (2008). The role of sleep in memory consolidation. In J. H. Byrne (Ed.), *Learning and memory: A comprehensive reference and concise learning and memory: The editor's selection* (pp. 547–569). Elsevier Press.

Ritts, V., Patterson, M. L., & Tubbs, M. E. (1992). Expectations, impressions, and judgments of physically attractive students: A review. *Review of Educational Research*, *62*(4), 413–426.

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, *5*(4), 296–320. 10.1207/S15327957PSPR0504_2

Schneid, E. D., Carlston, D. E., & Skowronski, J. J. (2015). Spontaneous evaluative inferences and their relationship to spontaneous trait inferences. *Journal of Personality and Social Psychology*, *108*(5), 681–696. 10.1037/a0039118

Schneid, E. D., Crawford, M. T., Skowronski, J. J., Irwin, L. M., & Carlston, D. E. (2015). Thinking about other people: Spontaneous trait inferences and spontaneous evaluations. *Social Psychology*, *46*(1), 24–35. 10.1027/1864-9335/a000218

Shimizu, Y. (2017). Why are negative behaviors likely to be immediately invoked traits? The effects of valence and frequency on spontaneous trait inferences. *Asian Journal of Social Psychology*, *20*(3–4), 201–210. 10.1111/ajsp.12183

Shimizu, Y., Lee, H., & Uleman, J. S. (2017). Culture as automatic processes for making meaning: Spontaneous trait inference. *Journal of Experimental Social Psychology*, *69*, 79–85.

Smith, E. R., & Zárate, M. A. (1992). Exemplar-based model of social judgment. *Psychological Review*, *99*,3–21. 10.1037/0033-295X.99.1.3

Stewart, D. D. (1998). Stereotypes, negativity bias, and the discussion of unshared information in decision-making groups. *Small Group Research*, *29*(6), 643–668.

Stickgold, R., & Walker, M. P. (2007). Sleep-dependent memory consolidation and reconsolidation. *Sleep Med*, *8*(4), 331–343.

Todorov, A., & Bargh, J. A. (2002). Automatic sources of aggression. *Aggression and Violent Behavior*, *7*, 53–68.

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science*, *308*, 1623–1626.

Todorov, A., & Uleman, J. S. (2002). Spontaneous trait inferences are bound to actors' faces: Evidence from a false recognition paradigm. *Journal of Personality and Social Psychology*, *83*(5), 1051–1065. 10.1037//0022-3514.83.5.1051

Uleman, J. S. (1999). Spontaneous versus intentional inferences in impression formation. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 141–160). Guilford.

Uleman, J. S., Hon, A., Roman, R. J., & Moskowitz, G. B. (1996). On-line evidence for spontaneous trait inferenes at encoding. *Personality and Social Psychology Bulletin*, *22*(4), 377–394. 10.1177/0146167296224005

Uleman, J. S., Newman, L. S., & Moskowitz, G. B. (1996). People as flexible interpreters: Evidence and issues from spontaneous trait inference. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 28, pp. 211–279). Elsevier. 10.1016/S0065-2601(08)60239-7

Uleman, J. S., & Saribay, S. A. (2012). Initial impressions of others. In K. Deaux & M. Snyder (Eds.), *Oxford library of psychology. The Oxford handbook of personality and social psychology* (pp. 337–366). Oxford University Press. 10.1093/oxfordhb/9780195398991.013.0014

Uleman, J. S., Saribay, S. A., & Gonzalez, C. M. (2008). Spontaneous inferences, implicit impressions, and implicit theories. *Annual Review of Psychology*, *59*(1), 329–360. 10.1146/annurev.psych.59.103006.093707

Van der Cruyssen, L., Van Duynslaeger, M., Cortoos, A., & Van Overwalle, F. (2009). ERP time course and brain areas of spontaneous and intentional goal inferences. *Social Neuroscience*, *4*(2), 165–184. 10.1080/17470910802253836

Van Overwalle, F., Van Duynslaeger, M., Coomans, D., & Timmermans, B. (2012). Spontaneous goal inferences are often inferred faster than spontaneous trait inferences. *Journal of Experimental Social Psychology*, *48*(1), 13–18. 10.1016/j.jesp.2011.06.016

Walker, M. P., & Stickgold, R. (2004). Sleep-dependent learning and memory consolidation. *Neuron*, *44*(1), 121–133. 10.1016/j.neuron.2004.08.031

Wang, M., & Yang, F. (2017). The malleability of stereotype effects on spontaneous trait inferences: The moderating role of perceivers' power. *Social Psychology*, *48*(1), 3–18. 10.1027/1864-9335/a000288

Wigboldus, D. H. J., Dijksterhuis, A., & Van Knippenberg, A. (2003). When stereotypes get in the way: Stereotypes obstruct stereotype-inconsistent trait inferences. *Journal of Personality and Social Psychology*, *84*(3), 470–484. 10.1037/0022-3514.84.3.470

Wigboldus, D. H., Sherman, J. W., Franzese, H. L., & van Knippenberg, A. (2004). Capacity and comprehension: Spontaneous stereotyping under cognitive load. *Social Cognition*, *22*(3), 292–309.

Willis, J. & Todorov, A. (2006) First impressions. Making up your mind after a 100-ms exposure to a face. *Psychological Science*, *17*(7), 592–598.

Winter, L., & Uleman, J. S. (1984). When are social judgments made? Evidence for the spontaneousness of trait inferences. *Journal of Personality and Social Psychology*, *47*(2), 237.

Zárate, M. A., & Enge, L. (2013). The role of memory consolidation during sleep in social perception and stereotyping. Chapter. In N. Ellemers, D. T. Scheepers, & B. Derks (Eds.), *The neuroscience of prejudice* (pp. 130–145). Psychology Press.

Zárate, M. A., & Smith, E. R. (1990). Person categorization and stereotyping. *Social Cognition*, *8*(2), 161–185.

Zárate, M. A., Uleman, J. S., & Voils, C. I. (2001). Effects of culture and processing goals on the activation and binding of trait concepts. *Social Cognition*, *19*(3: Special issue), 295–323.

Zebrowitz, L. A., Hall, J. A., Murphy, N. A., & Rhodes, G. (2002). Looking smart and looking good: Facial cues to intelligence and their origins. *Personality and Social Psychology Bulletin*, *28*(2), 238–249.

Zebrowitz, L. A. & Montepare, J. M. (2008). Social psychological face perception: Why appearance matters. *Social and Personality Psychology Compass*, *2/3*, 1497–1517.

21 Confronting First Impressions: Motivating Self-Regulation of Stereotypes and Prejudice through Prejudice Confrontation

Kimberly E. Chaney[1], *Diana T. Sanchez*[2], *and Jessica D. Remedios*[3]

[1]*University of Connecticut*
[2]*Rutgers University*
[3]*Tufts University*

In 2020, during a summer marked by large racial justice protests following the murder of George Floyd, Jeff Bezos, CEO of Amazon, encouraged employees to take off Juneteenth, the holiday commemorating the end of slavery in the U.S. workers at an Amazon warehouse in Chicago, who were primarily Black (85% of Amazon's warehouse employees identified as Black in 2014; Demmitt, 2015), were told by management, however, that they would be celebrating Juneteenth with a catered fried chicken and waffles meal (Palmer, 2020). Strong associations between American holidays and food (e.g., Thanksgiving and turkey) may have led the management to want to mark the celebration with a meal. Yet, racist depictions of Black Americans and fried chicken started back in 1915 in the movie *Birth of a Nation* and racist tropes associating Black Americans with fried chicken have continued (Billig, 2001; John, 2014). Thus, an attempt to mark a celebration of particular importance to Black Americans may have led management to consider foods associated with Black Americans, including fried chicken and waffles. This demonstrates automatic stereotype activation: accessibility of knowledge about the association between Black Americans and fried chicken (e.g., Devine, 1989). Moving forward with the planning of the event demonstrates stereotype application: the use of stereotype knowledge in shaping perception and judgement (in the present example, what kind of celebration would be enjoyed; see Gilbert & Hixon, 1991; Kunda & Spencer, 2003). While stereotype activation can shape both explicit and implicit judgements about people (e.g., Devine, 1989; Duncan, 1976; Wigboldus et al., 2003), stereotype activation does not inevitably lead to stereotype application.

How can stereotype application be prevented? An Amazon employee group responded to this incident online saying, "As people throughout the world are rising up against cops, corporations, and this anti-Black capitalist system we live under, Amazon mocks us with this racist form of 'celebration.'" Might this confrontation by the Amazon employee group prevent managers at that

DOI: 10.4324/9781003045687-24

warehouse or other CEOs like Jeff Bezos from applying stereotypical associations in the future? In this chapter, we review research on how and when people self-regulate stereotype activation and application. We focus on research related to *prejudice confrontations*, verbal challenges directed at the person (or persons) who commits a blatant, subtle, or unspoken act of discrimination. We begin by highlighting the difficulties in detecting biases in ourselves and others, before reviewing the affective and cognitive mechanisms that occur after being confronted for using stereotypes. Next, we consider the broader implications of prejudice confrontations for observers and the confronted. Specifically, we consider whether being confronted motivates self-regulation of a person's biases more broadly (e.g., self-regulating racial and gender biases), and whether witnessing the prejudice confrontation of others facilitates self-regulation of one's own biases. Lastly, we explore the limitations of prejudice confrontation including highlighting the need for prejudice confrontation research to take an intersectional approach.

Awareness of Stereotypes

Research suggests prejudice is learned. Stereotypes are automatically activated and applied (Devine, 1989; Devine & Monteith, 1993; Monteith, 1993), and because of just how common the process of unknowingly applying a stereotype is, prejudice becomes a "habit." This habit can be broken, but requires individuals be motivated to do so. Motivation to "break the prejudice habit" involves a critical first step: awareness of bias. Put simply, in order for people to break a prejudice habit they must first be aware of that habit.[1] Awareness of bias has often been identified by individuals' awareness of the discrepancy between how one should behave (e.g., egalitarian thoughts and actions) and how one does behave (i.e., discriminatory thoughts and actions).

This seemingly straightforward first step is, however, difficult to achieve. People generally hold a bias-blind spot; they have an ability to see other people's biases more clearly than their own (Pronin et al., 2002). For example, participants indicated if they had ever engaged in a number of undesirable behaviors, including racist behavior (Bell et al., 2019). Many had. Months later, participants were presented with their own responses but told these were the responses of another student and were then asked to rank if this purported student was more racist than most students, and if they were more racist than most students. Participants consistently indicated that they were less racist than this purported student, *even though they were unknowingly rating their own racist behavior*. If the first step to "break the prejudice habit," is to become aware of one's own biases, intervening is particularly difficult to achieve.

Yet, some people may be more likely to detect their own biases than others. Recent work suggests that concerned awareness of one's own biases against Black Americans is an important individual difference among White Americans (Perry et al., 2015). Specifically, research by Perry et al. (2015) has demonstrated that willingness to recognize one's biases is independent from

one's actual levels of biases. Moreover, White Americans who are more concerned about their biases were more accepting of feedback about their biases and more likely to take action to reduce their biases (Perry et al., 2015). Yet, individuals who more strongly agreed with prejudiced statements also reported that those statements were less prejudiced, and this association was greatest for people who more strongly endorsed egalitarian standards (Fetz & Müller, 2020). This research demonstrates that people who are most biased may be the least aware of what constitutes bias despite believing they are egalitarian. As such, while some people may be more likely to be concerned about their use of stereotypes, they may also have a narrower definition of what counts as prejudice. Together, this research on the bias blind-spot and individual differences in concerns and awareness of one's own biases suggests spontaneous detection of one's own biases are rare and often inaccurate.

Detecting Bias through Prejudice Confrontations

Given that people are unlikely to spontaneously detect their own biases, social psychological research has primarily focused on using experimental introspection to make people aware of their biases (e.g., Monteith, 1993; Monteith et al., 2002). These interventions often present people with hypothetical scenarios and ask them to indicate both how they would behave and how they should behave (e.g., Devine et al., 1991). This reflection of should-would discrepancies serves as an intervention to highlight biases. Yet, the critical component of such an intervention is having someone else ask people to engage in this introspection (i.e., externally forced reflection). Unfortunately, the "should-would" reflection on potential behaviors makes these interventions difficult to implement and a step removed from day-to-day realism making them an unlikely route to increasing awareness outside of the lab.

We believe, however, that prejudice confrontations offer another pathway to increasing introspection that can be implemented in daily social interactions. Importantly, prejudice confrontations have been identified as an effective strategy to reduce the application of stereotypes (Chaney et al., 2015; Chaney & Sanchez, 2018; Chaney et al., 2021; Czopp et al., 2006). For example, White American male participants were confronted for using a negative Black stereotype; they assumed a Black man who wandered the streets was homeless, rather than a tourist. One week later, these participants used fewer negative stereotypes about Black and Latinx Americans when asked to make inferences about a person about whom they had little information compared to participants who had not been confronted (Chaney et al., 2021). Why were participants who were confronted for using stereotypes about Black Americans less likely to use stereotypes seven days later? Research suggests that prejudice confrontations may be uniquely well suited for helping people become aware of, identify, and self-regulate their biases (e.g., Ashburn-Nardo et al., 2008; Chaney & Sanchez, 2018; Monteith et al.,

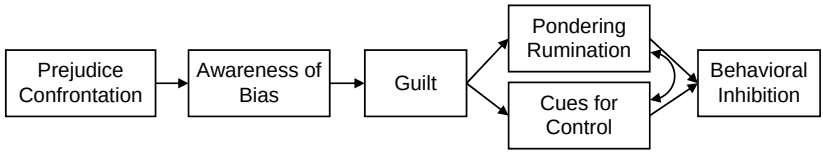

Figure 21.1 Proposed process model of self-regulation of prejudices following a prejudice confrontation.

2002). Indeed, if prejudice is truly a habit, with stereotypes guiding judgements and disparate treatment, an effective prejudice reduction strategy needs to motivate people to change that automatic habit. We outline this process, from prejudice confrontation to behavioral inhibition of biases, in Figure 21.1.

Frequency of Prejudice Confrontations

People are generally reluctant to confront others' biases. For example, despite 81% of women indicating they would confront a hypothetical perpetrator of sexism, only 45% of women did when presented with a real scenario (Swim & Hyers, 1999). Norms of cooperation and "getting along" as well as concerns about the cost of calling out someone's biases (e.g., Good et al., 2012; Shelton & Stewart, 2004) contribute to these relatively low rates of actual, compared to anticipated, prejudice confrontations. Thus, an important avenue for research will be identifying ways to increase individuals' introspection of their own biases and people's willingness to confront others' biases. Changing norms of politeness (Brown, 2015) to norms of speaking out in the face of inequality may therefore be an important step in increasing awareness of bias. Similarly, decreasing the perceived costs of confronting prejudice by increasing acceptance of constructive feedback and identifying optimal ways to confront prejudice (see Chaney & Sanchez, in press) may increase the frequency of prejudice confrontations.

The Role of Guilt in Self-Regulating Stereotypes

If a person's biases have been made evident, research suggests individuals will likely feel negative self-directed affect, sometimes referred to as compunction or guilt (Devine et al., 1991; Monteith, 1993; Monteith et al., 2002). People tend to report feeling guilty after learning about their own biases (Chaney & Sanchez, 2018; Czopp et al., 2006). Threats of being labeled "racist" appear to be an identity threat for White Americans during interracial interactions (Goff et al., 2008; Vorauer et al., 2000). Yet, despite expecting that being confronted would make them feel more angry and less guilty, these affective forecasts do not align with responses to actually being confronted (Kroeper,

2020). As a negative affective state, people are motivated to alleviate their sense of guilt. Indeed, guilt is widely viewed as a critical ingredient for acknowledgement of wrongdoing, acceptance of responsibility, and a desire to improve (Hall & Fincham, 2005; Fisher & Exline, 2010; Woodyatt & Wenzel, 2014). Moreover, people who are more prone to feeling guilt also self-report greater empathy, altruism, and a desire to repair the consequences of one's failures (Cohen et al., 2011). Within the research on prejudice confrontations, research has frequently demonstrated that people confronted for their prejudice report feeling guilty, which is in turn associated with self-regulation of biases in the future, including preventing stereotype activation (Chaney & Sanchez, 2018; Czopp et al., 2006).[2] Being confronted increases perceptions that one's behavior was prejudiced and promotes a desire to self-correct, regardless of initial beliefs about the acceptability of prejudice (Kroeper, 2020).

Classic research suggested people alleviate their guilt by avoiding the source of the guilt (Baumeister et al., 1995) or by correcting the behavior that led to the guilt (Monteith et al., 1993). However, recent work has demonstrated a two-step process of avoidance followed by approach and correction (Amodio et al., 2007; 2008). Social-neuroscience evidence demonstrated that when individuals learned they held negative implicit racial attitudes, they felt guilty (Amodio et al., 2007). This guilt led to the reduction in approach motivations, as indicated by left-sided frontal asymmetry activity reduction (see Coan & Allen, 2003 for review). However, when given a chance to compensate for their racial attitudes by reading newspaper articles about prejudice reduction strategies, neural responses indicated an activation of an approach motivation, as evidenced by increased left-frontal asymmetry activity. Relatedly, a meta-analysis found that guilt was associated with a motivation to constructively approach failure and its consequences (Leach & Cidam, 2015). Thus, people may demonstrate initial avoidance behavior in which they disengage to prevent further interpersonal damage and to employ self-regulation of their biases, but when given the chance, people engage in compensatory behaviors aimed at relieving the guilt and repairing their self-image.

Development of Cues for Control

How exactly does guilt lead to changed attitudes and stereotype use? People experiencing guilt may seek out information about prejudice reduction strategies (Amodio et al., 2007), allowing them to learn about methods for self-regulation approaches, suggesting an effortful, longitudinal effort. Guilt may also create an automatic detection system to prevent future reactions that could elicit guilt. Specifically, guilt may make people more vigilant for situations in which stereotypes may be activated. For example, to avoid feeling additional or future guilt, people may become vigilant for "cues for control," defined as cues which signal that one should regulate the automatic

activation and application of stereotypes (Monteith et al., 2002). Cues for control are stimuli which signal to individuals a potential occasion to evoke a stereotypical impression. For example, learning that a Black student has a scholarship might evoke automatic stereotypes of this student as an athlete for a professor due to Black stereotypes in the United States. However, if the professor has been confronted previously, guilt might activate a "bias alarm bell," and cause them to inhibit the automatic application of an "athlete" stereotype, in order to consider alternative possible attributions for this student (e.g., academic scholarship).

Indeed, research has demonstrated that upon detection of cues for control, people who feel guilty about their biases engage in behavioral inhibition; they briefly pause their behavior in order to prevent a future biased response (Monteith et al., 2002; Moskowitz & Ignarri, 2009; Moskowitz & Li, 2011). Individuals with chronic egalitarian goals (Moskowitz & Ignarri, 2009) or individuals made aware of their racial biases (Monteith et al., 2002; Moskowitz & Li, 2011) "put the brakes on," by responding slower on tasks in which stereotypes may be automatically activated in an effort to mitigate stereotyping. For example, in a study with White American undergraduates, participants completed an inference task meant to elicit the use of negative Black stereotypes (Chaney & Sanchez, 2018). After, half of the participants were confronted by a White experimenter for using a negative Black stereotype, while the other half received no feedback. Measures of guilt were completed and revealed that confronted participants felt more guilty for their responses during the task than participants who were not confronted. Seven days later, participants were contacted to complete an online survey described as an unrelated task. Participants completed a modified probe recognition task. During this task, participants were presented with a series of sentences paired with images of Black men, including some which presented behavior that was stereotypical of Black men in the United States. These stereotypical sentences should have been cues for control, alerting confronted participants that they should self-regulate their biases. After each sentence and image was presented, words (probes) were presented on the screen one at a time, and participants simply had to indicate if the word was in the previous sentence or not. Demonstrating behavioral inhibition after exposure to cues for control, participants who had been confronted one week earlier responded significantly slower when classifying the probes for stereotypical sentences compared to neutral sentences and compared to participants who had not been confronted. This research was the first to demonstrate behavioral inhibition in a classification task following confrontation, demonstrating that confronted participants actually "put the brakes on prejudice" to ensure they did not allow biases to influence their responses.

Critically, more work is needed to better understand if this self-regulation of stereotypes is associated with a meaningful understanding of why their past behavior was prejudiced or merely a suppression of biased responding to mitigate guilt. While self-regulation was associated with greater egalitarian motivation (Chaney & Sanchez, 2018; Chaney et al., 2021), future research

focusing on interracial interactions or anti-racist behavior will provide valuable insights into the motivations underlying this apparent self-regulation.

Understanding the Source and Endurance of Guilt

Despite centering guilt as a key emotional state for self-regulation of stereotypes, little work has sought to better understand the source of guilt following prejudice confrontations. For example, is this guilt derived from a violation of one's own moral standards and self-image, guilt for breaking society's moral standards and being negatively perceived, or guilt about the pain caused by one's biases? Research on motivation to self-regulate biases has focused on two dimensions: internal motivation, being egalitarian because it is personally important, and external motivation, being egalitarian because society has a norm of equality (Plant & Devine, 1998). Yet, when recalling experiences where one's prejudices were present (in behavior or thoughts), people reported feeling guilty, regardless of if they were high or low in internal motivation to regulate biases (Monteith et al., 2010). This suggests that guilt following prejudice confrontation may not be tied simply to a violation of personal (internal) or societal (external) standards of egalitarianism. Rather, we posit that feelings of guilt following a prejudice confrontation may stem from a combination of sources, suggesting a need for greater specificity in guilt assessments. Self-reports of guilt often simply ask how an individual is feeling, without regard to the source of guilt. While people high in self-reflection tendencies (Grant et al., 2002) may be more apt at determining the source of their guilt than others, we suspect greater understanding of the sources of guilt will likely be integral in understanding how people may self-regulate stereotype application.

Notably, some research has posited that being egalitarian is a goal, and once it is achieved, individuals will no longer be vigilant to cues for control (Moskowitz et al., 2011). For example, after writing about a time that they had failed to be egalitarian towards Black men, half of the participants were then given a chance to reaffirm their egalitarian self by writing about a time when they had been egalitarian towards Black men, while the other half of participants wrote a self-affirming paragraph that was not relevant to being egalitarian. Moskowitz et al. (2011) found that participants who had not had a chance to complete an egalitarian goal pursuit (wrote a self-affirmation essay) demonstrated greater attentional bias to faces of Black men in arrays compared to participants who had completed their egalitarian goal pursuit. From this viewpoint, being egalitarian was simply a one-time goal, such that individuals may only be motivated to self-regulate their biases until they can re-affirm their egalitarian self. Yet, we believe time may be needed for one's goal to self-regulate their biases to become internalized, meaning pursuing the goal because of personal importance rather than simply to avoid guilt (Deci & Ryan, 2000).

How Social Norms of Stereotype Use Influence Guilt

Would someone confronted for using a negative stereotype about an arsonist or other widely derogated group feel guilty and self-regulate their stereotype more in the future? Possibly. The normative window theory suggests that some groups are associated with unquestioning social rejection (e.g., drunk drivers, child molesters), while other groups are associated with unquestioning acceptance (e.g., firemen, librarians; Crandall & Warner, 2005; Crandall et al., 2002). For other social groups, the acceptability of prejudice directed towards them varies across contexts and time. Yet, regardless of the manipulated acceptability of negative attitudes towards smokers on a college campus participants felt more guilty when confronted than when they were not confronted for making derogatory comments about smokers (Kroeper, 2020). This suggests that perceived prejudice acceptability in a context may not have a strong effect on motivating stereotype self-regulation. Instead, confrontations may work to shift self-regulation of stereotypes because reactions (i.e., a confrontation) signal the social norm that is most important: the local norm that has been created within the dyad by the confrontation. Put another way, a confrontation may itself signal the acceptability of a prejudice, regardless of broader contextual social norms. Thus, awareness of offending others by using stereotypes may be sufficient to elicit guilt, regardless of perceptions of broader social norms about prejudice acceptability. It is important to note, however, that we would expect confrontations of socially acceptable prejudices to occur less frequently than less socially acceptable prejudices.

How Confronters' Social Identities Influence Confrontations and Guilt

Importantly, some individuals are more prone to detecting and confronting other's prejudice, and social identities are likely one dimension on which prejudice confrontation frequency differs. White Americans are less likely than Black Americans to attribute disparate treatment of Black Americans to discrimination, and men may be less likely than women to attribute disparate treatment of women to sexism (e.g., Fitzgerald & Ormerod, 1991; Hochschild, 1996; Pryor & Day, 1988). This means that White Americans and men are less likely to confront incidents of racism and sexism, respectively, due in part to being less likely to detect discrimination. Yet, perceived (and actual) costs of confronting prejudice are higher for targets (i.e., a woman confronting sexism targeting women) than allies (i.e., a man confronting sexism targeting women) in the form of backlash from a perpetrator (Alt et al., 2019; Good et al., 2012; Shelton & Stewart, 2004), which may impede target confrontation rates. Further, stereotypes about people's own social identities may influence their frequency of confronting bias. For example, men are less likely to confront instances of heterosexism than women, due in part to men's fears of being

labeled gay and women's proximity to gay men due to shared femininity stereotypes (Case et al., in press; Dickter, 2012).

Moreover, research has suggested that a confronter's social identities may change how the prejudiced comment is evaluated by witnesses. For example, men accepted that another man's actions were more sexist when they were confronted by a male observer compared to a female observer (Drury, 2013). Similarly, when White Americans confronted anti-Black bias, the incident was rated as more racist by witnesses compared to when a Black American confronted the racist comment (Rasinski & Czopp, 2010). Importantly, both studies assessed third-party observers' ratings of how offensive the sexist or racist comments were and did not actually examine the rated offensiveness or guilt of the perpetrator. Nevertheless, these findings suggest that because targets of discrimination are seen as complaining or overly sensitive when they confront bias (e.g., Kaiser & Miller, 2001), allies may be seen as more credible when labelling an action discriminatory.

Research that has examined the effect of confronter identity on perpetrators' guilt has been mixed. For example, while some research has found that Black confronters of racial bias elicit less guilt among White Americans than White confronters (Czopp & Monteith, 2003), other research suggested the opposite was true (Czopp et al., 2006), or that no such effect of confronter race emerged (Chaney et al., 2021, supplemental study). As such, it is not clear if a confronter's social identities have a significant effect on confrontation rates or perpetrators' guilt, though it does still appear to be an important factor in evaluation of the confrontation by witnesses.

The Role of Rumination in Self-Regulating Stereotypes

How might guilt promote a path towards prolonged self-regulation of stereotypes? Early work focusing on guilt in response to prejudice confrontations focused on changes in attitudes, behavior, and the use of stereotypes immediately after the confrontation, finding more favorable explicit attitudes and reduced stereotype use by perpetrators immediately after they were confronted due, in part, to guilt (e.g., Czopp et al., 2006; Mallett & Wagner, 2011). Yet, the self-regulation model of prejudice reduction theorized that just as prejudice is a habit that is learned over time, unlearning the habit through self-regulatory control would require prolonged effort. Thus, it was posited that feelings of guilt would promote retrospective reflection, or reflecting on one's past prejudiced behaviors and actions, a self-reflection process that may take time to unfold (Monteith et al., 2002).

When people act in a manner they regret, they often ruminate over their actions. Rumination is broadly conceptualized as recurrent thinking (Martin & Tesser, 1996; Segerstrom et al., 2003) about negative feelings and their causes, meanings, and consequences (Nolen-Hoeksema et al., 2008). Rumination has often been identified as a response to guilt in clinical psychology (Tangney & Fischer, 1995) and is a common cognition following

negative moods or events (Robinson & Alloy, 2003). Though rumination is most often studied as it relates to depression and anxiety (e.g., Robinson & Alloy, 2003), there are different types of rumination. Brooding rumination focuses on passive comparison of one's state compared to an unachieved standard, while reflective pondering rumination is a "purposeful turning inward to engage in cognitive problem-solving" (Treynor et al., 2003). This pondering rumination is centered on higher-level causes and consequences (e.g., why do I feel this way? why did this happen?), can be beneficial in generating plans during problem solving (Williams, 1996), and may be adaptive, promoting positive emotional regulation (Watkins, 2004; Watkins & Moulds, 2005). Further, people who ruminate about a trauma and its implications experience personal growth (Tedechi & Calhoun, 2004; Ullrich & Lutgendorf, 2002), and rumination is associated with greater active behavioral problem-solving efforts (Szabo & Lovibond, 2004; 2006). From a goal pursuit perspective, the goal-progress model of rumination (Martin & Tesser, 1996; 2006) has suggested that rumination occurs when unsatisfactory progress has been made towards goal completion. Importantly, ruminating on one's goals was associated with greater goal progress over a one-month period (Moberly & Dickson, 2016). Together, this research suggests that rumination can be an important form of self-reflection that may facilitate growth and self-regulation of stereotype application.

To determine if guilt promoted rumination that was critical to prolonged self-regulation of prejudice, we sought to examine this process in our work described earlier (Chaney & Sanchez, 2018). After assessing behavioral inhibition via a modified probe recognition task (described earlier), we asked participants how much they had been thinking about their responses in the last week, including if they had been thinking about their guilt. We found that participants who had been confronted immediately reported greater guilt than participants who had not been confronted. Importantly, these same participants also reported that they had spent more time ruminating on the experience in the lab during the last week, and this rumination was in turn associated with greater behavioral inhibition one week after the confrontation.

Understanding the Source and Onset of Rumination

Because rumination is a self-reflective process, it likely does not lead to behavioral change right away. Indeed, in our work, White American men who were confronted did not report differences in rumination 24–72 hours after the confrontation compared to White American men who were not confronted (Chaney et al., 2021). These confronted White American men did, however, still use fewer stereotypes compared to their un-confronted counterparts. It is possible, then, that guilt is sufficient to produce self-regulation in a short timeline after a confrontation, but that rumination may be important for solidifying and maintaining self-regulation over a longer timeline such as seven days, including in establishing and reinforcing cues for control.

Further examination of how and when guilt translates into pondering rumination over longer durations will be integral in better understanding how people can effectively self-regulate stereotypes.

Just as questions about the source of guilt are critical, we similarly believe the focus of the rumination is important. For example, pondering rumination could take the form of pondering about where one's biases come from or why the confronter chose to call them out. Some research suggests that rumination following a confrontation might focus on one's own biases and guilt. Individuals informed of their biases reported thinking about their biases and prejudice-related guilt more frequently than individuals who were not made aware of their biases (Monteith et al., 2002), though more work is needed to clearly identify the focus of such rumination. Yet, some individuals may be more prone to negative brooding rumination after being confronted, focusing instead simply on them being confronted and the negative affect associated with the experience, never moving on towards more productive self-reflection. The focus of this rumination may be critical in predicting a person's future use of stereotypes as well as whether they choose to become a confronter of bias. Specifically, people may seek to make amends for their own bias by calling out other's bias in an effort to mitigate the prejudice cycle. People who have themselves been confronted may be poised to be especially effective confronters as they have experienced a prejudice confrontation, potentially developing a script or "how to" on confronting. As such, we believe rumination is a critical factor in the enduring effect of prejudice confrontations on motivation to self-regulate stereotyping.

The Reach of Prejudice Confrontations across People and Biases

The Effect of Witnessing a Prejudice Confrontation

Could witnessing a confrontation as a neutral observer lead to the use of fewer stereotypes? One might assume no, as witnesses should not feel guilty for something they did not do. Yet, research on collective guilt suggests otherwise. Collective guilt is guilt due to the belief that one's group is responsible for an immoral act against another group (Doosje et al., 1998; Leach et al., 2002). This collective guilt can arise from "guilt by association" when one's membership in a social group that has harmed others becomes salient. For example, Dutch students reminded that the Dutch had abused and killed Indonesians during colonization in the past did not report feeling personally guilty for these wrongdoings but did report feeling guilty that the perpetrators were from their social ingroup (Doosje et al., 1998). Other research on collective guilt has found that individuals can experience guilt for current wrongdoings of their ingroup. White American participants reported feeling more guilty when they believed that White Americans discriminated against other social groups, though not when they simply read about Black

Americans facing racial discrimination (Iyer et al., 2003; Leach et al., 2002). This suggests that to elicit a collective group, people must focus on a perpetrator who has done wrong, whether directly, or as part of a social group, not merely that a wrong has occurred (see Iyer et al., 2003 for review). It is worth noting that collective guilt is likely a rare emotion (McGarty et al., 2002; Leach, 2002) because people may ignore or deny the group's responsibility for inequality (see Knowles et al., 2014).

But, when people do experience collective guilt, does it translate to self-regulation of stereotypes? Research on collective guilt has found that White Americans who experience collective guilt due to White Americans' discrimination against Black Americans were more supportive of affirmative action hiring in college admissions (Iyer et al., 2003; Swim & Miller, 1999). Indirect evidence of the role of collective guilt has also been demonstrated in the literature on prejudice confrontations. Hyers (2010) found that people who witnessed a confrontation of heterosexism subsequently made fewer anti-gay remarks than before the confrontation occurred. Similarly, reading about a confrontation of sexism in a classroom decreased participants' own reported sexism (Boysen, 2013). More research will be needed to better understand if these changes in behavior after witnessing a prejudice confrontation are due to collective guilt or social norm influence. For example, seeing someone openly condemn prejudice communicates social norm information (Blanchard et al., 1991, 1994; Boysen, 2013; Monteith et al., 1996) which may spread throughout social networks (Paluck, 2011; Paluck & Shepherd, 2012; Stangor et al., 2001). Indeed, Case (2012) found that White women committed to anti-racism indicated they would be more likely to confront prejudice if another bystander spoke up first because of social norms against racism being made salient. In contrast, seeing someone condone prejudice or not speak up in the face of prejudice encourages prejudice expression (Blanchard et al., 1994; Burkley et al., 2016; Jewell et al., 2015). While this research on witnessing prejudice confrontations does not measure collective guilt, we believe future research in this area will be important for establishing the importance of guilt in motivating self-regulation of stereotypes.

The Broad Self-Regulation of Stereotypes

Prejudice reduction strategies informed by social psychology theories often examine interventions that reduce the use of stereotypes directed towards *one* stigmatized group (e.g., Lai et al., 2016). This focus on stereotyping one social group goes against evidence that prejudices are generalized, such that someone who is prejudiced against one devalued social group is more likely to hold biases against multiple devalued groups (Allport, 1954; Duckitt & Sibley, 2007; Ekehammar & Akrami, 2003; Osborne et al., 2020). For example, individuals who hold negative attitudes towards Black people also hold negative attitudes towards other devalued social groups (e.g., Latinos,

women; Duckitt & Sibley, 2007; Sibley & Duckitt, 2008). Moreover, prejudice towards derogated social groups such as women and Black Americans stems from an underlying ideology supporting social hierarchies and inequality (Sidanius & Pratto, 1999; Osborne et al., 2020).

Our work has suggested that people also generally endorse a belief that prejudices co-occur in perpetrators, a belief we have called *lay theory of generalized prejudice*. For example, Black and Latina women believed that someone who held negative attitudes towards women would also hold negative attitudes towards their racial group (Chaney et al., in press a). Similarly, White men and women perceived someone who was racist would also be sexist, as did Black and Latino men (Sanchez et al., 2017). Together, our body of work has found that a lay theory of generalized prejudice can also influence individuals' cardiovascular stress when anticipating interacting with a perpetrator (Chaney et al., in press b), working memory in STEM contexts (Chaney et al., 2016), and anticipated inclusion or exclusion at organizations (Chaney et al., 2016; Sanchez et al., 2018).

Yet, this work primarily focused on the extent to which people perceived prejudices to go hand-in-hand *in other people*. We hypothesized that this belief could also be applied to one's own attitudes. Indeed, researchers have previously proposed that changing attitudes towards one devalued social group should result in changed attitudes towards another similarly devalued social group, referred to as secondary transfer effects (Pettigrew, 1997; Pettigrew, 2009). For example, when people had a positive interaction with an immigrant, they reported more positive explicit attitudes towards immigrants and a secondary outgroup, sexual minorities (Schmid et al., 2012). Thus, mitigating prejudices towards one group may provide an avenue to promote broader intergroup attitude change if the intervention promotes broad self-regulation of biases.

We tested the effect of prejudice confrontations on self-regulation of biases towards similarly stigmatized social groups; specifically, whether efforts to reduce stereotyping of one group may reduce the use of stereotypes of other stigmatized groups (Chaney et al., 2021). In one study, White American men engaged in an online conversation with, they believed, another White American man, but who was actually a pre-programmed set of responses. Participants responded to a number of moral dilemmas and provided and received feedback from the other participant. On one, participants read about an issue at a hospital involving a nurse. Half of our participants then received a response from their interaction partner saying, "I noticed you referred to the nurse as a 'she.' The nurse could also be a man. We shouldn't use stereotypies, you know?" The other half of the participants simply received neutral feedback. 24 hours later, we posted a second survey available only to these same participants but made no connection between the Time 1 survey and the present. During this second part, participants completed an inference-making paradigm, including trials that were meant to elicit

negative stereotypes about Black and Latino men. Specifically, at Time 2, participants were asked to make inferences about individuals based on only a picture and a descriptive sentence (e.g., This person spends time at shelters). The White American men who had been confronted for using a female gender role stereotype (referring to the nurse as a "she") 24–72 hours earlier, went on to use significantly fewer negative stereotypes about Black and Latino men (e.g., were less likely to say that a Latino man who spends time at shelters is homeless, and more likely to say he was a volunteer). In similar studies, we also found participants confronted for using negative Black stereotypes, compared to participants who had not been confronted, subsequently used fewer negative stereotypes about Latino Americans, female gender role stereotypes, and even positive Black stereotypes (assuming a Black student was on an athletic scholarship, not an academic scholarship).

While confronted participants reported feeling more guilty than participants who had not been confronted, this increase in guilt did not account for subsequent self-regulation of stereotyping. Guilt may therefore only guide self-regulation of biases for the congruent social group (the bias for which one was confronted for) and may not account for this spreading self-regulation of biases. Indeed, we found that when confronted for using a female gender role stereotype, participants' guilt was associated with greater reported racial egalitarian motivation 24–72 hours after confrontation, which in turn led to the use of fewer racial stereotypes compared to participants who were not confronted. Guilt, therefore, may spark broad egalitarian motivation, and this may be most likely among participants who endorse a lay theory of generalized prejudice. Notably, the discussed work focused on the self-regulation of stereotypes about similarly stigmatized social groups. For example, we do not expect that being confronted for using a negative Black stereotype would decrease the use of stereotypes about White Americans (who hold high status; Zou & Cheryan, 2017) or other groups who do not face similar stigma. We encourage future research to explore these important questions about the boundaries of prejudice confrontations' breadth.

A Need to Consider Prejudice Confrontations through an Intersectional Framework

Research on how prejudice confrontations can promote self-regulation of stereotypes has been limited, nearly always confronting sexism directed at White women or racism directed at Black men (e.g., Chaney & Sanchez, 2018; Chaney et al., 2021; Czopp et al., 2006). Thus, this area of research has rarely taken an intersectional approach (for an exception, see Dupree, this volume): examining and understanding how systems of oppression (e.g., racism, sexism, heterosexism) are interrelated (Crenshaw, 1989; Moradi & Grzanka, 2017). From an intersectional framework, people's multiple social identities (e.g., race, gender, sexual orientation) are also interrelated, creating meaning from one

another and reinforcing systems of inequality (Crenshaw, 1989). For example, stereotypes about Black women differ from stereotypes about White women and Black men (Ghavami & Peplau, 2013), and research on prejudice confrontations could explore if people confronted for making prejudiced statements about Black men subsequently self-regulate their use of stereotypes about Black women. Relatedly, the intersectional invisibility hypothesis has demonstrated that the experiences of women of color and multiple-stigmatized individuals are often erased or ignored (Cole, 2009; Remedios & Snyder, 2018; Sesko & Biernat, 2010). As such, people may have an even harder time detecting biases towards, or considering the discrimination faced by, women of color. Prejudice associated with invisibility stigma may ultimately be the least likely to be confronted due to lack of detection (see Neel & Lassetter, 2019). Even when confronted, prejudices associated with not seeing, hearing, or acknowledging stigmatized social groups in some contexts may similarly be less likely to evoke guilt as such biases are not prototypical forms of discrimination (e.g., Sommers & Norton, 2006).

Further, research on double jeopardy has demonstrated that people with multiple stigmatized social identities face additional discrimination compared to people with only one stigmatized social identity (Beale, 1970; Berdahl & Moore, 2006). Relatedly, women of color may attribute disparate treatment to either racism or sexism depending on the perpetrator (Remedios et al., 2020; Remedios & Snyder, 2015), and may infer that someone who is racist will discriminate against them for both their racial and gender identities (Chaney et al., in press a). This may create attributional ambiguity, such that people may not know whether to confront a perpetrator for being racist, sexist, or both, and may also provide a pathway for the perpetrator to deny being biased due to such attributional ambiguity.

Lastly, we believe it is important to note that the body of research on prejudice confrontations to date has primarily relied on the theoretical view of prejudice as a habit that needs to be broken (e.g., Chaney & Sanchez, 2018; Czopp et al., 2006; Monteith, 1993). While a predominant and well supported theory in social psychology, this framing may downplay the ways in which stereotypes are at times motivated inferences that aim to maintain current systems of oppression. Specifically, stereotypes can be used to reinforce and legitimize White Americans' positions of power and status by denigrating Black Americans (e.g., Collins, 2000). Future research on prejudice confrontations may be enriched by examining if a lay theory that prejudice is a habit versus a motivated cognition influence rates of confronting prejudice or an individual's response to being confronted. Overall, we believe research on prejudice confrontations and self-regulation of stereotypes will be greatly enriched by taking an intersectional perspective that considers how systems of oppression are interrelated (for additional hypothesis on prejudice confrontations derived from an intersectional perspective, see Remedios & Akhtar, 2019).

Conclusion

Prejudice confrontations make evident the confronted individual's use of stereotypes, a critical first step in promoting motivation to reduce biases (Monteith, 1993; Monteith et al., 2002). Moreover, confronted individuals report feeling guilty (Chaney & Sanchez, 2018; Czopp et al., 2006). Because people are motivated to alleviate guilt, they seek out ways to apologize and compensate (Mallett & Wagner, 2011). Additionally, this guilt develops cues for control, signaling to people when they may be at risk of applying stereotypes (Monteith et al., 2002). When encountering cues for control, people engage in behavioral inhibition to mitigate stereotype use (Monteith et al., 2002). Over time, people may engage in productive pondering rumination, which may lead to the reinforcement of cues for control and long-term behavioral inhibition (Chaney & Sanchez, 2018). Lastly, prejudice confrontations lead to broad self-regulation of biases, though more research is needed to better understand the underlying processes of this broad self-regulation (Chaney et al., 2021).

Importantly, this multistage process of self-regulating prejudice relies on the activation of guilt. The consequences of the elicitation of guilt are complex, but all work in tandem to reduce the expression of prejudice. Guilt arising from prejudice confrontations does not seem to be directly tied to the norms of prejudice expression (Kroeper, 2020). Instead, a prejudice confrontation may be sufficient to signal that the perpetrator violated the local norms. In doing so, prejudice confrontations may ultimately serve as the signals of norms in a specific context. In conjunction with signaling broader norms, witnessing a prejudice confrontation may create a feeling of collective guilt among observers who are part of the dominant social group. Collective guilt, much like individual guilt, is associated with a desire for restorative justice (Iyer et al., 2003). Prejudice confrontations may therefor motivate self-regulation of stereotype use in the broader context and among observers.

Important questions remain about prejudice confrontations and the role of guilt in self-regulation. For example, more research is needed to better understand how the social identities of a confronter influences the experiences of guilt and self-regulation of perpetrators (Czopp & Monteith, 2003; Czopp et al., 2006). Similarly, we believe greater specification about the source of guilt and rumination following a prejudice confrontation is needed. This specificity will likely influence research on prejudice confrontation styles (Chaney & Sanchez, in press; for review, see Monteith et al., 2019). Lastly, we believe future work integrating an intersectional framework will greatly enrich the field's understanding of prejudice confrontations (see Remedios & Akhtar, 2019).

Together, this body of research highlights the importance of affect in motivating self-regulation of stereotypes. The discomfort of guilt is important in making people slow down and resist the automatic application of stereotypes, and prejudice confrontations appear to be an effective way to elicit

guilt. Prejudice confrontations may be an underutilized strategy to combat prejudice, but we believe a better conceptual understanding of guilt as not merely a negative emotion, but a motivating emotion, may be integral in promoting prejudice confrontations.

Notes

1 While awareness is a necessary first step, it is not sufficient. Awareness of one's biases can have the unintended effect of making people more resistant to self-regulation if it makes them angry, fearful, or defensive (see Vitriol & Moskowitz, 2021). While these affective responses have primarily been documented in response to feedback on one's performance on the Implicit Association Test, research on prejudice confrontations has rarely explored these responses.
2 Notably, research on making people aware of their implicit biases has found that people can react with anger or defensiveness, rather than guilt (Howell & Ratliff, 2017; Vitriol & Moskowitz, 2021). Anger and defensiveness have been found to diminish self-regulation of biases.

References

Allport, G. W. (1954). *The nature of prejudice*. Addison-Wesley.

Alt, N. P., Chaney, K. E., & Shih, M. J. (2019). 'But that was meant to be a compliment!' Evaluative costs of confronting positive racial stereotypes. *Group Processes & Intergroup Relations*, *22*(5), 655–672.

Amodio, D. M., Devine, P. G., & Harmon-Jones, E. (2007). A dynamic model of guilt implications for motivation and self-regulation in the context of prejudice. *Psychological Science*, *18*(6), 524–530.

Amodio, D. M., Devine, P. G., & Harmon-Jones, E. (2008). Individual differences in the regulation of intergroup bias: The role of conflict monitoring and neural signals for control. *Journal of Personality and Social Psychology*, *94*(1), 60–74.

Ashburn-Nardo, L., Morris, K. A., & Goodwin, S. A. (2008). The confronting prejudiced responses (CPR) model: Applying CPR in organizations. *Academy of Management Learning & Education*, *7*(3), 332–342.

Beale, F. M. (1970). Double jeopardy: To be Black and female. In T. Cade (Ed.), *The Black woman* (pp. 90–100). New American Library.

Baumeister, R. F., Stillwell, A. M., & Heatherton, T. F. (1995). Personal narratives about guilt: Role in action control and interpersonal relationships. *Basic and Applied Social Psychology*, *17*(1–2), 173–198.

Bell, A. C., Burkley, M., & Bock, J. (2019). Examining the asymmetry in judgments of racism in self and others. *The Journal of Social Psychology*, *159*(5), 611–627.

Berdahl, J. L., & Moore, C. (2006). Workplace harassment: Double jeopardy for minority women. *The Journal of Applied Psychology*, *91*, 426–436.

Billig, M. (2001). Humour and hatred: The racist jokes of the Ku Klux Klan. *Discourse & Society*, *12*(3), 267–289.

Blanchard, F. A., Lilly, T., & Vaughn, L. A. (1991). Reducing the expression of racial prejudice. *Psychological Science*, *2*(2), 101–105.

Blanchard, F. A., Crandall, C. S., Brigham, J. C., & Vaughn, L. A. (1994). Condemning and condoning racism: A social context approach to interracial settings. *Journal of Applied Psychology*, *79*(6), 993–997.

Boysen, G. A. (2013). Confronting math stereotypes in the classroom: Its effect on female college students' sexism and perceptions of confronters. *Sex Roles*, 69(5–6), 297–307.

Brown, P. (2015). Politeness and language. In *The International Encyclopedia of the Social and Behavioural Sciences (IESBS)* (2nd ed., pp. 326–330). Elsevier.

Burkley, M., Andrade, A., & Burkley, E. (2016). When using a negative gender stereotype as an excuse increases gender stereotyping in others. *The Journal of Social Psychology*, *156*(2), 202–210.

Case, K. A. (2012). Discovering the privilege of whiteness: White women's reflections on anti-racist identity and ally behavior. *Journal of Social Issues*, *68*(1), 78–96.

Case, K. A., Rios, D., Lucas, A., Braun, K., & Enriquez, C. (in press). Intersectional patterns of prejudice confrontation by White, heterosexual, and cisgender allies. *Journal of Social Issues*. Advanced online publication.

Chaney, K. E. & Sanchez, D. T. (in press). Prejudice confrontation styles scale: A validated and reliable measure of how people confront prejudice. *Group Processes and Intergroup Relations*. Advanced online publication. https://doi-org.ezproxy.lib.uconn.edu/10.1177/13684302211005841

Chaney, K. E. & Sanchez, D. T. (2018). The endurance of interpersonal confrontations as a prejudice reduction strategy. *Personality and Social Psychology Bulletin*, *44*(3), 418–429.

Chaney, K. E., Sanchez, D. T., Alt, N. P., & Shih, M. (2021). The breadth of confrontations as a prejudice reduction strategy. *Social Psychological and Personality Science*, *12* (3), 314–322.

Chaney, K. E., Sanchez, D. T., Himmelstein, M. S., & Manuel, S. K. (in press b). Lay theory of generalized prejudice moderates cardiovascular stress responses to racism for White women. *Group Processes and Intergroup Relations*. Advanced online publication.

Chaney, K. E., Sanchez, D. T., & Remedios, J. D. (in press a). Dual cues: Women of color anticipate both gender and racial bias in the face of a single identity cue. *Group Processes and Intergroup Relations*. Advanced online publication.

Chaney, K. E., Sanchez, D. T., & Remedios, J. D. (2016). Organizational identity safety cue transfers. *Personality and Social Psychology Bulletin*, *42*(11), 1564–1576.

Chaney, K.E., Young, D.M., & Sanchez, D.T. (2015). Confrontation's Health Outcomes and Promotion of Egalitarianism (C-HOPE) Framework. *Translational Issues in Psychological Science*, *1*(4), 363–371.

Cohen, T. R., Wolf, S. T., Panter, A. T., & Insko, C. A. (2011). Introducing the GASP scale: A new measure of guilt and shame proneness. *Journal of Personality and Social Psychology*, *100*, 947–966.

Coan, J.A., & Allen, J.J.B. (2003). The state and trait nature of frontal EEG asymmetry in emotion. In K. Hugdahl, & R. J. Davidson (Eds.), *The asymmetrical brain*, (pp. 565–615). MIT Press.

Cole, E. R. (2009). Intersectionality and research in psychology. *The American Psychologist*, 64, 170–180.

Crandall, C. S., Eshleman, A., & O'Brien, L. (2002). Social norms and the expression and suppression of prejudice: The struggle for internalization. *Journal of Personality and Social Psychology*, *82*(3), 359–378.

Collins, P. H. (2000). *Black feminist thought: Knowledge, consciousness, and the politics of empowerment*. Routledge.

Crandall, C. S., & Warner, R. H. (2005). How a prejudice is recognized. *Psychological Inquiry*, *16*(2/3), 137–141.

Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: A Black feminist critique of antidiscrimination doctrine, feminist theory, and antiracist politics. *University of Chicago Legal Forum*, *1989*, 139–167.

Czopp, A. M., & Monteith, M. J. (2003). Confronting prejudice (literally): Reactions to confrontations of racial and gender bias. *Personality and Social Psychology Bulletin*, *29*(4), 532–544. 10.1177/0146167202250923

Czopp, A. M., Monteith, M. J., & Mark, A. Y. (2006). Standing up for a change: Reducing bias through interpersonal confrontation. *Journal of Personality and Social Psychology*, *90*(5), 784–803. 10.1037/0022-3514.90.5.784

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, *11*(4), 227–268.

Demmitt, J. (2015, June 11). *85 percent of Amazon's black U.S. workers hold unskilled jobs*. Puget Sound Biz Journal. https://www.bizjournals.com/seattle/blog/techflash/2015/06/85-percent-of-amazon-s-black-u-s-workers-hold.html

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*(1), 5–18.

Devine, P. G., & Monteith, M. J. (1993). The role of discrepancy-associated affect in prejudice reduction. In *Affect, cognition and stereotyping* (pp. 317–344). Academic Press.

Devine, P. G., Monteith, M. J., Zuwerink, J. R., & Elliot, A. J. (1991). Prejudice with and without compunction. *Journal of Personality and Social Psychology*, *60*(6), 817–830.

Dickter, C. L. (2012). Confronting hate: Heterosexuals' responses to anti-gay comments. *Journal of Homosexuality*, *59*(8), 1113–1130.

Doosje, B., Branscombe, N. R., Spears, R., & Manstead, A. S. R. (1998). Guilty by association: When one's group has a negative history. *Journal of Personality and Social Psychology*, *75*, 872–886.

Drury, B. J. (2013). Confronting for the greater good: Are confrontations that address the broad benefits of prejudice reduction taken seriously? Doctoral Dissertation, University of Washington.

Duckitt, J., & Sibley, C. G. (2007). Right wing authoritarianism, social dominance orientation and the dimensions of generalized prejudice. *European Journal of Personality*, *21*(2), 113–130.

Duncan, B. L. (1976). Differential social perception and attribution of intergroup violence: Testing the lower limits of stereotyping of blacks. *Journal of Personality and Social Psychology*, *34*(4), 590–598.

Ekehammar, B., & Akrami, N. (2003). The relation between personality and prejudice: A variable-and a person-centered approach. *European Journal of Personality*, *17*(6), 449–464.

Fetz, K., & Müller, T. S. (2020). Is one's own ethnic prejudice always subtle? The inconsistency of prejudice endorsement and prejudice awareness depends on self-related egalitarian standards and motivations. *Basic and Applied Social Psychology*, *42*(1), 1–28.

Fisher, M. L., & Exline, J. J. (2010). Moving toward self-forgiveness: Removing barriers related to shame, guilt, and regret. *Social and Psychology Compass*, *4*, 548–558.

Fitzgerald, L., & Ormerod, A. J. (1991). Perceptions of sexual harassment: The influence of gender and context. *Psychology of Women Quarterly*, *15*, 281–294.

Ghavami, N., & Peplau, L. A. (2013). An intersectional analysis of gender and ethnic stereotypes: Testing three hypotheses. *Psychology of Women Quarterly*, *37*(1), 113–127.

Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology*, *60*, 509 –517. 10.1037/0022-3514.60.4.509

Goff, P. A., Steele, C. M., & Davies, P. G. (2008). The space between us: Stereotype threat and distance in interracial contexts. *Journal of Personality and Social Psychology*, *94*(1), 91–107. 10.1037/0022-3514.94.1.91

Good, J. J., Moss-Racusin, C. A., & Sanchez, D. T. (2012). When do we confront? Perceptions of costs and benefits predict confronting discrimination on behalf of the self and others. *Psychology of Women Quarterly*, *36*(2), 210–226.

Grant, A. M., Franklin, J., & Langford, P. (2002). The self-reflection and insight scale: A new measure of private self-consciousness. *Social Behavior and Personality: An International Journal*, *30*(8), 821–835.

Hall, J. H., & Fincham, F. D. (2005). Self-forgiveness: The stepchild of forgiveness research. *Journal of Social and Clinical Psychology*, *24*, 621–637.

Hochschild, J. L. (1996). *Facing up to the American dream: Race, class, and the soul of the nation* (Vol. 51). Princeton University Press.

Hyers, L. L. (2010). Alternatives to silence in face-to-face encounters with everyday heterosexism: Activism on the interpersonal front. *Journal of Homosexuality*, *57*(4), 539–565.

Howell, J. L., & Ratliff, K. A. (2017). Not your average bigot: The better-than-average effect and defensive responding to Implicit Association Test feedback. *British Journal of Social Psychology*, *56*(1), 125–145.

Iyer, A., Leach, C. W., & Crosby, F. J. (2003). White guilt and racial compensation: The benefits and limits of self-focus. *Personality and Social Psychology Bulletin*, *29*(1), 117–129.

Jewell, J., Spears Brown, C., & Perry, B. (2015). All my friends are doing it: Potentially offensive sexual behavior perpetration within adolescent social networks. *Journal of Research on Adolescence*, *25*(3), 592–604. 10.1111/jora.12150

John, A. (2014, February 6). *Making fried chicken and watermelon racist*. The Atlantic. https://www.theatlantic.com/national/archive/2014/02/heres-why-your-fried-chicken-and-watermelon-lunch-racist/357814/

Kaiser, C. R., & Miller, C. T. (2001). Stop complaining! The social costs of making attributions to discrimination. *Personality and Social Psychology Bulletin*, *27*(2), 254–263.

Knowles, E. D., Lowery, B. S., Chow, R. M., & Unzueta, M. M. (2014). Deny, distance, or dismantle? How White Americans manage a privileged identity. *Perspectives on Psychological Science*, *9*(6), 594–609.

Kroeper, K. M. (2020). Expecting prejudice confrontation to backfire: prejudice norms and misalignment between forecaster expectations and experiencer realities. Doctoral Dissertation, University of Indiana.

Kunda, Z., & Spencer, S. J. (2003). When do stereotypes come to mind and when do they color judgment? A goal-based theoretical framework for stereotype activation and application. *Psychological Bulletin*, *129*, 522– 544. 10.1037/0033-2909.129.4.522

Lai, C. K., Skinner, A. L., Cooley, E., Murrar, S., Brauer, M., Devos, T., … & Simon, S. (2016). Reducing implicit racial preferences: II. Intervention effectiveness across time. *Journal of Experimental Psychology: General*, *145*(8), 1001–1016.

Leach, C. W. (2002). Democracy's dilemma: Explaining racial inequality in egalitarian societies. *Sociological Forum*, *17* (4), 681–690.

Leach, C. W., & Cidam, A. (2015). When is shame linked to constructive approach orientation? A meta-analysis. *Journal of Personality and Social Psychology*, *109*(6), 983–1002.

Leach, C. W., Snider, N., & Iyer, A. (2002). "Poisoning the consciences of the fortunate": The experience of relative advantage and support for social equality. In I. Walker & H. J. Smith (Eds.), *Relative deprivation: Specification, development, and integration* (pp. 136–163). Cambridge University Press.

Mallett, R. K., & Wagner, D. E. (2011). The unexpectedly positive consequences of confronting sexism. *Journal of Experimental Social Psychology*, *47*(1), 215–220.

Martin, L. L., & Tesser, A. (1996). Some ruminative thoughts. *Advances in Social Cognition*, *9*, 1–47.

Martin, L. L., & Tesser, A. (2006). Extending the Goal Progress Theory of Rumination: Goal Reevaluation and Growth. In L. J. Sanna & E. C. Chang (Eds.), *Judgments over time: The interplay of thoughts, feelings, and behaviors* (pp. 145–162). Oxford University Press.

McGarty, C., Pedersen, A., Leach, C.W. Mansell, T., Waller, J., & Bliuc, A.-M. (2002). *Collective guilt as a predictor of commitment to apology*. Unpublished Manuscript, Australian National University.

Moberly, N. J., & Dickson, J. M. (2016). Rumination on personal goals: Unique contributions of organismic and cybernetic factors. *Personality and Individual Differences*, *99*, 352–357.

Monteith, M. J. (1993). Self-regulation of prejudiced responses: Implications for progress in prejudice-reduction efforts. *Journal of Personality and Social Psychology*, *65*(3), 469–485.

Monteith, M. J., Devine, P. G., & Zuwerink, J. R. (1993). Self-directed versus other-directed affect as a consequence of prejudice-related discrepancies. *Journal of Personality and Social Psychology*, *64*(2), 198–210.

Monteith, M. J., Ashburn-Nardo, L., Voils, C. I., & Czopp, A. M. (2002). Putting the brakes on prejudice: On the development and operation of cues for control. *Journal of Personality and Social Psychology*, *83*(5), 1029–1050.

Monteith, M. J., Burns, M. D., & Hildebrand, L. K. (2019). Navigating successful confrontations. In R. K. Mallett & M. J. Monteith (Eds.), *Confronting prejudice and discrimination: The science of changing minds and behaviors* (pp. 225–248). Academic Press.

Monteith, M. J., Deneen, N. E., & Tooman, G. D. (1996). The effect of social norm activation on the expression of opinions concerning gay men and Blacks. *Basic and Applied Social Psychology*, *18*(3), 267–288.

Monteith, M. J., Mark, A. Y., & Ashburn-Nardo, L. (2010). The self-regulation of prejudice: Toward understanding its lived character. *Group Processes & Intergroup Relations*, *13*(2), 183–200.

Moskowitz, G. B., & Ignarri, C. (2009). Implicit volition and stereotype control. *European Review of Social Psychology*, *20*(1), 97–145.

Moradi, B., & Grzanka, P. R. (2017). Using intersectionality responsibly: Toward critical epistemology, structural analysis, and social justice activism. *Journal of Counseling Psychology*, *64*(5), 500–513.

Moskowitz, G. B., & Li, P. (2011). Egalitarian goals trigger stereotype inhibition: A proactive form of stereotype control. *Journal of Experimental Social Psychology*, *47*(1), 103–116.

Moskowitz, G. B., Li, P., Ignarri, C., & Stone, J. (2011). Compensatory cognition associated with egalitarian goals. *Journal of Experimental Social Psychology*, *47*(2), 365–370.

Neel, R., & Lassetter, B. (2019). The stigma of perceived irrelevance: An affordance-management theory of interpersonal invisibility. *Psychological Review*, *126*(5), 634–659.

Nolen-Hoeksema, S., Wisco, B. E., & Lyubomirsky, S. (2008). Rethinking rumination. *Perspectives on Psychological Science*, *3*(5), 400–424.

Osborne, D., Satherley, N., Little, T.D., & Sibley, C. G. (2020). Authoritarianism and social dominance predict annual increases in generalized prejudice. *Social Psychological & Personality Science*, *12*(7), 1136–1145.

Palmer, A. (2020, June 19). *Amazon workers in Chicago angered by 'tokenized' Juneteenth celebration offering chicken and waffles*. CNBC. https://www.cnbc.com/2020/06/19/chicago-amazon-workers-angered-by-tokenized-juneteenth-celebration.html?__source=sharebar|email&par=sharebar

Paluck, E. L. (2011). Peer pressure against prejudice: A high school field experiment examining social network change. *Journal of Experimental Social Psychology*, *47*(2), 350–358.

Paluck, E. L., & Shepherd, H. (2012). The salience of social referents: A field experiment on collective norms and harassment behavior in a school social network. *Journal of Personality and Social Psychology*, *103*(6), 899–915.

Perry, S. P., Murphy, M. C., & Dovidio, J. F. (2015). Modern prejudice: Subtle, but unconscious? The role of bias awareness in Whites' perceptions of personal and others' biases. *Journal of Experimental Social Psychology*, *61*, 64–78.

Pettigrew, T. F. (1997). Generalized intergroup contact effects on prejudice. *Personality and Social Psychology Bulletin*, *23*(2), 173–185.

Pettigrew, T. F. (2009). Secondary transfer effect of contact: Do intergroup contact effects spread to noncontacted outgroups?. *Social Psychology*, *40*(2), 55–65.

Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, *75*(3), 811–832.

Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin*, *28*(3), 369–381.

Pryor, J. B., & Day, J. D. (1988). Interpretations of sexual harassment: An attributional analysis. *Sex Roles*, *18*, 405–417.

Rasinski, H. M., & Czopp, A. M. (2010). The effect of target status on witnesses' reactions to confrontations of bias. *Basic and Applied Social Psychology*, *32*(1), 8–16.

Remedios, J. D., & Akhtar, M. (2019). Intersectional approaches to the study of confronting prejudice. In R. K. Mallett, & M. J. Monteith (Eds.), *Confronting prejudice and discrimination: The science of changing minds and behaviors* (pp. 179–200). Academic Press.

Remedios, J. D., Reiff, J. S., & Hinzman, L. (2020). An identity-threat perspective on discrimination attributions by women of color. *Social Psychological and Personality Science*, *11*(7), 889–898.

Remedios, J. D., & Snyder, S. H. (2018). Intersectional oppression: Multiple stigmatized identities and perceptions of invisibility, discrimination, and stereotyping. *Journal of Social Issues*, *74*(2), 265–281.

Remedios, J. D., & Snyder, S. H. (2015). How women of color detect and respond to multiple forms of prejudice. *Sex Roles*, 73, 371–383.

Robinson, M. S., & Alloy, L. B. (2003). Negative cognitive styles and stress-reactive rumination interact to predict depression: A prospective study. *Cognitive Therapy and Research*, *27*(3), 275–291.

Sanchez, D. T., Chaney, K. E., Manuel, S. K., & Remedios, J. D. (2018). Theory of prejudice and American identity threat transfer for Latino and Asian Americans. *Personality and Social Psychology Bulletin*, *44*(7), 972–983.

Sanchez, D. T., Chaney, K. E., Manuel, S. K., Wilton, L. S. & Remedios, J. D. (2017). Stigma by prejudice transfer: Racism threatens White women and sexism threatens men of color. *Psychological Science*, *28*(4), 445–461.

Schmid, K., Hewstone, M., Küpper, B., Zick, A., & Wagner, U. (2012). Secondary transfer effects of intergroup contact: A cross-national comparison in Europe. *Social Psychology Quarterly*, *75*(1), 28–51.

Segerstrom, S. C., Stanton, A. L., Alden, L. E., & Shortridge, B. E. (2003). A multidimensional structure for repetitive thought: what's on your mind, and how, and how much?. *Journal of Personality and Social Psychology*, *85*(5), 909–921.

Sesko, A. K., & Biernat, M. (2010). Prototypes of race and gender: The invisibility of Black women. *Journal of Experimental Social Psychology*, *46*(2), 356–360.

Shelton, J. N., & Stewart, R. E. (2004). Confronting perpetrators of prejudice: The inhibitory effects of social costs. *Psychology of Women Quarterly*, *28*(3), 215–223.

Sibley, C. G., & Duckitt, J. (2008). Personality and prejudice: A meta-analysis and theoretical review. *Personality and Social Psychology Review*, *12*(3), 248–279.

Sidanius, J., & Pratto, F. (1999). *Social dominance: An intergroup theory of social hierarchy and oppression*. Cambridge University Press.

Sommers, S. R., & Norton, M. I. (2006). Lay theories about White racists: What constitutes racism (and what doesn't). *Group Processes & Intergroup Relations*, 9(1), 117–138.

Stangor, C., Sechrist, G. B., & Jost, J. T. (2001). Changing racial beliefs by providing consensus information. *Personality and Social Psychology Bulletin*, *27*(4), 486–496.

Swim, J. K., & Hyers, L. L. (1999). Excuse me—What did you just say?! Women's public and private responses to sexist remarks. *Journal of Experimental Social Psychology*, *35*(1), 68–88.

Swim, J. K., & Miller, D. L. (1999). White guilt: Its antecedents and consequences for attitudes toward affirmative action. *Personality and Social Psychology Bulletin*, *25*(4), 500–514.

Szabo, M., & Lovibond, P. F. (2004). The cognitive content of thought-listed worry episodes in clinic-referred anxious and nonreferred children. *Journal of Clinical Child and Adolescent Psychology*, *33*, 613–622.

Szabo, M., & Lovibond, P. F. (2006). Worry episodes and perceived problem solving: A diary-based approach. *Anxiety, Stress, & Coping*, *19*, 175–187.

Tangney, J. P. & Fischer, K. W. (Eds.). (1995). *Self-conscious emotions: Shame, guilt, embarrassment, and pride* (pp. 174–197). Guilford.

Tedechi, R. G., & Calhoun, L. G. (2004). Posttraumatic growth: Conceptual foundations and empirical evidence. *Psychological Inquiry*, *15*, 1–18.

Treynor, W., Gonzalez, R., & Nolen-Hoeksema, S. (2003). Rumination reconsidered: A psychometric analysis. *Cognitive Therapy and Research*, *27*(3), 247–259.

Ullrich, P. M., & Lutgendorf, S. K. (2002). Journaling about stressful events: Effects of cognitive processing and emotional expression. *Annals of Behavioral Medicine*, *24*, 244–250.

Vitriol, J., & Moskowitz, G.B. (2021). Reducing defensive responding to implicit bias feedback: On the role of perceived moral threat and efficacy to change. *Journal of Experimental Social Psychology*, *96*. 10.1016/j.jesp.2021.104165

Vorauer, J. D., Hunter, A. J., Main, K. J., & Roy, S. A. (2000). Metastereotype activation: Evidence from indirect measures for specific evaluative concerns experienced by members of dominant groups in inter-group interactions. *Journal of Personality and Social Psychology*, *78*, 690 –707.

Watkins, E. (2004). Adaptive and maladaptive ruminative self-focus during emotional processing. *Behaviour Research and Therapy*, *42*, 1037–1052.

Watkins, E., & Moulds, M. (2005). Distinct modes of ruminative self-focus: Impact of abstract versus concrete rumination on problem solving in depression. *Emotion*, *5*, 319–328.

Wigboldus, D. H., Dijksterhuis, A., & van Knippenberg, A. (2003). When stereotypes get in the way: Stereotypes obstruct stereotype-inconsistent trait inferences. *Journal of Personality and Social Psychology*, *84*(3), 470–484.

Williams, J. M. G. (1996). Depression and the specificity of autobiographical memory. In D. C. Rubin (Ed.), *Remembering our past: Studies in autobiographical memory* (pp. 244–267). Cambridge University Press.

Woodyatt, L., & Wenzel, M. (2014). A needs-based perspective on self-forgiveness: Addressing threat to moral identity as a means of encouraging interpersonal and intrapersonal restoration. *Journal of Experimental Social Psychology*, *50*, 125–135.

Zou, L. X., & Cheryan, S. (2017). Two axes of subordination: A new model of racial position. *Journal of Personality and Social Psychology*, *112*(5), 696–717.

22 Implicit Person Memory: Domain-General and Domain-Specific Processes of Learning and Change[1]

Benedek Kurdi[1] *and Mahzarin R. Banaji*[2]
[1]*Yale University*
[2]*Harvard University*

Imagine that you learn that a co-worker of yours recently rear-ended a car. How do you update your impression of this person? Does it make a difference whether you observe the accident directly, or you hear about it from a friend? Or perhaps from the office gossip who has been known to spread rumors even if they lack any basis in reality? Does the gender or the race of the co-worker make any difference? What if this is the third time that the same co-worker has caused an accident? And what if you also know that the road was icy each time?

Person memory,[2] a subdiscipline of social cognition research, is synonymous with the birth of social cognition as a field of study. Work on person memory emerged from a small conference in the late 1970s, organized by a group of social psychologists who recognized a new opportunity to advance their field by using existing methods of cognition, especially measures of explicit memory, to study the structure and organization of knowledge about other humans (Hastie et al., 1980).

Person memory researchers investigated whether and how humans would update their beliefs about other humans when confronted with new knowledge about them, including in relatively mundane cases (such as driving ability), and also considerably more complex ones. In each study, a situation involving a single individual would be presented, with experimental designs that included multiple conditions varying information about the person and the context. Judgments of the target individual and memory for the presented information served as the main dependent measures. At its core, person memory as a field took up questions of how knowledge about other individuals is acquired, stored, retrieved, and updated (e.g., Higgins et al., 1977; Kerpelman & Himmelfarb, 1971; Srull & Wyer, 1979; Winter & Uleman, 1984).

Along with person memory's focus on the individual, another subfield of social cognition, social group cognition, also made rapid progress. Here the focus was on representations of individuals as members of social groups, such as age, gender, race/ethnicity, and sexuality. Interestingly, person memory and social group cognition remained, to a large degree, theoretically and

DOI: 10.4324/9781003045687-25

methodologically independent of each other, in spite of their close conceptual connection. Notably, from the earliest days, the idea of automaticity took center stage in the study of social group cognition; as such, measures of implicit rather than explicit memory were adapted to investigate group-based attitudes and stereotypes.

Among the first such paradigms used by social cognition researchers was sequential priming, a procedure originally designed to explore the organization of conceptual knowledge in human memory (Meyer & Schvaneveldt, 1971; Neely, 1976). These methods were then specifically adapted to the study of social categories, such as gender and race, with a focus on both implicit attitudes and stereotypes (Devine, 1989; Dovidio et al., 1986; Fazio et al., 1986; Gaertner & McLaughlin, 1983). The goal of demonstrating the presence of automatic attitudes and stereotypes, both generally and in intergroup contexts more specifically, characterized early research (Banaji et al., 1993; Banaji & Hardin, 1996). Presumably due to the dominant theoretical view that implicit social group cognition was resistant to new information (Bargh, 1999; Devine, 1989; Rydell & McConnell, 2006; Smith & DeCoster, 2000; Wilson et al., 2000), most relevant research tended to stay away from questions of change, or the acquisition and updating of representations given new knowledge. Notable exceptions emphasized the goal-dependent nature of implicit social cognitive processes, including implicit stereotyping (Moskowitz, 1996; Moskowitz et al., 1999).

In this chapter, we focus on a body of work that has used implicit measures, such as sequential priming (Fazio et al., 1986, 1995), the Implicit Association Test (Greenwald et al., 1998), the Affect Misattribution Procedure (Payne et al., 2005), and their variants, to study how evaluations of and beliefs about individual human targets are acquired and how they shift in the face of new information. As such, these studies provide insights into phenomena and processes of *implicit person memory*, i.e., knowledge about individuals that is retrieved under conditions of automaticity. By use of the term, we do not mean to suggest that any of the authors whose work we discuss below would have subscribed to this label themselves or, importantly, that their work would have been guided by a shared set of theoretical assumptions. In fact, one of the conclusions emerging from this brief overview points to the need for a stronger focus on theory building if research on implicit person memory is to make progress.

It is odd, in retrospect, that these two lines of research, one on person memory and the other on social group cognition, both pursued similar goals of applying measures of memory to the study of social entities, and yet operated in parallel, with little cross-fertilization. The fact that one engaged with explicit forms of memory and the other with implicit forms of memory hardly seems to be a sufficient reason for these lines of research to have remained separate, with little to no cross-talk. This is not to say that important exceptions did not exist. Notably, work by Uleman and colleagues as well as Moskowitz and colleagues on spontaneous trait inferences was devoted to the study of automatic processes in person memory from the earliest

days of the field (e.g., Moskowitz, 1993; Moskowitz & Roman, 1992; Uleman et al., 1996; Winter & Uleman, 1984; see also Newman, 1991). Similarly, Bargh and colleagues used subliminal priming to investigate whether the accessibility of certain constructs (including traits such as honesty or hostility) could influence processes of learning and judgment about individual social targets even outside of conscious awareness (e.g., Bargh & Pietromonaco, 1982; Bargh & Thein, 1985; Pratto & Bargh, 1991).

Research on implicit social group cognition and person memory connected in more profound and far-reaching ways in the mid-2000s when papers on implicit person memory started to appear in larger numbers (Castelli et al., 2004; DeCoster et al., 2006; McConnell et al., 2008; Meersmans et al., 2005; Rydell et al., 2006). From person memory, the study of implicit person memory inherited its core question—an interest in knowledge about individual humans; from implicit social group cognition, it inherited its core method—an emphasis on memory and judgment that occurs in automatic form.

The remainder of this chapter is structured as follows. In the first part of the chapter, we review what is already known about implicit person memory. For the sake of clarity and as a first tentative step toward theory building, we present existing implicit person memory research as belonging to one of two basic categories.

In the first category, we discuss implicit person memory work that does not emphasize the uniquely human nature of human targets or the importance of uniquely social processes of reasoning. Instead, such work uses human targets incidentally to explore how implicit evaluations and beliefs are acquired and how they change. In doing so, this research does not assume that processes of acquisition or change differ depending on the targets of learning. Rather, the tacit understanding underlying these experiments seems to be that products and brands, the self, social categories, abstract concepts, significant others, or political parties are fundamentally interchangeable with each other and with single individuals as the targets of learning. This subset of implicit person memory work emphasizes questions about the inputs to and the processes contributing to attitude and belief acquisition and change. For example, among the inputs investigated are approach/avoidance training, evaluative conditioning, and verbal statements of different kinds. When it comes to process, much attention has been devoted to the distinction between association formation mechanisms registering merely that two stimuli go together in the environment and propositional processes also encoding the specific types of relationships that stimuli can share with each other.

In the second category, we review studies that have investigated implicit person memory by attempting to identify processes specific to learning about social targets. The themes emerging from this subset of implicit person memory work include the interplay between individual-level and category-level information in implicit attitude acquisition and change, the role of facial cues, diagnostic narrative information, and the reinterpretation of previously encountered behavioral evidence about a person. This latter body of

experimental work, which operates under an assumption of the uniqueness of social learning processes, raises the complementary theoretical issue of whether these inputs and mechanisms are, in fact, unique to the social domain.

At a first glance, the approaches taken by these two sets of studies seem intrinsically incompatible: Learning about novel social targets cannot at the same time be essentially equivalent to learning about a brand or an abstract idea and also fundamentally different from it. Competing assumptions of domain-general vs. domain-specific processes in human learning and memory are, of course, not specific to the study of implicit person memory; rather, they are ubiquitous across social psychology and the cognitive sciences more broadly. Domain-general accounts posit that the computations characterizing human cognition are fundamentally the same no matter whether someone is thinking about Reese Witherspoon, the number line, or high-calorie foods (e.g., Banaji & Bhaskar, 2000; Ruff & Fehr, 2014); meanwhile, domain-specific theories suggest that human thought cannot be properly characterized without adequately considering the type of object that the person is thinking about (e.g., Cosmides & Tooby, 1994; Sperber, 1994).

Against this general theoretical backdrop as well as the apparent contradiction between the two sets of empirical studies reviewed above, we devote considerable space in the third (and final) section of this chapter to the issue of whether a domain-specific account of implicit person memory is worth proposing and defending. We also address other important topics that are yet to be settled in this area. These topics include differing definitions of what it means for a learning process to be effective, conditions of encoding, and probably the thorniest issue of all: the content and format of the mental representations mediating implicit person memory and, more generally, implicit social cognition.

Implicit Person Memory as a Case Study of Domain-General Processes

Early implicit social group cognition work inherited from studies of conceptual organization in the human mind the method of sequential priming (e.g., Meyer & Schvaneveldt, 1971; Neely, 1976). In sequential priming studies, researchers measure participants' speed and accuracy in responding to a target (e.g., the word "butter") after exposure to different primes, some assumed to be semantically related to the target (e.g., the word "bread") and some assumed to be unrelated (e.g., the word "democracy").

Along with the sequential priming paradigm, early associative theories of implicit social cognition (e.g., Devine, 1989; Rydell & McConnell, 2006; Smith & DeCoster, 2000; Wilson et al., 2000) also adopted the theoretical framework commonly used to interpret findings from this paradigm: spreading activation models of semantic memory (Collins & Loftus, 1975). These models assume that the human mind encodes concepts (such as "butter", "bread", "good", "calculating", "African American", and "democracy") via a set of nodes in a vast semantic network. The closer two concepts are associated with each

other in meaning, the stronger the connections between them in the network and, as such, the more likely encountering one is to automatically co-activate the representation of the other. The strength of connections, in turn, is assumed to be driven by something akin to a simple form of associative learning, or the Hebbian principle of activity-dependent synaptic plasticity whereby concurrent firing of neurons strengthens their connection—the idea captured by the mnemonic "what fires together wires together" (Hebb, 1949). It then follows that concepts frequently encountered in close temporal and spatial proximity (such as "butter" and "bread") will come to be strongly connected, whereas concepts infrequently or never encountered together (such as "butter" and "democracy") will be relatively weakly, or not at all, connected.

Importantly, in the early days of implicit social cognition research, the dual assumptions of (a) associative representations and (b) low-level, trial-by-trial associative learning seemingly obviated the need to study the acquisition and change of implicit evaluations and beliefs in the lab. After all, if implicit attitudes merely reflect the piecemeal shift of associative strengths in response to the long-term co-occurrence statistics of the environment, then lab-based learning paradigms may not be particularly informative for at least two reasons. First, learning processes were assumed to be too mechanical and simple to be worth studying at all. Second, relatively minor manipulations of the kind implemented in the lab were not expected to be impactful in shifting a lifetime of experience tracking co-occurrences.

However, by the mid-2000s, theoretical work in implicit social cognition emancipated itself from spreading activation models of memory and, importantly, from the assumption of purely associative learning giving rise to implicit evaluations and beliefs. Notably, the associative–propositional evaluation (APE) model by Gawronski and Bodenhausen (2006) still assumed that implicit evaluations are subserved by conceptual associations stored in long-term memory. At the same time, it also began to stake out the idea that these associations can be sensitive not only to co-occurrences experienced in the environment but also, indirectly, to the relational content of propositions.[3] That is, they are assumed to encode not only the fact that two stimuli are associated with each other and the degree of their relatedness but also the type of relationship that they share with each other. For instance, under strict associative theories, exposure to statements such as "Donald is not delusional" is expected to produce an ironic effect of strengthening the connection between the conceptual nodes "Donald" and "delusional" in long-term memory. By contrast, under the APE model, at least in certain cases, implicit evaluations can reflect the propositional content of the statement, thus strengthening the conceptual connection between "Donald" and "rational" rather than the purely associative "Donald" and "delusional".

Later, De Houwer and his colleagues formulated an even more radical proposal (e.g., De Houwer, 2007, 2014; Mitchell et al., 2009), which has since gained much empirical traction. Specifically, they posited that associative processes of learning and representation are not necessary to account

for the acquisition and change of implicit evaluations at all. Rather, similar to their explicit counterparts, implicit evaluations were assumed to be able to shift quickly and dynamically (rather than only in response to vast numbers of stimulus co-occurrences). This idea represented a radical departure from previous thinking according to which implicit cognition taps associative structures and is therefore immune to propositional reasoning. As such, propositional theories went further in expanding the scope of potential inputs to implicit social cognition than even the most flexible dual-process accounts available at the time, such as the APE model mentioned above. Moreover, propositional accounts did away with the idea of associative representation. Instead, implicit and explicit evaluations were both thought to emerge from propositional representations (e.g., "Donald is delusional") and assumed to differ only in terms of the conditions of their retrieval. Specifically, propositional accounts suggest that implicit evaluations are characterized by relatively more automatic and explicit evaluations by relatively more controlled processes of activating the same type of propositional knowledge stored in long-term memory.

These new theoretical developments have fueled innovative empirical work on the acquisition and change of implicit evaluations and beliefs for at least three reasons. First, the APE model and, to a considerably larger extent, propositional accounts, popularized the idea that implicit social cognition may be amenable to the same basic processes of flexible updating as explicit social cognition, including its propensity for quick and dynamic revision in the face of relational information. If this is the case, then those processes of updating needed to be explored experimentally. Second, with theoretical disagreement between associative accounts, dual-process accounts, and propositional accounts regarding the processes of learning and representation underlying implicit evaluation came the desire to advance the debate and to reach a satisfactory resolution. Third, the APE model and propositional accounts are both yet to be formulated with sufficient computational specificity to derive falsifiable predictions from them. As such, efforts to constrain these theories using empirical data, and to eventually develop versions specific enough to be falsifiable, have been ongoing ever since these accounts were first introduced. Given interest in and methods available for formal modeling of mental processes, we are cautiously optimistic about the likelihood of success at this time.

Against this theoretical backdrop, a considerable number of implicit person memory studies have attempted to answer two distinct but related questions. First, what types of input are capable of producing change in implicit evaluations of or beliefs about novel human targets? Researchers have studied different types of input that can roughly be divided into the following categories: (a) approach/avoidance training (e.g., Van Dessel et al., 2015, 2016); (b) attribute conditioning, that is, repeated pairings of a target with stimuli related to a semantic category (e.g., Förderer & Unkelbach, 2015, 2016); (c) evaluative conditioning, that is, repeated pairings of a target with intrinsically valenced stimuli (e.g., Förderer & Unkelbach, 2013;

Gast & Rothermund, 2011a, 2011b; Rydell & Jones, 2009); and (d) behavioral statements (e.g., Boucher & Rydell, 2012; Cone et al., 2019, 2021; Moran et al., 2015, 2017; Olcaysoy Okten et al., 2019; Olcaysoy Okten & Moskowitz, 2020; Peters & Gawronski, 2011; Rydell et al., 2007). Second, what kind of learning processes mediate learning via the different manipulations mentioned previously? Specifically, are learning processes uniquely sensitive to associative information (co-occurrences of stimuli in the environment), or do they also encode relational information (the different types of relationship that those stimuli can share with each other)?

Responsiveness of Implicit Person Memory to Different Types of Learning

The first major finding emerging from this literature, which seems robust if not incontrovertible given the strength of the evidence, is that implicit person memory is flexible (that, is capable of changing) in the face of a variety of different inputs, including the types of information described above. Such inputs include approach/avoidance training, attribute conditioning, evaluative conditioning, and behavioral statements. For example, participants in the studies by Van Dessel et al. (2015) updated implicit evaluations of novel human targets that they approached in a positive direction and those that they avoided in a negative direction. Likewise, participants in the studies by Förderer and Unkelbach (2015) updated implicit beliefs of targets on trait dimensions such as athleticism as a result of repeated pairings of the targets with material semantically related to those trait dimensions. Participants have also been shown to adjust implicit evaluations of targets paired with positive stimuli in a positive direction and those paired with negative stimuli in a negative direction (Rydell & Jones, 2009). Finally, Rydell et al. (2007) found that participants revised their implicit evaluations of targets in both positive and negative directions in a lawful manner in response to verbal statements. Taken together, this body of work demonstrates that implicit evaluations of and beliefs about novel human targets are subject to change, including in response to relatively minimal experimental manipulations.

This basic result, which has been replicated dozens of times, seems fundamentally incompatible with the idea that implicit evaluations and beliefs require vast amounts of information to form and then to change. After all, the experiments referenced above involved exposure to information about novel individuals for relatively short periods of time ranging from a few minutes to no more than an hour. Arguably, this time frame is insufficient for the types of protracted learning processes posited by traditional associative theories (e.g., Devine, 1989; Rydell & McConnell, 2006; Smith & DeCoster, 2000; Wilson et al., 2000) to be essential for implicit attitude acquisition and change to unfold. By contrast, this body of evidence is considerably easier to reconcile with more flexible dual-process theories (e.g., Gawronski & Bodenhausen, 2006) and single-process propositional theories (e.g., De Houwer, 2007, 2014;

Mitchell et al., 2009), which allow for the possibility that implicit evaluations could be updated dynamically in the face of relatively small amounts of information.

The Role of Associative vs. Propositional Processes in Implicit Person Memory

At the same time, it is notable that at least three of the four types of manipulations described above, including approach/avoidance training, attribute conditioning, and evaluative conditioning, are commonly assumed to be associative in at least two senses of the word. First, these paradigms create learning via repeated co-occurrences of a target with a stimulus or action. Second, they are usually thought to reflect the products of such learning by strengthening conceptual associations in long-term memory. Arguably, the behavioral statements used in the paradigms described above (such as "Mike cheated during a poker game") can also be interpreted in associative terms given that they include co-occurrences of a target (e.g., "Mike") with valenced words (e.g., "cheat"; Caliskan et al., 2017; Kurdi & Dunham, 2021). As such, based on these results alone, it seems that the only minor change required to make traditional associative theories compatible with the data on learning and change is to allow for the possibility that associative learning can unfold quickly, perhaps after as few as a dozen trials or even in response to a single, highly potent, stimulus pairing. This possibility is by no means incompatible with theories and empirical findings on associative learning from outside the social cognition context (e.g., Drew et al., 2010; Gershman, 2015; Rescorla & Wagner, 1972).

However, another finding, now also broadly replicated, appears to be even more fundamentally incompatible with a purely associative notion of implicit person memory (for reviews, see Cone et al., 2017; De Houwer et al., 2020; Kurdi & Dunham, 2020). Specifically, under associative accounts, implicit evaluations and beliefs are thought to reflect exclusively the fact that two things go together in the environment and the number of times that they have been paired with each other. However, in direct contradiction to this idea, implicit evaluations and beliefs have been robustly demonstrated to also reflect *how* two pieces of information are related to each other.

Here we mention only a few cases in which implicit evaluations were found to encode relational information in a way that seems fundamentally incompatible with associative accounts. For example, implicit evaluations of novel targets in the studies by Peters and Gawronski (2011) and Boucher and Rydell (2012) were sensitive to whether the content of statements about those targets was affirmed or negated: A person presented along with the behavior "continually yells at his wife in public" was evaluated negatively when the behavior was revealed to be characteristic of him; however, when it was revealed to be uncharacteristic, implicit evaluations shifted in a positive direction. The idea that abstract knowledge of this kind would be crucial, or even

relevant, to implicit person memory would have been difficult to entertain in a predominantly associative framework. Notably, Kurdi and Dunham (2021) even found that the updating of implicit evaluations of a novel target depended on whether participants made normative errors in propositional inference, such as denying the antecedent, providing further evidence for the importance of high-level reasoning processes. Together, these results strongly suggest that associative processes alone are insufficient to account for the patterns of learning and updating observed in implicit person memory.

Domain-Specific Processes in Implicit Person Memory

The implicit person memory studies reviewed in the previous section share the important commonality that they have been designed to test relatively domain-general theories of implicit social cognition. These theories assume, more or less tacitly, that processes of learning and representation cut across different types of human (and even non-human) targets and that, therefore, different types of target stimuli used to investigate such processes are relatively interchangeable with each other. In fact, in our own work, we have conducted learning studies involving existing social categories and non-social targets (e.g., Kurdi & Banaji, 2017), novel social groups (e.g., Kurdi & Dunham, 2021), and novel individuals (e.g., Mann et al., 2020) without systematically investigating whether these targets differ from each other in theoretically relevant ways. Nevertheless, convergent results obtained across different categories of stimuli suggest that the underlying learning process is sufficiently general to produce similar outcomes. Such a result may be seen as surprising from the perspective of theories across the cognitive sciences that have emphasized the importance of domain-specific processes to human learning and memory (e.g., Cosmides & Tooby, 1994; J. P. Mitchell et al., 2005; Saxe & Kanwisher, 2003; Sperber, 1994).

By contrast, the studies discussed in this section have focused on inputs to and processes of implicit attitude acquisition and change that are relatively specific to the domain of person memory. Such domain specificity is usually related to one of two aspects of studies: the types of information being presented and the types of information processing assumed to occur. Of course, these two aspects are intertwined with each other more often than not, but here we discuss each of them separately for ease of presentation.

On the one hand, some studies have relied on information about novel social targets that would not be meaningfully interpretable outside the social domain. Such studies have included experiments probing the interplay of individual-level and category-level information (e.g., Cao & Banaji, 2016; Gawronski et al., 2003; McConnell et al., 2008; Rubinstein et al., 2018; Rubinstein & Jussim, 2019) and the effects of facial cues on implicit evaluation (e.g., Gawronski & Quinn, 2013; Shen et al., 2020). On the other hand, studies have also presented information to participants that was assumed to give rise to domain-specific processes of social reasoning: diagnostic

information about a target's true moral character (e.g., Cone et al., 2019, 2021; Cone & Ferguson, 2015) or information prompting participants to reinterpret a target's previously encountered behaviors (e.g., Kurdi et al., 2021b; Mann & Ferguson, 2015, 2017; Olcaysoy Okten et al., 2019).

Additional Evidence for Flexibility and the Role of Relational Information

The studies reviewed here differ from the studies reviewed previously in their emphasis on uniquely social types of information and inference. However, at the same time, similar to the relatively domain-general studies discussed previously, they can provide evidence on the flexibility of implicit person memory in the face of different types of input as well as on the role of relational information in processes of updating. Indeed, these relatively domain-specific studies have provided ample evidence for the flexible updating of implicit evaluations. As such, their findings largely converge with the experiments relying on relatively domain-general information reviewed above in suggesting that, contrary to influential early conceptualizations of implicit social cognition as resistant to updating (e.g., Bargh, 1999; Devine, 1989; Rydell & McConnell, 2006; Smith & DeCoster, 2000; Wilson et al., 2000; but see Moskowitz et al., 1999; Blair, 2002), implicit person memory is remarkably flexible in response to ever-changing informational inputs.

Moreover, similar to the set of domain-general studies reviewed above, updating in many of these experiments seems to have unfolded in a way that is difficult to reconcile with notions of slow and piecemeal associative learning to the exclusion of propositional processes of reasoning, which is a central assumption of associative accounts of implicit social cognition. For example, Cone and Ferguson (2015) demonstrated that implicit evaluations of a novel target formed from dozens of behavioral statements can reverse in valence from positive to negative as a result of a single piece of highly diagnostic information about that target (e.g., the person having mutilated a small, defenseless animal). Given the extremity of the valence of the novel information, this result in and of itself may be explained by a particularly potent form of associative learning. However, the finding that the strength of updating tracked the extent to which participants believed that the novel information was believable and diagnostic of the target's moral character (Cone et al., 2019, 2021) seems even more fundamentally incompatible with purely associative accounts.

Studies relying on the idea of reinterpretation have produced findings that are similarly difficult to reconcile with piecemeal association formation mechanisms. In these studies, unlike in most work on the updating of implicit evaluations, attitude change is not achieved by presenting entirely novel information about the target; rather, participants are prompted to reconsider the evaluative implications of already known information. For example, Kurdi et al. (2021b) have shown that exposure to excerpts from a real-world

podcast, containing a mix of positive and negative information, can lead to considerable updating of initially highly negative evaluations of an individual. Furthermore, similar to the studies relying on diagnostic information reviewed above, the amount of updating was predicted by the extent to which participants found the novel counterattitudinal information persuasive. Such ubiquitous involvement of higher-order reasoning processes in implicit person memory seems difficult if not impossible to reconcile with the idea of a cognitive system that merely tracks co-occurrences of targets with valenced information in the environment.

Evidence of Domain-Specific Processes

Notably, given their reliance on certain types of (social) information and (social) reasoning, these experiments also additionally inform about relatively domain-specific processes of implicit person memory. These insights concern the relative importance of category-level and individual-level information, use of facial cues, and the role of diagnostic information and reinterpretation. As mentioned in the introduction, we see an inherent contradiction between the approaches of (a) treating human targets as fundamentally interchangeable with other classes of stimuli (such as brands or abstract concepts) in implicit person memory work vs. (b) assuming that implicit evaluations of and beliefs about human targets are sensitive to a unique set of inputs and learning processes. As such, we hope that placing these two groups of studies side by side and critically reviewing both sets of underlying assumptions will prove helpful in reaching a resolution and achieving theoretical integration.

Category-Level Information vs. Individuating Information

A relatively large number of studies have investigated the interplay between and relative importance of category-level information (e.g., information that a target is a man or Iranian American) and individuating information (e.g., information that they rear-ended a car or took money from a donation box) in implicit person memory. Given the uniquely social nature of both the social category information and the individuating information used in these studies, this work can reasonably be interpreted as informing about domain-specific processes of social learning and memory.

In an early study, Gawronski et al. (2003) demonstrated that implicit evaluations of social categories can bias the process of forming an impression of individuals belonging to those categories. Specifically, participants in these studies interpreted ambiguous behaviors performed by a Black target more negatively than the same ambiguous behaviors performed by a White target but only to the extent that they had relatively positive implicit attitudes toward White Americans and relatively negative implicit attitudes toward Black Americans. Given the uniquely social nature of both the category-level and individual-level information used by Gawronski et al. (2003), this

experiment seems to provide early evidence for the involvement of uniquely domain-specific processes in implicit person memory.

More recent work has investigated the formation of implicit evaluations of and implicit beliefs about novel social targets more directly by presenting category-level and individual-level information to participants that were contradictory in their evaluative or semantic implications. Similar to the Gawronski et al. (2003) study, given the uniquely social nature of both types of information, these experiments are broadly assumed to inform about domain-specific inputs to implicit person memory.

For example, Cao and Banaji (2016) introduced participants to a male and a female target (category-level information) and then described the former as a nurse and the latter as a doctor (individuating information). Although implicit beliefs shifted significantly relative to baseline, they were still indicative of the persistence of stereotype-congruent associations of the female target with the category "nurse" and the male target with the category "doctor." Other work using different designs and different targets has produced results ranging from complete reliance of implicit person memory on category-level information to the exclusion of individual-level information (McConnell et al., 2008) to complete reliance on individual-level information to the exclusion of category-level information (Rubinstein et al., 2018; Rubinstein & Jussim, 2019). As such, further empirical and conceptual work will be necessary to reconcile these seemingly contradictory findings with each other. However, crucially, as a set, these studies seem to provide compelling evidence for the role of uniquely social types of input in the updating of implicit evaluations and beliefs.

The Role of Facial Information

A second, considerably smaller, set of studies have investigated the influence of the human face on implicit person memory. The effects of different facial cues, such as the shape of the face, the distance between the eyes, and the height of the forehead, on impression formation have been well documented using self-report measures (for reviews, see Todorov et al., 2008, 2015). Crucially from our perspective, similar to social category information, facial features are widely seen as a source of uniquely social information. As such, studies investigating the effects of facial cues on implicit measures of evaluation and belief can also provide information on relatively domain-specific mechanisms of learning and change in implicit person memory.

In a first relevant study by Gawronski and Quinn (2013), participants read positive and negative behavioral statements about novel targets (presented as faces) and then completed implicit measures of attitude toward previously unseen targets whose faces were manipulated to appear similar to the targets about whom participants had learned earlier. Implicit evaluations generalized to these novel target faces, thus providing initial evidence for the idea that facial cues can influence implicit person memory. In more recent work, Shen

et al. (2020) produced a conceptually similar finding, demonstrating that targets whose faces were manipulated to appear extremely untrustworthy engendered highly negative implicit evaluations. At the same time, diagnostic behavioral information about the same targets (see below) led to the revision, and sometimes full reversal, of the face-based negative evaluations. Again, these studies seem to provide evidence for the operation of relatively domain-specific processes of learning and updating in implicit person memory.

Diagnostic Information and Reinterpretation

Research by Ferguson, Cone, Mann, and colleagues has investigated in detail two seemingly uniquely social forms of updating in implicit person memory: the first relying on diagnostic information (Cone et al., 2019, 2021; Cone & Ferguson, 2015) and the second on the reinterpretation of previously encountered behavioral information (Kurdi et al., 2021b; Mann et al., 2020; Mann & Ferguson, 2015, 2017).

As alluded to earlier, the first type of paradigm tends to pit two types of information against each other: a large number of behavioral statements implying a positive evaluation of a novel target and a single piece of extremely negative and diagnostic behavioral information about the same target. Notably, these studies are theoretically well integrated with, and have directly expanded upon, a long line of work relying on explicit measures of impression formation (e.g., Reeder & Brewer, 1979; Reeder & Coovert, 1986; Trafimow & Schneider, 1994). Specifically, they show that negative behavioral information, especially extremely negative behavioral information, tends to give rise to particularly strong dispositional inferences and that these inferences, in turn, influence not only explicit but also implicit evaluations. Given that dispositional inferences are widely seen as uninterpretable outside a social context, these studies can also be construed as providing evidence for the operation of uniquely social processes in implicit person memory.

Similar to studies involving diagnostic behavioral information, studies relying on the idea of reinterpretation are also usually assumed to demonstrate the flexibility of implicit evaluations in the face of uniquely social information. The typical design of reinterpretation experiments involves presenting an initial narrative that is rich in negative episodic details (e.g., Mann & Ferguson, 2015, 2017; but see Olcaysoy Okten et al., 2019). For example, participants in Mann and Ferguson (2015) were introduced to a novel target called Francis West and read a relatively long vignette about him breaking into his neighbors' homes to remove "precious things" from them. Based on this initial information, participants construed West's actions as a burglary and evaluated him negatively on both explicit and implicit measures. Subsequently, in the reinterpretation condition, participants learned that West entered the houses because they were on fire and the "precious things" that he removed (saved) were actually the neighbors' children. Although the second piece of information is minimal in length and

detail compared with the narrative presented at the outset of the study, it was sufficient to induce revisions to, and often full reversals of, the initially formed negative evaluations. As such, this line of work provides additional evidence for the effective and rapid revision of implicit evaluations in the face of relatively domain-specific forms of reasoning about social information.

Interim Summary: The Flexibility of Implicit Person Memory

To summarize the insights gained from the work reviewed previously, evidence for the possibility of rapidly and dynamically revising implicit evaluations of novel social targets seems overwhelming. Processes of revision can unfold in response to information that could be characterized as relatively domain-general (including actions to approach or avoid targets, pairings of targets with intrinsically pleasant or unpleasant stimuli, and exposure to valenced verbal descriptions of targets), or in response to information that could be characterized as more specific to social targets (including competing category-level and individuating information, facial cues, diagnostic behavioral information, and information giving rise to the reinterpretation of previously encountered behaviors). These inputs to the updating of implicit evaluations clearly go beyond simple stimulus pairings; moreover, learning can emerge highly effectively, within a matter of a few minutes. Finally, even learning from seemingly simple paradigms involving the repeated presentation of stimulus pairings has been shown to be modulated by the meaning with which participants imbue those stimulus pairings, either spontaneously or as a result of relational information provided by the experimenter.

Overall, these results are difficult to reconcile both with most early conceptualizations of implicit social cognition, inherited from spreading activation models of memory, as well as the associative accounts building on these early conceptualizations (e.g., Bargh, 1999; Devine, 1989; Rydell & McConnell, 2006; Smith & DeCoster, 2000; Wilson et al., 2000; but see Moskowitz et al., 1999; Blair, 2002). After all, according to these accounts, implicit evaluations and beliefs can be updated only as a result of protracted learning involving vast numbers of stimulus pairings. Moreover, under these theories, implicit cognition is thought to be sensitive exclusively to co-occurrence information experienced in the environment without reflecting the ways in which such information is construed by the reasoner. By contrast, these findings are compatible with propositional accounts of implicit evaluation (e.g., De Houwer, 2007, 2014; Mitchell et al., 2009) as well as other theories that do not posit a strict separation between an implicit system reflecting an associative mode of processing and an explicit system reflecting a propositional mode of processing (e.g., Cunningham et al., 2007; Fazio, 2007; Gawronski & Bodenhausen, 2006; Kurdi & Dunham, 2020).

In summary, the body of knowledge generated by the field of implicit person memory (and implicit social cognition more broadly) over the past two decades

has led to a fundamental revision of most early conceptualizations of implicit attitudes. Specifically, there has been a significant movement away from theories uniquely emphasizing the importance of associative processes of learning and representation toward theories emphasizing (additionally or exclusively) the importance of propositional processes of learning and representation. Although substantial disagreement still remains about the relative contributions of different types of processes to implicit evaluation (e.g., Gawronski & Bodenhausen, 2018; De Houwer, 2018; Kurdi & Dunham, 2020; McConnell & Rydell, 2014), we believe that this shift alone demonstrates the considerable promise of the experimental approaches taken since the mid-2000s in unraveling the nature of implicit cognition and social learning.

Open Empirical and Theoretical Questions in Implicit Person Memory

In spite of the theoretically rich insights that have emerged over the past 20 years of research on implicit person memory, a sizable number of issues, some of them of crucial theoretical importance, are yet to be resolved or even to be systematically addressed. Some of these issues are specific to implicit person memory but several of them apply, *mutatis mutandis*, to implicit social cognition more generally. In the remainder of the chapter, we offer a brief and subjective overview of these unresolved issues. We hope that this overview will provide an impetus for new theory development and empirical work, or at least serve as a basis for discussions about what directions new theoretical and empirical approaches should take.

Do We Need a Domain-Specific Theory of Implicit Person Memory?

The apparent inconsistency in basic theoretical assumptions between the two sets of studies reviewed above seems in need of resolution. Specifically, although these theoretical assumptions are rarely if ever discussed explicitly, the first set of studies, relying on paradigms such as approach/avoidance training, attribute conditioning, evaluative conditioning, and verbal statements, seem to assume that the processes and mechanisms of implicit evaluation are largely domain-general. From this assumption it follows that (a) paradigms originally developed in the context of animal learning, and only later adapted to the study of human cognition and human social cognition, are generally well-suited to the study of implicit attitude acquisition and change and (b) the stimuli used in these paradigms (be they non-words, shapes, products, social groups, abstract concepts, or single individuals) are generally interchangeable with each other. Virtually all theories of implicit evaluation discussed in this chapter so far seem to make exactly the same assumption simply by virtue of being silent on the possibility of any domain-specific processes of implicit evaluation.

By contrast, the second set of studies seem to make a fundamentally different assumption, namely that there are at least some processes of implicit person memory that cannot be described using domain-general mechanisms. These studies have provided evidence for learning unfolding via the interplay of individuating and social category information, facial cues, diagnostic behavioral information, and reinterpretation. Notably, to the degree that these studies are embedded in existing theoretical frameworks, they tend to emphasize accounts that have been formulated in the context of person memory, such as the continuum model of impression formation (Fiske & Neuberg, 1990), interpersonal transference (Chen & Andersen, 1990), and attribution theories (e.g., Reeder & Brewer, 1979).

As presently implemented, these two approaches seem fundamentally incompatible with each other. Implicit evaluations of individual targets cannot be subserved by exclusively, or at least mostly, domain-general processes and also, simultaneously, by exclusively, or at least mostly, domain-specific processes. At least three possibilities for resolving this apparent inconsistency are worth considering. First, it is conceivable that the processes underlying implicit evaluation are mostly domain-general. Second, it is conceivable that the processes underlying implicit evaluation are mostly domain-specific. Finally, it is also conceivable that this level of analysis is too coarse and different aspects of implicit evaluation should be investigated separately along the continuum from fully domain-general to fully domain-specific.

Without prejudging how these open questions will be resolved, we believe that substantial amounts of theoretical work and empirical evidence already exist to suggest that these issues are sufficiently important to be experimentally addressed. First, the most critical takeaway from propositional theories of implicit evaluation (e.g., De Houwer, 2007, 2014; Mitchell et al., 2009) and empirical work informed by these theories is that associative processes relying exclusively on the idea of registering co-occurrence information are insufficient to capture the processes of acquisition and change observed in experimental studies. However, if this is the case, and processes of high-level reasoning have a ubiquitous influence on implicit evaluation, then the possibility that humans might reason differently about social and non-social entities, and even different social entities, cannot be dismissed out of hand. The amount of existing research that has shown distinctions between social and non-social processing, especially when measures of neural activation are included (e.g., J. P. Mitchell et al., 2005; Saxe & Kanwisher, 2003; Young et al., 2007), strongly suggests that this issue should become a priority for testing.

As an example from recent theoretical work, Faure et al. (2020) have made a compelling case for studying implicit evaluation in the context of close relationships and for incorporating the findings from such studies into overarching theories of implicit social cognition. Specifically, these authors point out that these theories tend to treat race attitudes (and, to some degree, other intergroup attitudes) and attitudes toward novel experimental targets as ideal typical cases of implicit evaluation. However, unlike these two

attitude domains, highly consequential implicit evaluations of close others, such as family members and romantic partners, stem from rich, complex, and constant, or at least repeated, personal experience that fluctuates in both valence and intensity. If the processes giving rise to implicit evaluations of close others and other social entities differ from each other in such a clear and potentially consequential way, then beliefs and evaluations in other domains that are being routinely investigated using implicit measures of (social) cognition, such as consumer goods and brands (Dimofte, 2010), addictive substances (Lindgren et al., 2020), and other attitude objects relevant to psychopathology (Teachman et al., 2019), may also differ considerably from each of those areas and also from each other.

Moreover, a systematic comparative approach would presumably also prompt investigators to be more precise about the types of differences that may exist between social and non-social attitude objects that are relevant to processes of attitude acquisition and revision. For example, the studies investigating the effects of conflicting social group information and individuating information seem to tacitly assume that this distinction is uniquely relevant to human targets. Although certain aspects of such information may indeed be unique, cases of contradictory category-level and individual-level information can also be considered a specific instance of the effects of how categorical and exemplar-specific information are integrated with each other in long-term memory. This issue has been studied extensively across different subfields of psychology (e.g., Medin et al., 1984; Merriman et al., 1997; Schapiro et al., 2017).

Similarly, although studies investigating the updating of implicit evaluations via diagnostic behavioral information and reinterpretation seem to assume that reasoning on the basis of these two types of input is uniquely social, such reasoning may at least in part be supported by domain-general processes. For example, the diagnostic behavioral information used in studies by Cone, Ferguson, and colleagues tends to be both negative and extreme in valence and, as such, the more general phenomenon of negativity dominance (Rozin & Royzman, 2001) may contribute to the effect. Of course, one could make the argument that the mechanisms underlying updating of implicit evaluations via diagnostic behavioral information cannot be satisfactorily explained by simple negativity dominance because the effect does not arise if the target is only incidentally associated with negative information (Cone & Ferguson, 2015). However, the process of assigning observable outcomes in the world to hidden latent causes is by no means specific to the social domain (Gershman et al., 2015; Gershman & Niv, 2010). Moreover, Kurdi et al. (2021c) have provided tentative evidence for the operation of negativity dominance specifically in the context of the acquistion of non-social attitudes.

A potentially defining difference between implicit person memory and other forms of implicit evaluation may be the sensitivity of the former, but not of the latter, to reasoning about hidden mental states, such as goals, beliefs, and desires (J. P. Mitchell et al., 2005; Saxe & Kanwisher, 2003;

Young et al., 2007). For example, Saxe and Kanwisher (2003) have found differences in neural activation in response to scenarios involving false beliefs (e.g., Sally erroneously believing that an object is in one box rather than in the other) vs. scenarios involving other types of false representations (e.g., an outdated photograph, such as a photograph of an apple hanging from a tree, which has been blown to the ground by a strong wind since the picture was taken). Similarly, Mitchell et al. (2005) identified unique patterns of neural activation when participants were asked to consider targets' psychological states (e.g., "curious" or "energetic") rather than their physical features.

In line with the idea that mental state reasoning may be an important contributor to implicit person memory, a recent line of studies by Kurdi et al. (2020) have provided evidence for the sensitivity of implicit evaluations to targets' accurate and false beliefs about the world. Specifically, in these studies, implicit evaluations of novel targets were more negative when they caused positive rather than negative outcomes. However, holding the valence of the outcome constant, implicit evaluations deviated more strongly from neutrality when the outcomes were caused intentionally (e.g., putting poison in someone's coffee knowing that it is poison) rather than unintentionally (e.g., putting poison in someone's coffee erroneously believing that it is sugar).

However, although these results are suggestive, there remains much to be explored about the potential uniqueness of some types of input to, or processes underlying, implicit person memory. First, the vignettes used in the experiments by Kurdi et al. (2020) featured extremely negative outcomes. As such, the results may not generalize to more mundane cases, which would undercut the idea that implicit person memory universally involves mental state reasoning. Second, whether mental state reasoning consistently contributes to implicit evaluations of social targets is unclear. It is possible that mental state reasoning may be impactful only in cases where it is directly applicable to the problem at hand (such as diagnostic behavioral information). Alternatively, such reasoning may operate by default even in cases where seemingly no relevant information is being provided (such as evaluative conditioning). Third, not all human targets activate mental state reasoning to the same degree (e.g., Harris & Fiske, 2006; McLoughlin & Over, 2017), which may make the ubiquity of this potentially uniquely social input to the updating of implicit evaluations questionable.

What about Different Definitions of Effectiveness?

The overwhelming majority of studies reviewed in this chapter equate the effectiveness of different forms of person learning, at least tacitly, with their immediate capacity to modulate responding on an implicit measure of belief or evaluation. However, as suggested by recent investigations in the context of implicit race attitudes, the temporary malleability of implicit evaluations need not translate into enduring change (Lai et al., 2014, 2016). A handful of studies have already explored, and provided evidence for, the durability of

change in implicit person memory brought about by different interventions (Cone et al., 2021; Kurdi et al., 2021a; Mann et al., 2020; Mann & Ferguson, 2017; Ranganath & Nosek, 2008). However, we hope that such investigations will become more commonly implemented in the future given that they have the potential to inform both about basic mechanisms of learning and change and about the effectiveness of different interventions in producing long-term shifts in implicit evaluation.

Moreover, if learning is to be truly effective, it should generalize across different contexts. In the study of both animal learning (e.g., Bouton & Bolles, 1979; Bouton & Peck, 1992) and human memory (e.g., Schiller et al., 2010, 2013) it is commonplace to assume that novel information contradicting the evaluative implications of a prior learning episode does not usually erase the memory trace associated with the original experience. Rather, it tends to create a new memory trace, which now competes with the old memory trace for expression. Although a relatively large body of evidence exists to suggest that implicit evaluations can be context-dependent (Gawronski et al., 2018), the boundary conditions and replicability of such effects are not sufficiently well understood (Gawronski et al., 2015). Moreover, with the sole exception of a study by Brannon and Gawronski (2017), none of these studies have systematically investigated the relative context (in-)dependence of different types of input to implicit evaluation.

Finally, next to nothing is known about whether and to what degree implicit evaluations created via different types of knowledge differ from each other in terms of their resistance to novel counterattitudinal information (but see Kurdi et al., 2021a). In fact, related to the conclusions of the previous section, different criteria of relative effectiveness may not yield consistent results across different areas of attitude acquisition and change. As such, based on presently available data, it can be confidently concluded that processes of implicit person memory (and implicit social cognition more broadly) are momentarily malleable in the face of complex information that goes well beyond simple co-occurrence information, including reasoning about causes and effects, mental states, and the believability and diagnosticity of the evidence that one encounters. A small number of studies additionally suggest that, at least in the context of novel social targets, such effects can persist beyond a single experimental session. However, very little is known about the (relative) context-specificity of different learning modalities as well as the resistance of their outputs to countervalent information.

What about Different Encoding Conditions?

Similar to the dearth of research on different conditions under which newly learned information about social targets can be retrieved, the state of knowledge about the effects of different encoding conditions on processes of updating in implicit person memory is extremely limited. In most experimental studies reviewed above, and in most studies on the acquisition and

change of implicit evaluations more generally, participants are able and motivated to focus on the evaluative information presented to them. Moreover, they are usually specifically instructed to memorize the information to which they are exposed. A few studies have manipulated the availability of cognitive resources during learning (e.g., Boucher & Rydell, 2012; Mann & Ferguson, 2015; Shen et al., 2020); however, these studies have yielded conflicting results. Notably, recent work by Fan et al. (2021) suggests that when cognitive resources are available, implicit evaluations can spontaneously reflect the effects of relational information; however, under cognitive load, implicit attitudes seem to be uniquely sensitive to co-occurrence information. These results call into question the ubiquity of propositional influences on implicit evaluation and highlight the need for further inquiry into the boundary conditions of such effects.

Moreover, the problem may run too deep to be solved simply by placing participants under cognitive load while they are exposed to information designed to create or shift implicit evaluations. Remarkably, based on recent studies by Wimmer and Poldrack, the results of single-session learning studies in which participants encounter novel information in a highly massed fashion may not at all, or only under extremely limited conditions, generalize to more ecologically valid settings in which reasoners are usually exposed to information about the same target across multiple occasions over time (Wimmer et al., 2018; Wimmer & Poldrack, 2021).

Specifically, these experiments suggest that when information is presented in a massed way to participants, the effectiveness of even model-free processes of value-based learning, long assumed to be emerging in a purely stimulus-driven way, is highly correlated with individual differences in working memory capacity. However, when the same information was administered to participants across three occasions over a three-day period, the correlation between working memory capacity and value-based learning disappeared entirely. As such, based on these results, it seems premature to conclude that truly associative processes cannot contribute to evaluative learning (Corneille & Stahl, 2018). Rather, the paradigms routinely used to try to produce such effects may simply not create the appropriate psychological conditions for those very effects to emerge.

Finally, What about Mental Representations?

Based on the evidence reviewed previously, it is quite clear that implicit evaluations can (at least momentarily) reflect the effects of different types of relational information in a way that accounts relying on association formation mechanisms alone cannot explain. However, propositional accounts of implicit evaluation (e.g., De Houwer, 2007, 2014; Mitchell et al., 2009) are committed to a considerably stronger theoretical claim—namely, that implicit attitudes emerge from the automatic activation of propositional

representations. At present, there is no evidence to suggest that this more sweeping idea is accurate, either in implicit person memory or beyond.

Specifically, as discussed in more detail in Kurdi and Dunham (2020), explicit and implicit evaluations could be thought of as being sensitive to the same basic sources of information, including relational information, but encoding such information at different levels of compression. To take an example from the person memory domain, let's assume that individuals can differ from each other along three dimensions: warmth, competence, and physical attractiveness. Each individual receives a score on each of the three dimensions (akin to a probabilistic implementation of truth values) and a weighted sum of the three scores is used to calculate the overall evaluation.

In this setting, explicit evaluations may be conceptualized as encoding the scores on each dimension, the weights, as well as the resulting summary evaluation, whereas implicit evaluations may be conceptualized as encoding only the resulting summary evaluation without having access to the specific pieces of (propositional) information from which the summary evaluation emerged. In fact, some initial evidence obtained mainly in non-social contexts seems to suggest that the idea of compression, entirely absent from both dual-process and propositional theories, can be useful in understanding some patterns of flexibility and recalcitrance in the updating of implicit evaluations (e.g., Kurdi et al., 2019, 2021c). This foray notwithstanding, the issue remains open for further exploration.

Conclusion

In this chapter, we provided a brief overview of a field we refer to as implicit person memory. Implicit person memory encompasses experimental studies investigating the acquisition (learning) and revision (updating) of implicit attitudes toward and implicit beliefs about novel social targets in response to different types of information. The evidence reviewed here seems to provide unequivocal support for the immediate malleability of implicit evaluations in the face of multiple sources of information, some of which are routinely regarded as emerging from domain-general mechanisms (e.g., evaluative conditioning) and others of which are routinely regarded as emerging from domain-specific processes of social reasoning (e.g., mental state inferences). These results are difficult to reconcile with most early conceptualizations of implicit evaluation as (a) purely associative and (b) generally resistant to updating.

The effects of relational information on implicit person memory are extremely well-established, and support for the possibility of rapid revisions to implicit attitudes is equally robust. However, considerably less is known about (a) the domain-specific vs. domain-general nature of the processes by which implicit evaluations are updated; (b) the generalizability of and mechanisms underlying updating across different domains; (c) the persistence of updating effects over time, their context-specificity, and the resistance of updating to counterattitudinal information; (d) the scope of encoding

conditions under which implicit evaluations can exhibit sensitivity to relational information; and, finally, (e) the mental representations mediating the effects of co-occurrence and relational information on implicit evaluation.

We hope that, by summarizing available evidence on these issues and by highlighting gaps in the existing literature, the present review will help create a coherent and systematic theory of implicit person memory and a more comprehensive, accurate, and easily falsifiable theory of implicit social cognition.

Notes

1 **Authors' Note:** Benedek Kurdi is a member of the Scientific Advisory Board of Project Implicit, a 501(c)(3) non-profit organization and international collaborative of researchers who are interested in implicit social cognition.
2 The term "person perception" is currently used considerably more frequently than the term "person memory." However, the term can cause confusion when used in psychology broadly because, with few exceptions, "person perception" research does not investigate any truly perceptual processes. Therefore, and to maintain connection with the history of the field, we have opted to use the term "person memory," although it may seem anachronistic to some readers.
3 As such, the APE model is not a fully propositional account of implicit evaluation but rather a flexible dual-process model that, like propositional accounts, allows for a role of propositional processes but, unlike propositional accounts, retains the idea of associative learning and representation from early theories of implicit social cognition.

References

Banaji, M. R., & Bhaskar, R. (2000). Implicit stereotypes and memory: The bounded rationality of social beliefs. In D. L. Schacter, & E. Scarry (Eds.), *Memory, brain, and belief* (pp. 139–175). Harvard University Press.

Banaji, M. R., & Hardin, C. D. (1996). Automatic stereotyping. *Psychological Science*, *7*(3), 136–141. 10.1111/j.1467-9280.1996.tb00346.x

Banaji, M. R., Hardin, C., & Rothman, A. J. (1993). Implicit stereotyping in person judgment. *Journal of Personality and Social Psychology*, *65*(2), 272–281. 10.1111/j.1467-9280.1996.tb00346.x

Bargh, J. A. (1999). The cognitive monster: The case against the controllability of automatic stereotype effects. In S. Chaiken, & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 361–382). The Guilford Press.

Bargh, J. A., & Pietromonaco, P. (1982). Automatic information processing and social perception: The influence of trait information presented outside of conscious awareness on impression formation. *Journal of Personality and Social Psychology*, *43*(3), 437–449. 10.1037/0022-3514.43.3.437

Bargh, J. A., & Thein, R. D. (1985). Individual construct accessibility, person memory, and the recall-judgment link: The case of information overload. *Journal of Personality and Social Psychology*, *49*(5), 1129–1146. 10.1037/0022-3514.49.5.1129

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review*, *6*(3), 242–261. 10.1207/s15327957pspr0603_8

Boucher, K. L., & Rydell, R. J. (2012). Impact of negation salience and cognitive resources on negation during attitude formation. *Personality and Social Psychology Bulletin*, *38*(10), 1329–1342. 10.1177/0146167212450464

Bouton, M. E., & Bolles, R. C. (1979). Contextual control of the extinction of conditioned fear. *Learning and Motivation*, *10*(4), 445–466. 10.1016/0023-9690(79)90057-2

Bouton, M. E., & Peck, C. A. (1992). Spontaneous recovery in cross-motivational transfer (counter conditioning). *Animal Learning & Behavior*, *20*(4), 313–321. 10.3758/bf03197954

Brannon, S. M., & Gawronski, B. (2017). A second chance for first impressions? Exploring the context-(in)dependent updating of implicit evaluations. *Social Psychological and Personality Science*, *8*(3), 275–283. 10.1177/1948550616673875

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, *356*(6334), 183–186. 10.1126/science.aal4230

Cao, J., & Banaji, M. R. (2016). The base rate principle and the fairness principle in social judgment. *Proceedings of the National Academy of Sciences*, *113*(27), 7475–7480. 10.1073/pnas.1524268113

Castelli, L., Zogmaister, C., Smith, E. R., & Arcuri, L. (2004). On the automatic evaluation of social exemplars. *Journal of Personality and Social Psychology*, *86*(3), 373–387. 10.1037/0022-3514.86.3.373

Chen, S., & Andersen, S. M. (1990). Relationships from the past in the present: Significant-other representations and transference in interpersonal life. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 31, pp. 123–190). Academic Press. 10.1016/s0065-2601(08)60273-7

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428. 10.1037/0033-295x.82.6.407

Cone, J., & Ferguson, M. J. (2015). He did *what*? The role of diagnosticity in revising implicit evaluations. *Journal of Personality and Social Psychology*, *108*(1), 37–57. 10.1037/pspa0000014

Cone, J., Flaharty, K., & Ferguson, M. J. (2019). Believability of evidence matters for correcting social impressions. *Proceedings of the National Academy of Sciences*, *116*(20), 9802–9807. 10.1073/pnas.1903222116

Cone, J., Flaharty, K., & Ferguson, M. J. (2021). The long-term effects of new evidence on implicit impressions of other people. *Psychological Science*. Advance online publication. 10.1177/0956797620963559

Cone, J., Mann, T. C., & Ferguson, M. J. (2017). Changing our implicit minds: How, when, and why implicit evaluations can be rapidly revised. In James M. Olson (Ed.), *Advances in experimental social psychology* (Vol. 56, pp. 131–199). Elsevier. 10.1016/bs.aesp.2017.03.001

Corneille, O., & Stahl, C. (2018). Associative attitude learning: A closer look at evidence and how it relates to attitude models. *Personality and Social Psychology Review*, *23*(2), 161–189. 10.1177/1088868318763261

Cosmides, L., & J. Tooby (1994). Origins of domain specificity: The evolution of functional organization. In L. A. Hirschfeld, & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 85–116). Cambridge University Press.

Cunningham, W. A., Zelazo, P. D., Packer, D. J., & Van Bavel, J. J. (2007). The iterative reprocessing model: A multilevel framework for attitudes and evaluation. *Social Cognition*, *25*(5), 736–760. 10.1521/soco.2007.25.5.736

DeCoster, J., Banner, M. J., Smith, E. R., & Semin, G. R. (2006). On the inexplicability of the implicit: Differences in the information provided by implicit and explicit tests. *Social Cognition*, *24*(1), 5–21. 10.1521/soco.2006.24.1.5

De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *The Spanish Journal of Psychology*, *10*(2), 230–241. 10.1017/s1138741600006491

De Houwer, J. (2014). A propositional model of implicit evaluation. *Social and Personality Psychology Compass*, *8*(7), 342–353. 10.1111/spc3.12111

De Houwer, J. (2018). A functional–cognitive perspective on the relation between conditioning and placebo research. *Neurobiology of the Placebo Effect Part I*, *138*, 95–111. Elsevier. 10.1016/bs.irn.2018.01.007

De Houwer, J., Van Dessel, P., & Moran, T. (2020). Attitudes beyond associations: On the role of propositional representations in stimulus evaluation. In B. Gawronski (Ed.), *Advances in experimental social psychology* (Vol. 61, pp. 127–183). Elsevier. 10.1016/bs.aesp.2019.09.004

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*(1), 5–18. 10.1037//0022-3514.56.1.5

Dimofte, C. V. (2010). Implicit measures of consumer cognition: A review. *Psychology & Marketing*, *27*(10), 921–937. 10.1002/mar.20366

Dovidio, J. F., Evans, N., & Tyler, R. B. (1986). Racial stereotypes: The contents of their cognitive representations. *Journal of Experimental Social Psychology*, *22*(1), 22–37. 10.1016/0022-1031(86)90039-9

Drew, M. R., Denny, C. A., & Hen, R. (2010). Arrest of adult hippocampal neurogenesis in mice impairs single- but not multiple-trial contextual fear conditioning. *Behavioral Neuroscience*, *124*(4), 446–454. 10.1037/a0020081

Fan, X., Bodenhausen, G. V., & Lee, A. Y. (2021). Acquiring favorable attitudes based on aversive affective cues: Examining the spontaneity and efficiency of propositional evaluative conditioning. *Journal of Experimental Social Psychology*. Advance online publication. 10.1016/j.jesp.2021.104139

Faure, R., McNulty, J. K., Hicks, L. L., & Righetti, F. (2020). The case for studying implicit social cognition in close relationships. *Social Cognition*, *38*(Supplement), s98–s114. 10.1521/soco.2020.38.supp.s98

Fazio, R. H. (2007). Attitudes as object–evaluation associations of varying strength. *Social Cognition*, *25*(5), 603–637. 10.1521/soco.2007.25.5.603

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, *69*(6), 1013–1027. 10.1037//0022-3514.69.6.1013

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, *50*(2), 229–238. 10.1037/0022-3514.50.2.229

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in Experimental Social Psychology* (Vol. 21, pp. 1–74). 10.1016/s0065-2601(08)60317-2

Förderer, S., & Unkelbach, C. (2013). On the stability of evaluative conditioning effects: The role of identity memory, valence memory, and evaluative consolidation. *Social Psychology*, *44*(6), 380–389. 10.1027/1864-9335/a000150

Förderer, S., & Unkelbach, C. (2015). Attribute conditioning: Changing attribute assessments through mere pairings. *The Quarterly Journal of Experimental Psychology B*, *68*(1), 144–164. 10.1080/17470218.2014.939667

Förderer, S., & Unkelbach, C. (2016). Changing US attributes after CS–US pairings changes CS attribute assessments. *Personality and Social Psychology Bulletin*, *42*(3), 350–365. 10.1177/0146167215626705

Gaertner, S. L., & McLaughlin, J. P. (1983). Racial stereotypes: Associations and ascriptions of positive and negative characteristics. *Social Psychology Quarterly*, *46*(1), 23–30. 10.2307/3033657

Gast, A., & Rothermund, K. (2011a). I like it because I said that I like it: Evaluative conditioning effects can be based on stimulus-response learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *37*(4), 466–476. 10.1037/a0023077

Gast, A., & Rothermund, K. (2011b). What you see is what will change: Evaluative conditioning effects depend on a focus on valence. *Cognition & Emotion*, *25*(1), 89–110. 10.1080/02699931003696380

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*(5), 692–731. 10.1037/0033-2909.132.5.692

Gawronski, B., & Bodenhausen, G. V. (2018). Evaluative conditioning from the perspective of the associative–propositional evaluation model. *Social Psychological Bulletin*, *13*(3), e28024. 10.5964/spb.v13i3.28024

Gawronski, B., Geschke, D., & Banse, R. (2003). Implicit bias in impression formation: Associations influence the construal of individuating information. *European Journal of Social Psychology*, *33*(5), 573–589. 10.1002/ejsp.166

Gawronski, B., Hu, X., Rydell, R. J., Vervliet, B., & De Houwer, J. (2015). Generalization versus contextualization in automatic evaluation revisited: A meta-analysis of successful and failed replications. *Journal of Experimental Psychology: General*, *144*(4), e50–e64. 10.1037/xge0000079

Gawronski, B., & Quinn, K. A. (2013). Guilty by mere similarity: Assimilative effects of facial resemblance on automatic evaluation. *Journal of Experimental Social Psychology*, *49*(1), 120–125. 10.1016/j.jesp.2012.07.016

Gawronski, B., Rydell, R. J., De Houwer, J., Brannon, S. M., Ye, Y., Vervliet, B., & Hu, X. (2018). Contextualized attitude change. In James M. Olson (Ed.), *Advances in Experimental Social Psychology* (Vol. 57, pp. 1–52). Elsevier. 10.1016/bs.aesp.2017.06.001

Gershman, S. J. (2015). A unifying probabilistic view of associative learning. *PLoS Computational Biology*, *11*(11), e1004567-20. 10.1371/journal.pcbi.1004567

Gershman, S. J., & Niv, Y. (2010). Learning latent structure: Carving nature at its joints. *Current Opinion in Neurobiology*, *20*(2), 251–256. 10.1016/j.conb.2010.02.008

Gershman, S. J., Norman, K. A., & Niv, Y. (2015). Discovering latent causes in reinforcement learning. *Current Opinion in Behavioral Sciences*, *5*, 43–50. 10.1016/j.cobeha.2015.07.007

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480. 10.1037//0022-3514.74.6.1464

Harris, L. T., & Fiske, S. T. (2006). Dehumanizing the lowest of the low: Neuroimaging responses to extreme out-groups. *Psychological Science*, *17*(10), 847–853. 10.1111/j.1467-9280.2006.01793.x

Hastie, R., Ostrom, T. M., Ebbesen, E., Wyer, R., Hamilton, D., & Carlston, D. (1980). *Person memory: The cognitive basis of social perception*. Psychology Press.

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. Wiley.

Higgins, E. T., Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology*, *13*(2), 141–154. 10.1016/S0022-1031(77)80007-3

Kerpelman, J. P., & Himmelfarb, S. (1971). Partial reinforcement effects in attitude acquisition and counterconditioning. *Journal of Personality and Social Psychology*, *19*(3), 301–305. 10.1037/h0031447

Kurdi, B., & Banaji, M. R. (2017). Repeated evaluative pairings and evaluative statements: How effectively do they shift implicit attitudes? *Journal of Experimental Psychology: General*, *146*(2), 194–213. 10.1037/xge0000239

Kurdi, B., & Dunham, Y. (2020). Propositional accounts of implicit evaluation: Taking stock and looking ahead. *Social Cognition*, *38*(Supplement), s42–s67. 10.1521/soco.2020.38.supp.s42

Kurdi, B., & Dunham, Y. (2021). Sensitivity of implicit evaluations to accurate and erroneous propositional inferences. *Cognition*. Advance online publication. 10.1016/j.cognition.2021.104792

Kurdi, B., Gershman, S. J., & Banaji, M. R. (2019). Model-free and model-based learning processes in the updating of explicit and implicit evaluations. *Proceedings of the National Academy of Sciences*, *116*(13), 6035–6044. 10.1073/pnas.1820238116

Kurdi, B., Krosch, A. R., & Ferguson, M. J. (2020). Implicit evaluations of moral agents reflect intent and outcome. *Journal of Experimental Social Psychology*, *90*, 103990–12. 10.1016/j.jesp.2020.103990

Kurdi, B., Mann, T. C., Axt, J. R., & Ferguson, M. J. (2021a). *The prospect of long-term implicit attitude change: Tests of spontaneous recovery and reinstatement*. [Manuscript in preparation]. Department of Psychology, Yale University.

Kurdi, B., Mann, T. C., & Ferguson, M. J. (2021b). Persuading the implicit mind: Changing negative implicit evaluations with an 8-minute podcast. *Social Psychological and Personality Science*. Online first publication. 10.1177/19485506211037140

Kurdi, B., Morris, A., & Cushman, F. A. (2021c). *The role of causal structure in implicit evaluation*. PsyArXiv. 10.31234/osf.io/r7cfa

Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J.-E. L., Joy-Gaba, J. A., Ho, A. K., Teachman, B. A., Wojcik, S. P., Koleva, S. P., Frazier, R. S., Heiphetz, L., Chen, E. E., Turner, R. N., Haidt, J., Kesebir, S., Hawkins, C. B., Schaefer, H. S., Rubichi, S., … Nosek, B. A. (2014). Reducing implicit racial preferences: I. A comparative investigation of 17 interventions. *Journal of Experimental Psychology: General*, *143*(4), 1765–1785. 10.1037/a0036260

Lai, C. K., Skinner, A. L., Cooley, E., Murrar, S., Brauer, M., Devos, T., Calanchini, J., Xiao, Y. J., Pedram, C., Marshburn, C. K., Simon, S., Blanchar, J. C., Joy-Gaba, J. A., Conway, J., Redford, L., Klein, R. A., Roussos, G., Schellhaas, F. M. H., Burns, M., … Nosek, B. A. (2016). Reducing implicit racial preferences: II. Intervention effectiveness across time. *Journal of Experimental Psychology: General*, *145*(8), 1001–1016. 10.1037/xge0000179

Lindgren, K. P., Baldwin, S. A., Peterson, K. P., Wiers, R. W., & Teachman, B. A. (2020). Change in implicit alcohol associations over time: Moderation by drinking history and gender. *Addictive Behaviors*, *107*, 106413. 10.1016/j.addbeh.2020.106413

Mann, T. C., & Ferguson, M. J. (2015). Can we undo our first impressions? The role of reinterpretation in reversing implicit evaluations. *Journal of Personality and Social Psychology*, *108*(6), 823–849. 10.1037/pspa0000021

Mann, T. C., & Ferguson, M. J. (2017). Reversing implicit first impressions through reinterpretation after a two-day delay. *Journal of Experimental Social Psychology*, *68*(C), 122–127. 10.1016/j.jesp.2016.06.004

Mann, T. C., Kurdi, B., & Banaji, M. R. (2020). How effectively can implicit evaluations be updated? Using evaluative statements after aversive repeated evaluative pairings. *Journal of Experimental Psychology: General*, *149*(6), 1169–1192. 10.1037/xge0000701

McConnell, A. R., & Rydell, R. J. (2014). The systems of evaluation model: A dual-systems approach to attitudes. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 204–217). Guilford Press.

McConnell, A. R., Rydell, R. J., Strain, L. M., & Mackie, D. M. (2008). Forming implicit and explicit attitudes toward individuals: Social group association cues. *Journal of Personality and Social Psychology*, *94*(5), 792–807. 10.1037/0022-3514.94.5.792

McLoughlin, N., & Over, H. (2017). Young children are more likely to spontaneously attribute mental states to members of their own group. *Psychological Science*, *28*(10), 1503–1509. 10.1177/0956797617710724

Medin, D. L., Altom, M. W., & Murphy, T. D. (1984). Given versus induced category representations: Use of prototype and exemplar information in classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(3), 333–352. 10.1037/0278-7393.10.3.333

Meersmans, T., De Houwer, J., Baeyens, F., Randell, T., & Eelen, P. (2005). Beyond evaluative conditioning? Searching for associative transfer of nonevaluative stimulus properties. *Cognition & Emotion*, *19*(2), 283–306. 10.1080/02699930441000328

Merriman, J., Rovee-Collier, C., & Wilk, A. (1997). Exemplar spacing and infants' memory for category information. *Infant Behavior and Development*, *20*(2), 219–232. 10.1016/s0163-6383(97)90024-2

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*(2), 227–234. 10.1037/h0031564

Mitchell, J. P., Banaji, M. R., & Macrae, C. N. (2005). General and specific contributions of the medial prefrontal cortex to knowledge about mental states. *NeuroImage*, *28*(4), 757–762. 10.1016/j.neuroimage.2005.03.011

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences*, *32*(02), 183–198. 10.1017/s0140525x09000855

Moran, T., Bar-Anan, Y., & Nosek, B. A. (2015). Processing goals moderate the effect of co-occurrence on automatic evaluation. *Journal of Experimental Social Psychology*, *60*(C), 157–162. 10.1016/j.jesp.2015.05.009

Moran, T., Bar-Anan, Y., & Nosek, B. A. (2017). The effect of the validity of co-occurrence on automatic and deliberate evaluations. *European Journal of Social Psychology*, *46*(6), 1101–1116. 10.1002/ejsp.2266

Moskowitz, G. B. (1993). Person organization with a memory set: Are spontaneous trait inferences personality characterizations or behaviour labels? *European Journal of Personality*, *7*(3), 195–208. 10.1002/per.2410070305

Moskowitz, G. B. (1996). The mediational effects of attributions and information processing in minority social influence. *British Journal of Social Psychology*, *35*(1), 47–66. 10.1111/j.2044-8309.1996.tb01082.x

Moskowitz, G. B., & Roman, R. J. (1992). Spontaneous trait inferences as self-generated primes: Implications for conscious social judgment. *Journal of Personality and Social Psychology*, *62*(5), 728–738. 10.1037/0022-3514.62.5.728

Moskowitz, G. B., Gollwitzer, P. M., Wasel, W., & Schaal, B. (1999). Preconscious control of stereotype activation through chronic egalitarian goals. *Journal of Personality and Social Psychology*, *77*(1), 167–184. 10.1037/0022-3514.77.1.167

Neely, J. H. (1976). Semantic priming and retrieval from lexical memory: Evidence for facilitatory and inhibitory processes. *Memory & Cognition*, *4*(5), 648–654. 10.3758/bf03213230

Newman, L. S. (1991). Why are traits inferred spontaneously? A developmental approach. *Social Cognition*, *9*(3), 221–253. 10.1521/soco.1991.9.3.221

Olcaysoy Okten, I., & Moskowitz, G. B. (2020). Easy to make, hard to revise: Updating spontaneous trait inferences in the presence of trait-inconsistent information. *Social Cognition*, *38*(6), 571–625. 10.1521/soco.2020.38.6.571

Olcaysoy Okten, I., Schneid, E. D., & Moskowitz, G. B. (2019). On the updating of spontaneous impressions. *Journal of Personality and Social Psychology*, *117*(1), 1–25. 10.1037/pspa0000156

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, *89*(3), 277–293. 10.1037/0022-3514.89.3.277

Peters, K. R., & Gawronski, B. (2011). Are we puppets on a string? Comparing the impact of contingency and validity on implicit and explicit evaluations. *Personality and Social Psychology Bulletin*, *37*(4), 557–569. 10.1177/0146167211400423

Pratto, F., & Bargh, J. A. (1991). Stereotyping based on apparently individuating information: Trait and global components of sex stereotypes under attention overload. *Journal of Experimental Social Psychology*, *27*(1), 26–47. 10.1016/0022-1031(91)90009-u

Ranganath, K. A., & Nosek, B. A. (2008). Implicit attitude generalization occurs immediately; explicit attitude generalization takes time. *Psychological Science*, *19*(3), 249–254. 10.1111/j.1467-9280.2008.02076.x

Reeder, G. D., & Brewer, M. B. (1979). A schematic model of dispositional attribution in interpersonal perception. *Psychological Review*, *86*(1), 61–79. 10.1037/0033-295x.86.1.61

Reeder, G. D., & Coovert, M. D. (1986). Revising an impression of morality. *Social Cognition*, *4*(1), 1–17. 10.1521/soco.1986.4.1.1

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black, & W. F. Prokasy (Eds.), *Classical conditioning II: Current theory and research* (pp. 64–99). Appleton-Century-Crofts.

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, *5*(4), 296–320. 10.1207/s15327957pspr0504_2

Rubinstein, R. S., & Jussim, L. (2019). Stimulus pairing and statement target information have equal effects on stereotype-relevant evaluations of individuals. *Journal of Theoretical Social Psychology*, *111*(1), 256–19. 10.1002/jts5.53

Rubinstein, R. S., Jussim, L., & Stevens, S. T. (2018). Reliance on individuating information and stereotypes in implicit and explicit person perception. *Journal of Experimental Social Psychology*, *75*, 54–70. 10.1016/j.jesp.2017.11.009

Ruff, C. C., & Fehr, E. (2014). The neurobiology of rewards and values in social decision making. *Nature Publishing Group*, *15*(8), 549–562. 10.1038/nrn3776

Rydell, R. J., & Jones, C. R. (2009). Competition between unconditioned stimuli in attitude formation: Negative asymmetry versus spatio-temporal contiguity. *Social Cognition*, *27*(6), 905–916. 10.1521/soco.2009.27.6.905

Rydell, R. J., & McConnell, A. R. (2006). Understanding implicit and explicit attitude change: A systems of reasoning analysis. *Journal of Personality and Social Psychology*, *91*(6), 995–1008. 10.1037/0022-3514.91.6.995

Rydell, R. J., McConnell, A. R., Mackie, D. M., & Strain, L. M. (2006). Of two minds: Forming and changing valence-inconsistent implicit and explicit attitudes. *Psychological Science*, *17*(11), 954–958. 10.1111/j.1467-9280.2006.01811.x

Rydell, R. J., McConnell, A. R., Strain, L. M., Claypool, H. M., & Hugenberg, K. (2007). Implicit and explicit attitudes respond differently to increasing amounts of counterattitudinal information. *European Journal of Social Psychology*, *37*(5), 867–878. 10.1002/ejsp.393

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in "theory of mind." *NeuroImage*, *19*(4), 1835–1842. 10.1016/s1053-8119(03)00230-1

Schapiro, A. C., McDevitt, E. A., Chen, L., Norman, K. A., Mednick, S. C., & Rogers, T. T. (2017). Sleep benefits memory for semantic category structure while preserving exemplar-specific information. *Scientific Reports*, *7*(1), 14869. 10.1038/s41598-017-12884-5

Schiller, D., Kanen, J. W., LeDoux, J. E., Monfils, M.-H., & Phelps, E. A. (2013). Extinction during reconsolidation of threat memory diminishes prefrontal cortex involvement. *Proceedings of the National Academy of Sciences*, *110*(50), 20040–20045. 10.1073/pnas.1320322110

Schiller, D., Monfils, M.-H., Raio, C. M., Johnson, D. C., LeDoux, J. E., & Phelps, E. A. (2010). Preventing the return of fear in humans using reconsolidation update mechanisms. *Nature*, *463*(7277), 49–53. 10.1038/nature08637

Shen, X., Mann, T. C., & Ferguson, M. J. (2020). Beware a dishonest face? Updating face-based implicit impressions using diagnostic behavioral information. *Journal of Experimental Social Psychology*, 86, 103888. 10.1016/j.jesp.2019.103888

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, *4*(2), 108–131. 10.1207/s15327957pspr0402_01

Sperber, D. (1994). The modularity of thought and the epidemiology of representations. In L. A. Hirschfeld, & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 29–67). Cambridge University Press.

Srull, T. K., & Wyer, R. S. (1979). The role of category accessibility in the interpretation of information about persons: Some determinants and implications. *Journal of Personality and Social Psychology*, *37*(10), 1660–1672. 10.1037/0022-3514.37.10.1660

Teachman, B. A., Clerkin, E. M., Cunningham, W. A., Dreyer-Oren, S., & Werntz, A. (2019). Implicit cognition and psychopathology: Looking back and looking forward. *Annual Review of Clinical Psychology*, *15*(1), 123–148. 10.1146/annurev-clinpsy-050718-095718

Todorov, A. T., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology*, *66*(1), 519–545. 10.1146/annurev-psych-113011-143831

Todorov, A. T., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, *12*(12), 455–460. 10.1016/j.tics.2008.10.001

Trafimow, D., & Schneider, D. J. (1994). The effects of behavioral, situational, and person information on different attribution judgments. *Journal of Experimental Social Psychology*, *30*(4), 351–369. 10.1006/jesp.1994.1017

Uleman, J. S., Hon, A., Roman, R. J., & Moskowitz, G. B. (1996). On-line evidence for spontaneous trait inferences at encoding. *Personality and Social Psychology Bulletin*, *22*(4), 377–394. 10.1177/0146167296224005

Van Dessel, P., De Houwer, J., & Gast, A. (2015). Approach–avoidance training effects are moderated by awareness of stimulus–action contingencies. *Personality and Social Psychology Bulletin*, *42*(1), 81–93. 10.1177/0146167215615335

Van Dessel, P., De Houwer, J., Roets, A., & Gast, A. (2016). Failures to change stimulus evaluations by means of subliminal approach and avoidance training. *Journal of Personality and Social Psychology*, *110*(1), e1–e15. 10.1037/pspa0000039

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review*, *107*(1), 101–126. 10.1037//0033-295x.107.1.101

Wimmer, G. E., Li, J. K., Gorgolewski, K. J., & Poldrack, R. A. (2018). Reward learning over weeks versus minutes increases the neural representation of value in the human brain. *Journal of Neuroscience*, *38*(35), 7649–7666. 10.1523/jneurosci.0075-18.2018

Wimmer, G. E., & Poldrack, R. A. (2021). Reward learning and working memory: Effects of massed versus spaced training and post-learning delay period. *Memory & Cognition*. Online first publication. 10.3758/s13421-021-01233-7

Winter, L., & Uleman, J. S. (1984). When are social judgments made? Evidence for the spontaneousness of trait inferences. *Journal of Personality and Social Psychology*, *47*(2), 237–252. 10.1037//0022-3514.47.2.237

Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences*, *104*(20), 8235–8240. 10.1073/pnas.0701408104

Afterward

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

23 Impressions of Impression Formation

James S. Uleman
New York University

With Gratitude and Joy

This book was a complete surprise when the editors proposed it to me, after they'd lined up a publisher and many contributors, and it prompted some reflection on my career. I'd already announced my retirement, after almost 50 years at NYU and 55 as a social psychologist. Like most careers in academia, mine has had its ups and downs. This book and all it represents is a huge "up."

I transferred out of Caltech to Michigan in 1959 after a proverbial sophomore slump and my mismatch with physics became apparent. At Caltech, Abe Maslow convinced me that psychology had a place for both science and humanism. My 1966 Harvard dissertation produced a measure of the need for power (n Power) based on the Thematic Apperception Test (TAT), but efforts to validate it as a motive measure largely failed. I was denied tenure at Michigan State in 1970, gained it at NYU in 1974, and directed the social psychology doctoral program from 1975 to 1997. But I floundered otherwise until social cognition came along, and to NYU in 1981 in the persons of colleagues Bargh, Chaiken, Higgins, and Ruble. Attending Bibb Latané's Nag's Head Conferences on social cognition, starting in 1984, gave me my reference group, supplemented over the years by Dave Hamilton and Eliot Smith's Person Memory Interest Group meetings and Don Carlston's Duck Conferences. Mel Manis was the editor at the *Journal of Personality and Social Psychology* who accepted the first paper on spontaneous trait inferences (STIs, Winter & Uleman, 1984). I have been gifted with many terrific students, colleagues, collaborators, and critics, some of whom contributed to this book (Bargh, Ferguson, Ferreira, Garcia-Marques, Hamilton, Henderson, Moskowitz, Newman, Olcaysoy Okten, Orghian Ramos, Shen, Skowronski, Todorov, Trope, and Zárate). Others have grabbed my attention with their research, opinions, and challenges, and in one case her sheer enthusiasm. So, my gratitude is great, for this book and for all those who have found something of interest in my work, and who have helped build the intellectual tradition of social cognition that I call home.

I must also acknowledge that Marilyn, my wife and love of 60 years, my three children, and now three grandchildren have sustained me throughout.

DOI: 10.4324/9781003045687-27

The chapters contained in this volume are a joy because they show how the social cognitive tradition is flourishing. Stereotypes and stereotyping are major topics, as they have been in social psychology since the 1930s, reflecting social psychology's concern with the brutal and tragic American traditions of colonialism, slavery, and racism. Stereotypes and their relation to spontaneous trait inferences (STIs) are also my focus, in another chapter on research with Alex Todorov and Celia Gonzalez. But these chapters are much more wide ranging. Reading them is like being in a seminar with a bunch of very smart scholars, and friends in many cases. So as in a seminar, I summarize them briefly below. The exchanges they invite must await another venue. As you read the chapters in this volume, pay attention to the clear thinking behind them, the incisive questions asked, the innovative methods developed to address them, and the many exciting avenues for future research. So many studies to do and so little time!

The Future of Impression Formation

Where is the field of impression formation going? I have written an overview of where it has been (Uleman & Kressel, 2013), but it keeps surprising me. And predicting the future of the field is much harder that predicting the outcome of any study I have done. Future developments will certainly be more interesting, as new models and methods are developed. (Almost anything will be more interesting than the ancient dispute over whether an additive or averaging model best predicts evaluative impressions of others (Anderson, 1974). It has been said that the field of "person memory" (Hastie et al., 1980) was launched in reaction against the sterility of Anderson's approach, and it became a foundation of social cognition.) So, knowing that my predictions are as likely to be wrong as not, here are some thoughts.

We are both perceivers and the "objects" being perceived. This makes person perception unlike the perception or cognition of other objects. Categorizing simple objects (e.g., small red triangles and large green circles) and solving problems in logic (e.g., if her action is high in consensus, distinctiveness, and consistency, is the locus of causality in her or not?) are unlike inferring things about other people. We can imagine ourselves in other people's place and draw on those experiences to "understand" them. Although only one chapter in this book focuses on this, there is abundant research that does. Research on simulation theory (Shanton & Goldman, 2010), which includes perspective taking (e.g., Epley et al., 2004), is based on the idea that perceivers assume similarity with others in several distinct ways. Such simulation of others' experiences can provide the bases for understanding other people in ways that we do not use for inanimate objects (unless we anthropomorphize them as some do about "history," social systems, and nations).

Social stereotypes have played a large role in research on impression formation and will continue to do so. Both Brewer (1988) and Fiske and Neuberg (1990) claimed that we first categorize others by the social categories of gender,

race, and age, and only then attend to other things including individuating information such as traits. Although these claims have been challenged (Kunda & Thagard, 1996; Monroe et al., 2018), stereotypes have been of interest to social psychologists since the journalist Walter Lippmann (1922) first used the term (e.g., Katz & Braly, 1933). Since then, major efforts have been devoted to describing the contents of various stereotypes and how they impact impressions, prejudice, and discrimination. After all, when we meet new people, their gender, race, and age are usually obvious immediately. But social psychologists also look for general processes and effects of stereotyping at the individual and interpersonal levels that cut across particular content (e.g., Schneider, 2004). Every stereotype is freighted with its own history and role in the relations between oppressor and oppressed. The distribution of melanin in our species, the predation of western colonialism, the rise of racialized enslavement and its legacy of white supremacy have ensured that the stereotype of African Americans serves as a basic category of impression formation in the United States today. Stereotypes of Native Americans, East Asians, Southern Europeans, Puerto Ricans, Mexicans, and many others have their own content and histories. Patriarchy has its own consequences. And there are stereotypes of occupational groups, socio-economic groups, handicapped groups, and regional groups in most societies; caste is important in some.

The chapters in this book prompted more specific thoughts. Here they are alphabetically by author. Further thoughts prompted by other research programs follow.

Arndt and Henderson consider gender's impact on impressions when gender is not obvious. This opens up a host of interesting issues, even within a binary (male/female) framework. One is whether or not gender is a "*mandatory inference*," a category that most people always infer or assume because ambiguity is inconceivable or intolerable in some way. Mandatory inferences may include that others are sane (whatever that means), moral (trustworthy), and rational (conventional?) until proven otherwise (see also Tamir et al., 2016). Such categories are broadly informative and diagnostic of narrower traits of interest, and provide the framework for most social interaction and the starting points for thinking about others. They may exist at the top of a hierarchy of characteristics, akin to the Big Five in personality. They may include race and age as well. But there is little research in social psychology on the concept of mandatory inferences—default values that always accompany perceptions of personhood—and how to measure them. A second set of issues concerns the socio-political conditions that make some categories "basic" or mandatory. Gender, race, and age are all bases on which it is illegal to discriminate in the United States, but it has not always been so. How much does the basicness of these categories depend on particular social groups' and their allies' relation to political power? In some communities, religious affiliation is basic and often signaled by clothing and other customs. Caste is critical in some. In contemporary American society, political party affiliation seems basic to many. Social dominance theory (Sidanius & Pratto, 2011), for instance, takes age and

gender as basic and treats other categories such as race as relatively arbitrary, from an evolutionary point of view. It should be possible to operationalize mandatory inferences and empirically study how they vary with social conditions, including with histories such as slavery in the United States.

Several chapters deal with changes in impressions over time—an important emerging area of research (that I have neglected completely in my own). **Bray, Armenta, and Zárate** describe effects of memory consolidation over hours or days, on impressions learned explicitly from photos and "news stories," and tested later with lexical decisions primed with actor photos. When memories integrate with existing knowledge over time, their accessibility shifts depending (e.g., on whether the actor is associated with an ingroup or outgroup). More generally, other possible changes as a result of "integration" should be explored with implicit measures such as lexical decision. Integrated memories seem to be more accessible, and more likely to generalize to new information. This research suggests the importance of spontaneous changes in memory for impressions over time, and their dependence on preexisting and subsequently encountered information. Such nonconscious and unintentional changes in memory are a particularly rich area for future research and could shed light on the spontaneous activity of cognitive structures. Are these results restricted to impressions of others? Do they depend on self relevance?

Chaney, Sanchez, and Remedios review research on people's ability to intentionally control prejudice, and how to motivate them to do so. While there are many lines of research on reducing stereotyping and discrimination, this work is among the most interesting because interpersonal processes are at its center. The focus is on how people respond to being told that they were prejudiced, i.e., prejudice confrontations. On the *in vivo* versus *in vitro* spectrum, this research is clearly *in vitro*. Guilt is the most common response to confrontation, but probably only in social contexts where the norms are anti-prejudice. Confrontations about racism on the IAT can be counterproductive, producing anger and denial, perhaps because the behavior is so unintended. Out-group members, relative to members of the group offended, are more effective confronters. Guilt over time, along with "reflective pondering rumination" (not brooding rumination), reduces prejudice. Intersectional research is just beginning. As in most *in vitro* research, there are many variables in multiple contexts, and few neat theories. But this work is critical in laying out pathways to a less brutal and more just society.

Chen, Quinn, and Maddox describe several ways in which trait and stereotype inferences are similar, how they differ, and how they do and might affect the processing of information about others. They advance many testable ideas about their interactions and functions, especially over time, in creating spontaneous impressions of others. Because I am interested in inferences of behaviors' meanings, we also need research on the ways that the meanings of stereotypes can affect the meanings of behaviors. Whereas "He said that his car is the very best model on the road" might imply *boastful*,

"The salesman said that his car …" might imply *pushy* or *aggressive*. "The banker offered to loan him $5,000" implies different traits from "The loan shark offered…" Older research has shown that explicit interpretations of actions can depend on actors' identities, as in the kinds of "aggression" expected of lawyers versus teamsters (Kunda & Sherman-Williams, 1993; Kunda et al., 1997). But there are no such explorations of spontaneous trait inferences. This chapter also highlights the importance and benefits of increasing the racial diversity both of stimuli and of scientists in this area.

"Impression management" is documented in a large research literature that is usually not considered in conjunction with "impression formation." But **Dupree** uniquely does this, in the specific context of multi-racial and multi-gendered groups. She is particularly engaged by situations that prompt us to behave in counter-stereotypic ways, according to the meta-stereotypes that we hold about other people's beliefs about us. Sometimes this is in the service of "upshifting" their impressions of us along some dimension, but other situations (audiences, groups, status positions) prompt efforts at downshifting. Impressions form and are managed and form again until, in the best case, some tentative equilibrium is reached. Stereotypes of race and gender, which have changed over the past 200 years, make this dance particularly interesting. Measuring dynamically shifting states in dyadic interactions, especially without being intrusive, is extremely challenging. But evolving methods for the study of synchrony (e.g., McAssey et al., 2012), as well as conceptual developments (e.g., Semin, 2007; Smith & Collins, 2009), open exciting avenues for future research on impression formation within social groups.

Garcia-Marques, Ferreira, Hagá, Marcelo, Ramos, and Orghian report on the latest developments from the Cognition in Context program at the University of Lisbon. They provide a lovely introduction to the historic roots of STI that makes STI seem almost inevitable, but it did not feel that way at the time. The CogCon group's emphasis on language, text comprehension, and linguistics is exciting, already fruitful, and close to my own thinking. Virtually all my STI sentences use direct action verbs (Semin & Fiedler, 1988) in the active voice, because intuitively comprehension seems easiest. The CogCon group has discovered effects of the active versus passive voice, and adverbs of manner (Marcelo), and suggests that varying other linguistic features can reveal important cognitive processes when tied to relevant linguistic theory. Orghian et al. (2017) developed a new measure of STI that is completely implicit, unlike the false recognition and savings paradigms. (Incidentally, I informally explored Latent Semantic Analysis (http://lsa.colorado.edu/), looking for clues to how people extract meaning from brief phrases or sentences. It did not produce trait inferences.) I am sure Garcia-Marques et al. are on the right track and well equipped to follow these paths. This chapter piqued my interest in linguistics and language again. Maybe listening to my fado recordings will help me think. Finally, I must mention that STI and I found a home away from home in Lisbon. I have visited

Lisbon several times, and the chapter authors Leonel, Tania, Diana, and Sara have spent significant time in New York. These exchanges have been great.

Gawronski, Brannon, and Luke describe research on unintended effects of the mere co-occurrence of valenced stimuli on intentional impression evaluations. They expected implicit measures to reflect mere co-occurrence, whereas explicit measures (judgments) would reflect propositional relations (e.g., causes, is part of, etc.) between the two (e.g., lotion X prevents skin cancer). While the phenomena occur, results in the literature are inconsistent. Multinomial modeling of responses on a single task (rather than on implicit and explicit measures) resolved some but not all of these inconsistencies. These studies and theorizing illustrate the power of bringing together clear psychological theories (the associative-propositional evaluation, APE, and the integrative propositional model, IPM), mathematical modeling of possible results, and a sequence of studies that vary relevant parameters (e.g., study time, repetition, response time, delays in responding, etc.). By focusing on an apparently simple phenomenon along with articulate theories, this work shows an unusual degree of theoretical rigor and sophistication, as contrasted with the more common pattern in impression formation research that merely demonstrates phenomena. A comparable paper on spontaneous trait inferences that comes to mind is McCarthy and Skowronski (2011).

Hamilton and Thurston, after a brief review of research on impression formation, describe Hamilton's discovery (Hamilton et al., 2015) of spontaneous trait inferences about groups (STIGs), inferred from groups' behaviors in the same way that STIs are inferred from individuals' behaviors. They propose that these may be the beginnings (seeds) of stereotypes. Of course, this is an idealization because all (or most) of our stereotypes are communicated to us by those who transmit our cultures and history, long since formed in our linguistic communities. But the idealization is important, and it was tested and confirmed with fictional groups. One of the surprising findings in Hamilton et al. (2015, Exps. 3 and 4) is that group entitativity had no effect on STIG formation. Neither Dave nor I predicted this. I thought that just as it is difficult to infer traits from the behavior of actors who have disintegrated personalities, i.e., are schizophrenic (Levey et al., 1995), it should be difficult to infer traits about low entitativity groups. Perhaps there are trait-implying "behaviors" that are implausible for very low entitativity groups, but that imply traits for individuals. More research…

Kurdi and Banaji review theory and research on the formation and revision of implicit evaluations of people, and find that simple associative processes fall short and that more complex propositional processes are called for. They then outline unresolved questions for the future. (Do they assume associative processes are implicit, and propositional processes are explicit?—a conflation best avoided.) Their first question is whether and if there are domain-specific processes for social information that differ from domain-general processes. Their exploration of this question is provocative and illuminating, and highlights the need for more theoretical clarity that generates new empirical work.

What role does mentalizing play? How stable over time and context are implicit evaluations? What might the role be of reconsolidation processes? What effects do encoding conditions have, including goals, cognitive capacity, and massed versus distributed learning? And what kinds of mental representation are involved? These are important questions. Amassing relevant research in this chapter is an important step in answering them. However, if we judge from research by winners of the Ostrom Award given by the *Person Memory* Interest Group (https://pmigconference.com/ostrom-award) or the Hastie et al. (1980) book by that same name, "implicit *person memory*" is too broad a designation for the implicit *evaluations* covered here.

Ludwin-Peery and Trope focus on predictive coding models, which they essentially treat as the modern version of expectancies (schema, stereotypes, scripts, etc.). They then apply this metaphor (because it remains that, as this level of generality) to findings from Construal Level Theory and other phenomena including perceptual and cognitive illusions. Implications for impression formation are described. The predictive coding metaphor potentially unites these traditionally diverse areas and is exciting for that reason. The next steps will be to go from metaphor to model by, for example, figuring out how to gather data that such models can test rigorously. Quantitative models are almost as old as social psychology and have played an important role in impression formation. They may merely attempt to predict behavioral outcomes (e.g., Anderson, 1974) or model processes as well. They can be used to convey the plausibility of theoretical arguments by, for example, showing that mere associations can produce output that mimics both STI and STT (spontaneous trait transfer) effects (Orghian et al., 2015). They can be used to separate effects of multiple processes, as in Jacoby's PDP model (McCarthy & Skowronski, 2011) or multinomial modeling (Gawronski, Brannon, & Luke, this volume). And they can model cognitive processes themselves. Predictive coding models attempt to do this in ways that are neurologically plausible. That is their promise and power. But to realize this promise, researchers must formalize such models and design studies to collect data that allow for tests of competing versions. Until we do that, predictive coding remains a metaphor.

Melnikoff and Bargh use the metaphor of a painting accidentally inverted when mounted on a museum wall to suggest that STI research has reoriented research on automaticity. It reminds me of one of Bargh's early studies of priming, a wall-mounted experiment designed to explore interspecies boundary conditions. In his own whimsical way of which I'm well aware, Bargh mounted small printed primes at ankle level on the walls (to protect humans in the area). We had mice in the building, and the primes were DIE and DEATH. It didn't work. No mice died (but neither did humans).

But seriously, Melnikoff and Bargh's chapter touches on two important ideas. "Predictive coding" is in the wind (see Ludwin-Peery & Trope, this volume), along with "model free and model-based learning." Once you admit the importance of inferences, beyond mere associations, the question

becomes how the inferences are made, and where the models come from. McKoon and Ratcliff (1986) showed spontaneous predictive inferences in text comprehension long ago, inspiring some of the STI work. Quantitative predictive coding models are prominent in theories of perception along with model-based learning. They blur the boundaries between perception and cognition. All such models assume conditionally automatic inferences from inputs, whether observed (Foulk et al., 2016) or described (STI) behaviors, or sensory input. The $64K question is how the inferences are made. I am sure the details will differ by domain (and the details are essential for adequate quantitative tests). But in the social domain, language is surely involved. Bayesian inference has been offered as a model (e.g., Tenenbaum et al., 2011) but it has serious limitations (Marcus & Davis, 2013, 2015). Inferences about social events occur but we do not know how. I was amazed that Foulk et al. (2016) found spontaneous activation of a trait concept from simply observing behavior. Is there a "language of behavior," or of music, with similar properties to the language of words? How do words encapsulate complex meanings so well, and so ambiguously (Uleman, 2005)? There are decades of research questions here.

Moskowitz, Olcaysoy Okten, and Schneid offer an unusually comprehensive review of research on updating initial impressions, including those based on traits, stereotypes, and affect (attitude). They distinguish implicit from explicit impressions; and inferences of traits, states, behavior gists, goals, roles, predictions, and evaluations. They consider blocks against updating. They distinguish judgments from memory measures, and distinguish changes in memory from adding information, negating information, and reconsolidation. Research on reconsolidation is a growing area in cognitive neuroscience, and applying these insights to the social cognition of initial impressions is particularly promising. The same can be said of the detection of inconsistencies from prior expectations (e.g., Na & Kitayama, 2011) and the consequences of inconsistency detection, including additional information processing. Research on perceivers' motives and cognitive capacity is also reviewed. Finally, the section on semantic versus evaluative systems updates work on this long-standing and important distinction within social psychology and impression formation research, in particular.

Newman and Marsden provide a detailed, critical review of culture's effects on STI, framed in terms of individualism-collectivism (I/C), including studies comparing samples from different nations, measures of individual differences, and priming. They note that nations differ on many variables besides I/C (geography is not a psychological variable); measures of individual differences correlate poorly with each other; and priming studies have disappointed. Uleman et al. (2000) found that "closeness" is the major component of most I/C scales, and that how collectivist or individualist one is with others depends on who they are, i.e., which group one considers (family, relatives, or friends). Although I am convinced that cultural differences exist in STI, I have become skeptical that the psychological variable

at work is I/C. Takano and Osaka (2018) published an extensive review of the I/C literature comparing North Americans and Japanese, and found no reliable difference. This raises the possibility that the continuing interest in I/C, almost to the exclusion of other cultural variables, is largely based on researchers' stereotypes (Uleman, 2018). It may be more fruitful to look for other mediators of the "culture" → STI relationship, such as attentional processes (Shimizu & Uleman, 2021), and perhaps to focus on the contributions of automatic processes (Shimizu et al., 2017). In any case, this chapter provides an excellent review and set of suggestions for future work.

Ratliff describes research on acquiring evaluations by association with evaluated others, and related phenomena: "attitude generalization, spreading attitude effect, higher-order conditioning, transference, stigma-by-association, etc." Implicit attitude transfer is based on similarity; group entitativity as well as family relationships affect perceived similarity. Both associative and propositional accounts of attitude transfer are discussed. Implicit and explicit measures of evaluation converge over time, reminiscent of work by Bray and Zárate (Chapter 20, this volume) on memory consolidation. Implicit evaluations generalize readily, and are related to stereotyping. In short, attitude transfer is important, ubiquitous and basic to many impression formation processes. It might be fruitful to explicitly compare evaluative and semantic inferences, i.e., evaluative transfer and STI, both of which have been discussed in terms of associations and propositions, explicit and implicit. Noting similarities and differences between measures might lead to theoretical and empirical advances.

Sands and Harris propose research on another sensory modality—auditory stimuli—and explore impressions based on persons' voices. They remind us of a model in use before the cognitive revolution in social psychology, Brunswick's lens model. It suggests assessing (a) the cues or features that a target person emits (voice qualities in this case such as pitch and shimmer) in order to convey particular impressions, (b) their validity as indicators of what the sender intends and what the perceiver is to judge (which are forced to be the same in this case), (c) the ability of the perceiver to detect them, and (d) the way the perceiver combines them to form a judgment. This model has been fruitfully used in many domains of person perception, as Hammond's work demonstrates (Hammond et al., 1966). Such a research program would add important information to our understanding of impressions based on voices. This approach is refreshing in getting out of perceivers' heads (interpretations and other cognitive processes) and assessing the objective utility of cues for judgments from particular modalities.

Shen and Ferguson focus on changes in evaluations of others conveyed by faces. Inferences from appearance seem to have high and immediate credibility, perhaps because there is no discernable "line of thought" from appearance to conclusion that can be reported or challenged. In fact, some dual process theories assert that mere associations rather than propositional knowledge characterize the links between appearance and evaluation. Such theories are

not consistent with the evidence that propositional information can change the evaluative implications of facial appearance, by affecting its diagnosticity, its reliability, and its interpretation. And these changes are more persistent than some dual process theories predict. Part of the importance of this experimental work lies in its demonstrations of how verbal and visual information can interact in producing impressions of others. How you "see" another person's scarred face—whether as an essentialistic feature or as the tragic result of an act of heroism—makes a large difference in the impression you form. Dual process theories that neglect or deny the interaction between various kinds of information are inadequate.

Sherman focuses on the cognitive processes involved in stereotypes' acquisition, their operation, and their dependence on inferences from the stereotyped group's behaviors. This is summarized by his Encoding Flexibility Model, a *tour de force* in analyzing the details of process over content, and the dependence of these processes on cognitive organization. Critical in this work is the distinction between what is revealed by recall versus recognition memory. Also relevant to stereotyping, but unmentioned here, is his application of Kruschke's attention theory of category learning to the development of stereotypes (Sherman et al., 2009). Among other things, it describes why the content of stereotypes so often emphasizes the features on which perceivers and others differ.

Shopshire, Gillespie, and Johnson bring two unique perspectives to the topic of categorizing others: JDM and self-relevance. JDM processes are largely conscious and intentional, whereas important impression formation processes are not. But this perspective may account for several known biases in explicitly categorizing others in terms of, e.g., sex, race, or sexual orientation—particularly those categorization biases that seem to involve self-relevance. It is also possible that conscious and spontaneous categorizations may differ, with the conscious ones more influenced by self-relevance. As I note elsewhere in this volume, Ham and van den Bos (2008) asked participants to make justice judgments about scenarios that did or did not involve themselves. Their conscious judgments did not distinguish between the two types; both versions—involving self or only other persons—were seen as unjust. But concepts of injustice were spontaneously (unconsciously and implicitly) activated only when the scenarios involved themselves. Future research should manipulate self-relevance and the stakes involved in a variety of ways, to explore the boundary conditions of such effects.

Skowronski and McCarthy describe the development of Carlston and Skowronski's (1994) savings-in-relearning paradigm for studying STI, and how it addressed two central issues that previous methods did not. It is a great example of how critical thinking about evidence for a phenomenon that you believe exists (or not) can produce novel research methods and findings, especially if confounds are vigorously pursued, which then lead to more hypotheses, etc. The chapter is full of interesting research suggestions, and describes the discovery of spontaneous trait transference (STT). Incidentally, a

recent meta-analysis of STI research (Bott et al., 2021) found no significant difference in effect sizes for the savings (0.62 [0.50, 0.74]) and the false recognition (0.75 [0.65, 0.86]) paradigms; they are equally sensitive to STIs. But the cognitive processes are different. Don and John's road trip was (is) an excellent adventure, and picked up other riders like Crawford, McCarthy, and many more along the way. This 30-year adventure also illustrates the importance of having interested collaborators, in it for the long haul. The small conferences that Don Carlston hosted at Duck, N.C., were invaluable. John S. and I have never co-authored a paper but have often reviewed each others' journal submissions. His reviews were always tough and fair (he often signed them), and I did my best to give as good as I got. We've both benefitted a lot, and from discussing science fiction early in our relationship, knew that the science was most important.

Todorov has done ground-breaking work on the information contained in (inferred from) faces (see also Todorov, 2017). He describes recent research on the neural encoding of both perceptual and social implications of faces; on the bases of the high degree of idiosyncratic inferences from faces; on faces' (vs. scenes') unique propensity to become associated with evaluative inferences; the impact of posture and clothing information on inferences from faces; etc. As usual, this work exemplifies the ways in which command of the tools of neuroscience and computational modeling extend the kinds of questions that can be conceived and empirically answered (see also the chapter by Shen and Ferguson). And it reminds me of the old questions I have had about the consequences of the mere presence (vs. absence) of faces. Do faces alone activate person-relevant or other social domain concepts, or not? Do they trigger preparatory motor responses? Good research always seems to stimulate interesting new questions.

The ability to form impressions of others is evident in the first year of life, well before infants have words! **Woo and Hamlin** critically review studies of the formation of impressions and evaluations of others in infancy, chiefly in the moral domain. Infants' ascription of mental states is crucial, e.g., in distinguishing intentional from accidental outcomes. Infants watch puppets helping or hindering or harming another puppet, or allocating resources to others. Their responses as measured by overt visual attention or preferential reaching are observed and provide evidence that others' mental states are inferred and affect the infants' reactions. Infants even distinguish between puppets that help intentionally versus only incidentally. Woo and Hamlin do not speculate directly on the representational system that infants use, although "theory of mind" (ToM) elements figure prominently in their account of these preverbal participants' behavior. But infants are spoken to in natural settings and understand language long before they produce it. Possible effects of caregivers' comments on infants' impressions while they observe others is worthy of future research (e.g., Shimizu et al., 2018). Developmental psychologists are making major contributions to our understanding of impression formation, particularly by infants and toddlers. As children develop,

simulation, ToM, language and concept development, and social processes must interact in complex ways (e.g., Baird & Astington, 2005). This is a vital area of future research. Finally, Hamlin's research always reminds me of the first time I met Kiley when she was a graduate student. I went to Yale to give a talk, and before I had time for anything else, she took me to her lab to see her puppets, helping or hindering other puppets. Her enthusiasm for this research, and its connections with my own, was contagious.

Comprehensive as the chapters in this handbook are, there are other developments in impression formation that are noteworthy. At the risk of appearing provincial, here are three from my own department at New York University that have my attention (by Rehder, Hackel, and Freeman).

Categories, Causes, and Bayes Nets

My colleague, Bob Rehder, has been working on categorization with Greg Murphy (a retired colleague) and Reid Hastie for the past three decades, using causal Bayes' nets to model categorization: how we infer categories from their features and vice versa, and the role of causal relations in defining categories. "If it builds nests and lays eggs but cannot fly, how likely is it to be a bird?" How do we infer categories from features and vice versa? Impression formation is all about inferences, categories, features, and causal relations among them. And reminiscent of STIs implicit character, Rehder and Burnett (2005) were "the first to address the specific role of causal knowledge in inferring the presence of an unobserved feature" (p. 299). Not unlike the discovery of the planet Neptune in 1846 by Le Verrier, through observing deviations in the orbit of the planet Uranus from that predicted by Newton's theory of gravitation, Rehder and Burnett (2005) "observed [in their data] a pervasive violation of one of the defining principles of Bayes' nets—the causal Markov condition—because the presence of characteristic features invariably led participants to infer yet another characteristic feature" (p. 264)—a feature that was never explicit, but which improved the modeling of their data on category and feature inferences when included in the causal model. That is, participants in their study (Experiment 5) made choices that indicated that they had inferred a causal feature in the net that was never explicit, but that made sense of the other information—not unlike trait inferences that make sense of behaviors. Although the empirical basis for concluding that participants inferred the existence of "an unobserved feature" is very different from the evidence for STIs, the idea is the same. Hidden causes like traits, which are never observed, can be inferred from other information, be the basis for predictions, and yet remain implicit. They can also be a basis for essentialism (Rehder, 2007).

The behaviors in STI studies are clearly categorized in trait terms, as generous, helpful, or honest, etc. The actors who perform them are inferred to have the traits of generosity, helpfulness, or honesty, etc., i.e., are categorized as generous, helpful or honest people. These categories (and others

such as stereotypes and their features) are causally related in perceivers' minds, in that traits cause behaviors (Kressel & Uleman, 2010, 2015). It is not clear where these categories come from, although language and culture surely play a large role. But given these categories and their features—which most adults have and use—and given people's beliefs in their causal relations with each other, we can use the power of the kind of Bayesian generative causal models that Rehder describes (Rehder & Kim, 2009) in impression formation research. Many of the familiar attribution phenomena involving multiple causes such as discounting and augmentation follow from these models easily, but they are much more precise and predictive than Kelley's formulations can be. They have more scope and power and can "discover" (but not name) implicit causes. They give precise meaning to the very fuzzy term "inference." We could also use them to incorporate other categories into our theories of impression formation and studies of STI, including stereotypes and their features, and theory of mind features such as belief, desire, and intention (e.g., Gopnik & Wellman, 2012). While the mathematics will intimidate most social psychologists (see Rehder, 2017a, 2017b), that is what colleagues and collaborators are for. The gist of these models is well conveyed by the diagrams and verbal descriptions, but their use in study designs and data analyses may require collaboration.

Interaction-Based Impressions in Evaluative and Semantic Memories

Leor Hackel, a former doctoral student in our program at New York University who collaborates with Dave Amodio, has been pursuing a research program that tracks how people evaluate interaction partners. Participants in the initial study (Hackel et al., 2015) had to choose to interact with one of several partners on each training trial and received rewards from the chosen partner's reward pool. Partners had various size pools, so that generosity (relative to the pool size) and actual rewards could vary orthogonally. Computational modeling showed that participants' choices of partners during subsequent test trials reflected both rewards received and implied trait generosity. Activity (fMRI) in different brain regions correlated with rewards received and implied generosity. Implied generosity also predicted the choice of partner for a future task involving cooperation but no rewards. The evidence supported the ideas that semantic (trait) and evaluative (reward) memories are stored in different places in the brain, consistent with older work on memories for different material in different places (Amodio & Ratner, 2011; Squire, 2004), and that interactions with others are critical in forming evaluations from rewards, i.e., when the interactions are instrumental. Importantly for me, participants were never asked to form impressions of their partners or to explicitly rate them on generosity or other traits. These inferences were implicit and spontaneous, and evidenced only through the design and data analyses.

Hackel and Amodio (2018) provide a fine primer on the kinds of evidence that computational modeling can reveal in social neuroscience. They focus particularly on impression formation that contrasts reinforcement (active, instrumental, operant) learning with observational (passive, inferential) learning. "Computational models allow researchers to probe trial-by-trial dynamics of learning and choice and to make precise quantitative predictions about behavior across time" (p. 92). This ability to track changes in impressions over time, and to do so unobtrusively from participants' responses on tasks without explicit reference to impression formation, is particularly attractive.

Hackel et al. (2020) used this framework to compare impressions of human verses slot machine partners. They found that during interactions, participants relied more on human's trait generosity than on rewards received, whereas with slot machines, received rewards were more important than trait generosity. This also held true for overall attitudes. Generous humans and rewarding slot machines were preferred, relative to rewarding humans and generous slot machines. Hackel, Mende-Siedlecki, and Amodio (2022), using the same trial-based data collection and analysis paradigm, looked at the extent to which trait-based preferences for partners were situation specific versus global. They found that "participants learned primarily from context-dependent traits gleaned from social interactions, secondarily from global traits, and least of all from rewards … traits, rather than rewards, provide a cognitive basis for forming stable, context-sensitive impressions from feedback in social interactions" (ms. p. 49).

Representational Similarity Analysis (RSA)

This method of data collection and analysis has become prominent in cognitive neuroscience. It first yields representational distance matrices (RDMs) depicting distances (or similarities or correlations) between responses (neural or behavioral or conceptual model) to pairs of stimuli of interest (see Figure 2, Popel et al., 2019). These can be based on individual participants with multiple trials per stimulus, or averages across participants. Then the similarity of these RDMs is computed, to see how well they represent each other. For example, one can ask how much the differences in responses to the stimuli of interest in one domain (e.g., neural) resemble differences in responses in another domain (e.g., behavioral). Thus, one *can* compare apples and oranges—not directly by some objective metric, but by a "second moment" derivative pattern of responses (Kriegeskorte et al., 2008). Researchers can collect, for example, taste ratings of apples and oranges, and data on purchases of apples and oranges, and then compute the resemblance of these two RDMs. This is useful in impression formation research because one

> can conceive of representations in social perception (e. g., social categories, emotions, traits) as points in a multidimensional space. The space can be measured using a variety of different modalities, such that the

> dimensions consist of neurons, fMRI voxels, or nodes in a computational model…. Although these multidimensional spaces from different modalities may be radically different in an absolute sense, it is valuable to estimate the extent to which a shared representational geometry (i.e., the pairwise distances among representations) is preserved.
>
> (Freeman et al., 2018, p. 83)

Spatial representations of social categories and traits are also evident in Freeman's mouse-tracking method, and seem especially useful in light of the inherent ambiguity of social categories, traits, etc. (Uleman, 2005). They are quite different from the discrete concept nodes connected by the activating and inhibitory links in Anderson and Bower's (1973) human associative memory (HAM) model, still prominent in social cognition.

My former colleague Jon Freeman uses RSA, along with his own mouse-tracking paradigm and the reverse correlation method for revealing images people have in mind (Dotsch et al., 2008), to develop support for his dynamic interactive theory of person perception (Freeman & Ambady, 2011; Freeman et al., 2020). This theory describes how "social-conceptual knowledge in particular can have a fundamental structuring role in how we perceive others' faces" (Freeman et al., 2020, p. 237), through a connectionist architecture that allows top-down sources to shape bottom-up perceptual processes, blurring the traditional distinction between cognition and perception. It describes how social-conceptual structures affect the perception of social categories, emotions, and traits in others' faces.

A series of studies showed "substantial overlap between the structures of perceivers' conceptual and social perceptual trait spaces, across perceptual domains… and that conceptual associations directly shape trait space…" (Stolier et al., 2020, p. 1). Studies showed that individual differences in conceptual trait spaces are reflected in face trait spaces. Moreover, the research team manipulated conceptual spaces through faux science articles, and found corresponding changes in face trait spaces, which provided evidence on how conceptual trait spaces are learned. Altogether, conceptual trait spaces seem to provide a general framework from which to make inferences about others from faces, stereotypes, voice, and other cues. This is supported in more recent research. Meshar, Stolier, and Freeman (in press) focused on perceptually ambiguous social categories (PASCs) such as alcoholics or gun owners; when they were more stereotypically associated with a trait, as is the case for alcoholics who are stereotypically associated with the trait extroversion, perceivers were more likely to infer PASC membership from faces conveying that trait. Further, they demonstrate that individual differences in the strength of trait-PASC stereotypes predicted face-based judgments and have a causal role in these judgments. These results imply that people can form any number of social category judgments from facial appearance alone by drawing on their learned social-conceptual associations, a conclusion that

requires RSA. More generally, there is a way to study what my intuition says are perceivers' "dynamic interactive theory of person perception."

Nothing is more fundamental to social psychology than our impressions of those with whom we interact. These impressions are both explicit and implicit, conscious and unconscious, intended and spontaneous. They are the product of our cultures, our experiences, our innate attunement to and dependence on each other, and the ways our minds parse and integrate all of this, moment by moment. As a field of scientific research and scholarship, they seem inexhaustible. Most of the human drama can be seen through their lens. Unlike the knowledge offered by the arts, scientific knowledge of them is cumulative, falsifiable, and public. Uncovering their richness has only just begun.

References

Amodio, D. M., & Ratner, K. G. (2011). A memory systems model of implicit social cognition. *Current Directions in Psychological Science*, *20*(3), 143–148. 10.1177/0963721411408562

Anderson, J. R., & Bower, G. H. (1973). *Human associative memory*. Washington, DC: Winston & Sons.

Anderson, N. H. (1974). Cognitive algebra. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 7, pp. 1–101). New York: Academic Press.

Baird, J. A., & Astington, J. W. (2005). The development of the intention concept: From the observable world to the unobservable mind. In R. R. Hassin, J. S. Uleman, & J. A. Bargh (Eds.)., *The new unconscious* (pp. 256–276). New York: Oxford University Press.

Bott, A., Brockmann, L., Denneberg, I., Henken, E., Kuper, N., Kruse, F., & Degner, J. (2021). Spontaneous inferences from behavior: A systematic meta-analysis. Manuscript under review.

Brewer, M. B. (1988). A dual process model of impression formation. In T. Srull & R. Wyer (Eds.), *Advances in social cognition* (Vol. 1, pp. 1–36). Hillsdale, NJ: Erlbaum

Carlston, D. E., & Skowronski, J. J. (1994). Savings in the relearning of trait information as evidence for spontaneous inference generation. *Journal of Personality and Social Psychology*, *66*(5), 840–856. 10.1037/0022-3514.66.5.840

Dotsch, R., Wigboldus, D. H., Langner, O., & van Knippenberg, A. (2008). Ethnic outgroup faces are biased in the prejudiced mind. *Psychological Science*, *19*, 978–980.

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology*, *87*, 327–339.

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in Experimental Social Psychology*, *23*, 1–74. 10.1016/S0065-2601(08)60317-2

Foulk, T., Woolum, A., & Erez, A. (2016). Catching rudeness is like catching a cold: The contagion effects of low-intensity negative behaviors. *Journal of Applied Psychology*, *101*(1), 50–67.

Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review*, *118*, 247–279.

Freeman, J. B., Stolier, R. M., & Brooks, J. A. (2020). Dynamic interactive theory as a domain-general account of social perception. In B. Malle (Ed.), *Advances in experimental social psychology* (Vol. 61, pp. 237–287). Amsterdam, The Netherlands: Elsevier.

Freeman, J. B., Stolier, R. M., Brooks, J. A., & Stillerman, B. S. (2018). The neural representational geometry of social perception. *Current Opinion in Psychology*, *24*, 83–91. 10.1016/j.copsyc.2018.10.003

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the Theory Theory. *Psychological Bulletin*, *138*(6) 1085–1108. 10.1037/a0028044

Hackel, L. M., & Amodio, D. M. (2018). Computational neuroscience approaches to social cognition. *Current Opinion in Psychology*, *24*, 92–97. 10.1016/j.copsyc.2018.09.001

Hackel, L. M., Doll, B. B., & Amodio, D. M. (2015). Instrumental learning of traits versus rewards: dissociable neural correlates and effects on choice. *Nature Neuroscience*, *18*(9), 1233–1243. 10.1038/nn.4080

Hackel, L. M., Mende-Siedlecki, P., & Amodio, D. M. (2020), Reinforcement learning in social interaction: The distinguishing role of trait inference. *Journal of Experimental Social Psychology*, *88*, 1–8. 10.1016/j.jesp.2019.103948

Hackel, L. M., Mende-Siedlecki, P., & Amodio, D. M. (2022). Context-dependent learning in social interaction: Trait impressions support flexible social choices. *Journal of Personality and Social Psychology*. 10.1037/pspa0000296

Ham, J. & van den Bos, K. (2008). Not fair for me! The influence of personal relevance on social justice inferences. *Journal of Experimental Social Psychology*, *44*, 699–705.

Hamilton, D. L., Chen, J. M., Ko, D., Winczewski, L., Banerji, I., & Thurston, J. A. (2015). Sowing the seeds of stereotypes: Spontaneous inferences about groups. *Journal of Personality and Social Psychology*, *109*, 569–588.

Hammond, K. R., Wilkins, M. M., & Todd, F. J. (1966). A research paradigm for the study of interpersonal learning. *Psychological Bulletin*, *65*(4), 221–232. 10.1037/h0023103

Hastie, R., Ostrom, T. M., Ebbesen, E. B., Wyer, R. S. Jr., Hamilton, D. L., & Carlston, D. E. (Eds.). (1980). *Person memory: The cognitive basis of social perception*. Hillsdale, NJ: Erlbaum.

Katz, D., & Braly, K. (1933). Racial stereotypes in one hundred college students. *Journal of Abnormal and Social Psychology*, *28*, 280–290.

Kressel, L., & Uleman, J. S. (2010). Personality traits function as causal concepts. *Journal of Experimental Social Psychology*, *46*, 213–216.

Kressel, L. M., & Uleman, J. S. (2015). The causality implicit in traits. *Journal of Experimental Social Psychology*, *57*, 51–54. 10.1016/j.jesp.2014.11.005

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis – Connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*(4), 1–28. 10.3389/neuro.06.004.2008

Kunda, Z., & Sherman-Williams, B. (1993). Stereotypes and the construal of individuating information. *Personality and Social Psychology Bulletin*, *19*(1), 90–99. 10.1177/0146167293191010

Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychological Review*, *103*(2), 284–308.

Kunda, Z., Sinclair, L., & Griffin, D. (1997). Equal ratings but separate meanings: Stereotypes and the construal of traits. *Journal of Personality and Social Psychology*, *72*(4), 720–734. 10.1037/0022-3514.72.4.720

Levey, S., Howells, K., & Levey, S. (1995). Dangerousness, unpredictability and the fear of people with schizophrenia, *Journal of Forensic Psychiatry*, 6(1), 19–39, 10.1080/09585189508409874

Lippmann, W. (1922). Stereotypes. *Public opinion* (pp. 79–94). New York, NY: Harcourt, Brace & Co. 10.1037/14847-006

Marcus, G. F., & Davis, E. (2013). How robust are probabilistic models of higher-level cognition? *Psychological Science*, *24*(12), 2351–2360.

Marcus, G. F., & Davis, E. (2015). Still searching for principles: A response to Goodman et al. (2015). *Psychological Science*, *26*(4), 542–544.

McAssey, M. P., Helm, J., Hsieh, F., Sbarra, D. A., & Ferrer, E. (2012). Methodological advances for detecting physiological synchrony during dyadic interactions. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 9(2), 41–53. 10.1027/1614-2241/A000053

McCarthy, R. J., & Skowronski, J. J. (2011). The interplay of controlled and automatic processing in the expression of spontaneously inferred traits: A PDP analysis. *Journal of Personality and Social Psychology*, *100*(2), 229–240. 10.1037/a0021991

McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *12*(1), 82–91.

Meshar, M. B., Stolier, R. M., & Freeman, J. B. (in press). Facial stereotyping drives judgments of perceptually ambiguous social groups. *Social Psychological and Personality Science*.

Monroe, B. M., Koenig, B. L., Wan, K. S., Laine, T., Gupta, S., & Ortony, A. (2018). Re-examining dominance of categories in impression formation: A test of dual process models. *Journal of Personality and Social Psychology*, *115*(1), 1–30. 10.1037/pspa0000119

Na, J., & Kitayama, S. (2011). Spontaneous trait inference is culture-specific: Behavioral and neural evidence. *Psychological Science*, *22*(8), 1025–1032. 10.1177/0956797611414727

Orghian, D., Garcia-Marques, L., Uleman, J. S., & Heinke, D. (2015). A connectionist model of spontaneous trait inference and spontaneous trait transference: Do they have the same underlying processes? *Social Cognition*, *33*, 20–66.

Orghian, D., Smith, A., Garcia-Marques, L., & Heinke, D. (2017). Capturing spontaneous trait inference with the modified free association paradigm. *Journal of Experimental Social Psychology*, *73*, 243–258. 10.1016/j.jesp.2017.07.004

Popel, H., Wang, Y., & Olson, I. R. (2019). A guide to representational similarity analysis for social neuroscience. *Social Cognitive and Affective Neurosciences*, 1243–1253. 10:1053/scan/nsz099

Rehder, B. (2007). Essentialism as a generative theory of classification. In A. Gopnik & L. Schultz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 190–207). Oxford, UK: Oxford University Press.

Rehder, B. (2017a). Concepts as causal models: Classification. In M. R. Waldmann (Ed.), *The Oxford handbook of causal reasoning* (pp. 346–375). New York: Oxford University Press.

Rehder, B. (2017b). Concepts as causal models: Induction. In M. R. Waldmann (Ed.), *The Oxford handbook of causal reasoning* (pp. 377–413). New York: Oxford University Press.

Rehder, B., & Burnett, R. C. (2005). Feature inference and the causal structure of categories. *Cognitive Psychology*, *50*, 264–314. 10.1016/j.cogpsych.2004.09.002

Rehder, B., & Kim, S. W. (2009). Classification as diagnostic reasoning. *Memory and Cognition*, *37*(6), 715–729. 10.3758/MC.37.6.715

Schneider, D. J. (2004). *The psychology of stereotyping*. New York: Guilford Press.

Semin, G. R. (2007). Grounding communication: Synchrony. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles* (2nd ed., pp. 630–649). New York, NY: The Guilford Press.

Semin, G. & Fiedler, K. (1988). The cognitive functions of linguistic categories in describing persons: Social cognition and language. *Journal of Personality and Social Psychology*, *54*(4), 558–568. 10.1037/0022-3514.54.4.558

Shanton, K. & Goldman, A. I. (2010). Simulation theory. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(4), 527–538.

Sherman, J. W., Kruschke, J. K., Sherman, S. J., Percy, E. J., Petrocelli, J. V., & Conrey, F. R. (2009). Attentional processes in stereotype formation: A common model for category accentuation and illusory correlation. *Journal of Personality and Social Psychology*, *96*, 305–323.

Shimizu, Y., Lee, H., & Uleman, J. S. (2017). Culture as automatic processes for making meaning: Spontaneous trait inferences. *Journal of Experimental Social Psychology*, *69*(1), 79–85. 10.1016/j.jesp.2016.08.003

Shimizu, Y., Senzaki, S., & Uleman, J. S. (2018). The influence of maternal socialization on infants' social evaluation in two cultures. *Infancy*, *25*(3), 748–766. 10.1111/infa.12240

Shimizu, Y., & Uleman, J. S. (2021). Attention allocation is a possible mediator of cultural variations in spontaneous trait and situation inferences: Eye-tracking evidence. *Journal of Experimental Social Psychology*, *94*, 1–11. 10.1016/j.jesp.2021.104115

Sidanius, J., & Pratto, F. (2011). Social dominance theory. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories in social psychology*, (Vol. 2, pp. 418–438). Los Angeles: Sage.

Smith, E. R., & Collins, E. C. (2009). Contextualizing person perception: Distributed social cognition. *Psychological Review*, *116*(2), 343–364. 0.1037/a0015072

Squire, L. R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory*, *82*, 171–177. 10.1016/j.nlm.2004.06.005

Stolier, R. M., Hehman, E., & Freeman, J. B. (2020). Trait knowledge forms a common structure across social cognition. *Nature Human Behavior*. 1–11. 10.1038/s41562/019-0800-6.38/s41562-

Takano, Y., & Osaka, E. (2018). Comparing Japan and the U.S. on individualism/collectivism: A follow-up review. *Asian Journal of Social Psychology*, *21*, 301–316. 10.1111/ajsp.12322

Tamir, D. I., Thornton, M. A., Contreras, J. M., & Mitchell, J. P. (2016). Neural evidence that three dimensions organize mental state representation: Rationality, social impact, and valence. *Proceedings of the National Academy of Sciences*, *113*(1), 194–199. 10.1073/pnas.1511905112

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*, 1279–1285. 10.1126/science.1192788

Todorov, A. (2017). *Face value: The irresistible influence of first impressions*. Princeton, NJ.: Princeton University Press.

Uleman, J. S. (2018). I/C: Individualism/ collectivism or individuate/ categorize? *Asian Journal of Social Psychology*, *21*, 317–323. 10.1111/ajsp.12330

Uleman, J. S. (2005). On the inherent ambiguity of traits and other mental concepts. In B. F. Malle & S. D. Hodges (Eds.), *Other minds: How humans bridge the divide between self and others* (pp. 253–267). New York: Guilford Publications.

Uleman, J. S., & Kressel, L. M. (2013). A brief history of theory and research on impression formation. In D. E. Carlston (Ed.), *Oxford handbook of social cognition* (pp. 53–73). New York: Oxford University Press.

Uleman, J. S., Rhee, E., Bardoliwalla, N., Semin, G., & Toyama, M. (2000). The relational self: Closeness to ingroups depends on who they are, culture, and the type of closeness. *Asian Journal of Social Psychology*, *3*, 1–17.

Winter, L., & Uleman, J. S. (1984). When are social judgments made? Evidence for the spontaneousness of trait inferences. *Journal of Personality and Social Psychology*, *47*(2), 237–252.

Index

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com