The Social Psychology of Perceiving Others Accurately

Edited by Judith A. Hall, Marianne Schmid Mast and Tessa V. West

The Social Psychology of Perceiving Others Accurately

We are constantly forming impressions about those around us. Social interaction depends on understanding others' behavior – assessing one another's personality, emotions, thoughts and feelings, attitudes, deceptiveness, group memberships, and other personal characteristics through facial expressions, appearance, body language, voice, and spoken language. But how accurate are impressions and when does such accuracy matter? How is accuracy achieved, and are some people more successful at achieving it than others? This comprehensive overview summarizes classic and cutting-edge research on this fast-expanding field and will be essential reading for anyone interested in the psychology of interpersonal perception. A wide range of experts in the field explore topics including age and gender effects, psychopathology, culture and ethnicity, workplaces and leadership, clinicians' skills, empathy, metaperception, and training people to be more accurate in their perceptions of others.

Judith A. Hall is a university distinguished professor of psychology at Northeastern University, Boston.

Marianne Schmid Mast is a professor of organizational behavior at the University of Lausanne, Switzerland.

Tessa V. West is an associate professor of psychology at New York University.

The Social Psychology of Perceiving Others Accurately

Edited by

Judith A. Hall
Northeastern University

Marianne Schmid Mast
University of Lausanne

Tessa V. West
New York University

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107101517

First published 2016

A catalogue record for this publication is available from the British Library

Library of Congress Cataloguing in Publication data
Hall, Judith A., editor. | Mast, Marianne Schmid, editor. | West, Tessa V., editor.
The social psychology of perceiving others accurately / edited by Judith A. Hall, Northeastern University, Marianne Schmid Mast, University of Lausanne, Tessa V. West, New York University.
Cambridge, United Kingdom : Cambridge University Press, 2016. | Includes index.
LCCN 2015042970 | ISBN 9781107101517 (hardback)
LCSH: Social perception. | Social psychology.
LCC HM1041 .S635 2016 | DDC 302–dc23
LC record available at http://lccn.loc.gov/2015042970

ISBN 978-1-107-10151-7 Hardback

To Nalini Ambady, who brought great intellect and joy to the study of interpersonal accuracy

Contents

Figures

Contributors

RAVIN ALAEI is a PhD candidate in the Department of Psychology at the University of Toronto. His research investigates how people effectively navigate their social world using minimal nonverbal cues, such as from the face and synchronous body movement.

SUSAN A. ANDRZEJEWSKI is an assistant professor of marketing in the Martin V. Smith School of Business & Economics at California State University Channel Islands. Susan's research lies at the intersection of consumer behavior and social psychology. The foundation of much of this work stems from the idea that psychological principles strongly influence what happens in the consumer marketplace. This work has been published in numerous marketing and psychology journals.

CHRIS ASHWIN is a lecturer (assistant professor) in the Department of Psychology at the University of Bath, UK. He originally studied psychology in Canada where he completed his BA (Hons) degree at the University of Windsor, and then carried out his doctoral training at the University of Cambridge investigating social–emotional processing in autism. He has previously held a post at the University of Essex carrying out research on emotional processing in anxiety disorders, before moving to Bath. His research interests involve social and emotional processes in general and how they are different in various disorders, including autism, psychosis, schizophrenia, and anxiety.

MITJA D. BACK is a professor of psychology at the University of Münster. He is interested in the interplay of personality and social relationships with a focus on the underlying behavioral and perceptual processes. His recent work includes research on the accuracy of interpersonal judgments, personality self-knowledge, the determinants of interpersonal attraction and friendship, individual differences in dating processes, the conceptualization and social consequences of narcissism, and the joint development of peer networks and personality.

TANJA BÄNZIGER studied psychology in Switzerland (in the universities of Lausanne and Geneva). She defended a PhD on the nonverbal communication of emotion (supervised by Professor Klaus R. Scherer) in Geneva, in 2004. During her postdoctoral studies, she was involved in the development of tests to measure emotion recognition accuracy and worked on several other methodological issues related to the study of emotional expressions and interpersonal communication. She worked also on the development of a multimodal database of emotional expressions: the Geneva Multimodal Emotion Portrayals. After moving to Sweden with her family, she worked as a researcher at the University of Gävle and Uppsala and is today employed as a lecturer at Mid Sweden University. Her main research interests are still related to interpersonal perception and nonverbal communication of emotion.

MAXWELL BARRANTI received his master of arts from Wake Forest University in 2014 and is currently a doctoral student at the University of Toronto. His research is focused on understanding the ways in which people see themselves and their social world and has been presented at numerous national conferences.

DANIELLE BLANCH-HARTIGAN is a social scientist with interdisciplinary research interests in psychology and public health. Her research, broadly defined, aims to understand how psychological constructs influence the powerful relationship between healthcare providers and patients to foster patient-centered care and to improve the patient experience. She received her PhD in psychology from Northeastern University, where her dissertation focused on assessing and training healthcare providers' ability to accurately recognize patients' verbal and nonverbal emotion cues. Danielle then completed a Masters in Public Health with a concentration in Quantitative Methods from Harvard School of Public Health and postdoctoral training as a Cancer Prevention Fellow at the National Cancer Institute in Bethesda, Maryland. Danielle is currently an assistant professor of health studies at Bentley University in Waltham, MA.

R. THOMAS BOONE received his PhD in social developmental psychology from Brandeis University in 1996, and also holds joint MA in psychology and women's studies. Dr Boone is currently an associate professor at the University of Massachusetts, Dartmouth, and chair of the Psychology Department. His primary research interests include the development of nonverbal emotional communication skills and their behavioral consequences. Some of his work in this area has demonstrated that children as young as four and five years of age can successfully encode and decode emotional meaning in expressive body movement. Another related research area focuses on the development of cooperative systems. In doing this research, he developed a new variant of the prisoner's dilemma game that allows the independent manipulation of both resource and relational dependency. He has also expanded this research to include the role of the affective signaling of trustworthiness and the role of emotional expressivity and Machiavellianism in the development of cooperation.

JUDEE K. BURGOON is Director of Research for the Center for the Management of Information and Site Director, National Science Foundation Center for Identification Technology Research at the University of Arizona, where she holds titles as professor of communication, family studies and human development. She has authored or edited 15 books and monographs (*Nonverbal Communication, Interpersonal Adaptation, Small Group Communication, Mexican Americans and the Mass Media*) and over 300 published articles, chapters, and reviews. Her current research on deception, dyadic interaction, and technologies for automated analysis of nonverbal and verbal communication has been funded by the National Science Foundation, Gannett Foundation, Department of Defense, and Department of Homeland Security, among others. She is the recipient of the highest honors given by the International Communication Association and National Communication Association and has held several offices in NCA.

ERIKA N. CARLSON holds a BS in psychology from the University of Florida, an MA in experimental psychology from Wake Forest University, and a PhD in psychology from Washington University in St. Louis. She is currently an assistant professor at the University of Toronto. Her research focuses on self-knowledge and interpersonal perception. Broadly speaking, her work examines whether people's beliefs about themselves and of their social world are accurate. For example, what do people know about their own personality that others do not know, and which aspects of personality can other people see that the self cannot see? Do people understand how other people experience them and if not, why? Is self-knowledge and an understanding of how others experience the self adaptive, or is it best to be blind? If self-knowledge is adaptive, how can people learn more about what they are like? Her work has been published in broad psychology outlets, such as *Current Directions in Psychological Science* and the *Journal of Social and Personality Psychology* and has been featured in popular media outlets such as the *Huffington Post*.

NORAH E. DUNBAR is a professor of communication at the University of California Santa Barbara. She teaches courses in nonverbal and interpersonal communication, communication theory, and deception detection. She was the principal investigator of a $5.4 million contract from the Intelligence Advanced Research Projects Activity and has had her research funded by the National Science Foundation, the Central Intelligence Agency, and the NSF Center for Identification Technology Research. She has published over 40 peer-reviewed journal articles and book chapters and presented over 60 papers at national and international conferences. Her research has appeared in *Communication Research*, *Communication Monographs*, and *Journal of Computer-Mediated Communication* as well as interdisciplinary journals such as *Journal of Management Information Systems* and *Computers in Human Behavior*. She has served on the editorial board of six disciplinary journals and is Chair of the Nonverbal Division of the National Communication Association.

HILLARY ANGER ELFENBEIN is a business school professor at the Olin School of Washington University in St. Louis. She holds a PhD in organizational behavior, a master's degree in statistics, and undergraduate degrees in physics and Sanskrit, all from Harvard University. Before joining Washington University, Dr Elfenbein served for five years on the faculty at the University of California, two years as a senior researcher at the Harvard Business School, and two years as a management consultant with the Monitor Group. Her research focuses on emotion in the workplace, including emotional intelligence, interpersonal relationships, negotiation, and the role of personality. Her work has appeared in journals such as the *Academy of Management Annals, Academy of Management Journal, Current Directions in Psychological Science, Journal of Applied Psychology, Journal of Personality and Social Psychology, Psychological Bulletin*, and *Psychological Science*, and she has served as an associate editor at *Management Science*.

PIP GRIFFITHS studied applied psychology at Cardiff University, Wales, where he gained an interest in autism during a year as an assistant psychologist in a special educational needs school. He specialized in autism while earning his PhD at the University of Bath, in particular studying the cognitive mechanisms that underlie facial emotion processing in autism. His research aims to understand how people with autism differ in the way they process emotional expressions.

SARAH D. GUNNERY is a postdoctoral scholar in the Department of Occupational Therapy at Tufts University. She earned her PhD in psychology from Northeastern University with a focus on expression, perception, and social outcomes of smiling. Her research interests include measuring facial expression in people with nonverbal communication deficits due to health conditions, use of deliberate facial expression to compensate for a loss of spontaneous expression, and gender differences in the expression and perception of facial expressions.

JUDITH A. HALL is a university distinguished professor of psychology at Northeastern University in Boston, MA. She received her PhD in social psychology from Harvard University. She has dual, and overlapping, interests in nonverbal communication processes, especially interpersonal accuracy, and in quality of medical care with a focus on physician–patient communication. Themes in her research have included gender and social power in relation to nonverbal communication. She has been editor-in-chief of *Patient Education and Counseling* and the *Journal of Nonverbal Behavior*, and is currently an associate editor at that journal. Dr Hall has been an author or editor of several books on nonverbal behavior, interpersonal accuracy, and physician–patient communication.

KRISTA M. HILL is an interdisciplinary researcher with interests in psychology and marketing. Her research background is in service recovery. Specifically, she is interested in the factors that render apologies and service recovery strategies more effective. She earned her PhD in psychology from Northeastern University and has received training in the Marketing Departments at Babson College and the D'Amore McKim School of Business at Northeastern University. Krista is currently an assistant professor of marketing at Bridgewater State University in Massachusetts.

SARA D. HODGES is a professor of psychology at the University of Oregon, where she currently also serves as associate dean of the graduate school. She studies connections between the self and other people, with specific research interests in empathic accuracy (how accurate people are at inferring others' thoughts), comparisons between the self and others, and the overlap in self perceptions and perceptions of others. She has previously served as an associate editor for the *Personality and Social Psychology Bulletin* and for *Social Cognition*. She was the 2013 winner of the Thomas Herman Award for Distinguished Teaching at the University of Oregon.

TERRENCE G. HORGAN is an associate professor of psychology at the University of Michigan-Flint. He is a coauthor of *Nonverbal*

Communication in Human Interaction. His research focuses on appearance accuracy, behavioral mimicry, sexual objectification, and adaptive memory. He teaches statistics, research methods, social psychology, and nonverbal communication to undergraduates at UM-Flint.

WILLIAM ICKES is a distinguished professor of psychology at the University of Texas at Arlington. He is the editor of *Empathic Accuracy* (1997) and the coeditor (with Jean Decety) of the *Social Neuroscience of Empathy* (2009). He is the author of *Everyday Mind Reading: Understanding What Other People Think and Feel* (2003) and *Strangers in a Strange Lab: How Personality Shapes Our Initial Encounters with Others* (2009). His research on empathic accuracy has been recognized by three international research awards.

DEREK M. ISAACOWITZ is a professor of psychology and director of the Lifespan Emotional Development Lab (LEDlab) at Northeastern University. He was an undergraduate at Stanford and earned his PhD in psychology from the University of Pennsylvania in 2001. His research on attention – emotional links across adulthood – is funded by the National Institute on Aging, and has appeared in journals such as *Psychological Science*, *Psychology and Aging*, and *Emotion*. He is currently an associate editor at *Psychology and Aging* and was previously an associate editor of *Emotion*. He has received APA Division 20's Springer Early Career Achievement Award, and the Baltes Early Career Award from GSA.

IOANA LATU received her PhD in social-cognitive psychology from Georgia State University in 2010. She was a Swiss National Science Foundation Ambizione fellow at the University of Neuchâtel in Switzerland. Since 2014 she is an assistant professor of psychology at Rutgers University Camden. She conducts research at the intersection of social psychology and work and organizational psychology, by investigating how people's stereotypes influence actual behavior in workplace interactions.

ELIZABETH A. LUCKMAN is a doctoral candidate at the Olin Business School of Washington University in St. Louis, specializing in Organizational Behavior. Her research seeks to cultivate a better understanding of how cultural values, ideologies, and assumptions shape the work experience. Prior to entering academia, she worked as a merchant and manager in luxury retailing. She holds an MBA from the Fisher College of Business at the Ohio State University and a BA from Wellesley College.

NORA A. MURPHY is an associate professor of psychology at Loyola Marymount University in Los Angeles, CA. Her research interests include nonverbal behavior, first impressions, person perception accuracy, emotions, and meta-analysis. Much of her research involves the use of "thin slices," which are brief video clips of social interactions, to investigate how emotions, personality, and nonverbal behavior shape social outcomes. She is currently an associate editor of the *Journal of Nonverbal Behavior* and associate chair of the Psychology Department at LMU. Dr Murphy received a BS in psychology from Trinity College, Hartford, CT, and MA and PhD degrees in psychology from Northeastern University, Boston, MA.

MATTHEW W. E. MURRY obtained a bachelor of science degree as a psychology major from West Virginia University. He is currently a doctoral student working in the Lifespan Emotional Development Laboratory at Northeastern University and is pursuing a PhD in psychology with a specialization in social and personality psychology. His research focuses on how the social environment influences the emotion perception abilities of younger and older adults, as well as how emotion perception is correlated with performance in other interpersonal perception tasks including those that measure everyday communication skills.

STEFFEN NESTLER is a senior lecturer in psychology at the University of Münster. He is interested in how individuals form judgments, why they are prone to judgmental errors (such as the hindsight bias), and how accuracy and confidence of judgments are related. Furthermore, he works in quantitative psychology on estimation of structural equation models, the social relations model, and social network models.

MOLLIE A. RUBEN is a postdoctoral fellow at the US Department of Veterans Affairs in the Center for Healthcare Organization and Implementation Research (CHOIR). She earned her PhD in psychology from Northeastern University with a focus on the nonverbal expression of physical pain and the perception and assessment of others' physical pain. Her research interests include doctor–patient communication, patient-centered care, and patient health outcomes.

NICHOLAS O. RULE is an associate professor and Canada Research Chair in Social Perception and Cognition at the University of Toronto. He investigates nonverbal behavior and person perception with a focus on the accuracy and predictive validity of social judgments. Much of his work considers the processes involved in the expression and perception of social groups that are perceptually ambiguous (e.g., sexual orientation, political affiliation, and religious ideology) and how individuals' life outcomes can be perceived from minimal cues in their appearance and behavior (e.g., leaders' selection and success). He approaches these topics from multiple levels of analysis, ranging from neural substrates to cross-cultural differences. He has received several professional awards for his research, including the Early Career Award from the International Social Cognition Network and Sage Young Scholars Award from the Foundation for Personality and Social Psychology.

KATJA SCHLEGEL received a PhD in psychology from the University of Geneva (Switzerland) in 2013. She is currently a postdoctoral researcher at the Social Interaction Laboratory at Northeastern University in Boston, MA. Dr Schlegel's work focuses on the assessment of individual differences in emotional skills and traits, including emotion recognition ability and emotional intelligence. She is also interested in how these characteristics manifest in verbal and nonverbal behavior and how they can be improved.

PETRA C. SCHMID is an assistant professor of organizational behavior at the Department of Management, Technology, and Economics at the Swiss Federal Institute of Technology (ETH Zurich). Her PhD research focused on the influence of moods and information processing styles on interpersonal accuracy. Most recently, she has studied how social power affects social judgments, self-regulation, and goal pursuit and examined the underlying psychological and neural mechanisms.

MARIANNE SCHMID MAST is a full professor of organizational behavior at the Faculty of Business and Economics in the Department of Organizational Behavior at the University of Lausanne in Switzerland. In her research, she investigates how individuals in power hierarchies interact, perceive, and communicate (verbally and nonverbally), how first impressions affect interpersonal interactions and evaluations, how people form accurate impressions of others, and how physician communication affects patient outcomes. The methods she uses include immersive virtual environment technology and computer-based automatic sensing and analyzing of nonverbal behavior. She is an associate editor of the *Journal of Nonverbal Behavior* and a member of the editorial board of *The Leadership Quarterly*.

ISHABEL M. VICARIA is a doctoral candidate at Northeastern University. Her research focuses on how the expression of affective states influences the ways in which people understand, communicate with, and get along with others of their own age and of different age groups. She also studies age-related changes in the recognition of emotional expressions via the face and body, as well as the nonverbal, interpersonal behaviors that arise during dyadic interactions. Isha first discovered a passion for psychological research as an undergraduate at Stetson University, which was further developed in a master's program at the University of Edinburgh.

TESSA V. WEST is an associate professor of psychology at New York University. Dr West utilizes multiple theoretical perspectives, coupled with progressive analytical and methodological approaches, to understand the psychological underpinnings of interpersonal perception as it plays out during dyadic and group interactions. Her work examines the experience of intergroup interaction from the perspective of all partners (e.g., whites and racial minorities). Dr West has developed theoretical and analytical models of person perception that are applicable to many subareas within psychology; she has published over 30 articles on interpersonal perception, many of which are on accuracy in interpersonal encounters.

ADRIENNE A. P. WISE received her master's degree in psychology from the University of Oregon, studying interpersonal perception and empathic accuracy. She received her undergraduate degree at Reed College, where she researched how various self-perceptions relate to women's persistence in STEM (science, technology, engineering, and math) fields. Additionally, she is interested in how perceptions of effort versus natural ability in an academic context influence ability and aptitude beliefs, vary based on the field of study, and depend on gender.

Part I

Domains of accurate interpersonal perception

1 Accurate interpersonal perception

Many traditions, one topic

Judith A. Hall, Marianne Schmid Mast, and Tessa V. West

Abstract

Research on people's accuracy in perceiving other people's states, traits, and social attributes has existed for over 100 years. In the past few decades, however, it has exploded into a vibrant, interdisciplinary, and international pursuit with relevance to all areas of social, interpersonal, and intrapersonal life. However, researchers typically work within narrowly defined traditions within the field. The present volume brings these areas together to describe method, theory, and findings for seven content domains (judging emotions, thoughts and feelings, truth versus lie, personality, social attributes, others' views of self, and group attitudes). Correlates at the group, individual, and situational levels are discussed, as well as the basic question: how accurate are people in judging other people? The strengths, weaknesses, and gaps in this field are discussed, and directions for future research are offered.

One of the most ubiquitous activities in daily life – a compulsion, even – is to *figure out* the people one knows, meets, or simply passes on the street. Every day, a person makes countless inferences about others' states and traits, background, attitudes – in fact any characteristic they may have. People might wonder who the leader is of a group they observe as an outsider, whether the person they just met at a party is involved with the person standing right next to them, whether they believe their teenager's claims about not drinking alcohol at a party, or whether their new collaborator has the motivation to see a project through. Think of how many strangers, coworkers, friends, and family members one interacts with in a typical day. Add to this the people seen or heard in media – in movies, television, interviews, news programs, social networking sites, or advertisements in video or print. Every time, the person viewing or listening is drawing inferences about those people. Regardless of whether perceivers are aware of drawing inferences, or are even aware of noticing those people, perceivers are still constantly processing information about those people's physical characteristics, their clothing and adornments, how they talk, what they say, and a myriad of nonverbal cues conveyed by their faces, postures, movement style, gazing patterns, voice, and even how they smell if proximity permits.

There are few things about a person that people do not pay attention to, consciously or nonconsciously, though some of these features or behaviors are more relevant to some judgment goals than others, and different perceivers may pay attention to different things. But, paying attention to each other and trying to figure out others is an irresistible inclination, and for a good reason: how could complex social life exist if people did not engage in these activities?

Sometimes, noticing is an end in itself. It is better to have noticed a friend's size before heading for the clothing store to buy that friend a sweater. Much of the time, however, noticing things about others – their appearance, behavior, attitudes, preferences, or whatever – leads to judgments and inferences. What are they feeling? Where do they come from? How old are they? Are they sexually available? Are they conscientious, intelligent, good-natured? Do they hate members of my social group?

Of course, the study of person perception and impression formation is well developed, as are many other research traditions relating to how people think about, and draw inferences about, each other (for example, correspondent inference theory; Jones & Harris, 1967). The present volume concerns a very specific aspect of person perception that has not previously been discussed in a unified way: *interpersonal accuracy*. The study of interpersonal accuracy is about whether a social perception or inference about another person (or persons) is correct. The authors of the chapters ask whether people are *accurate* in the conclusions they draw, either in general or in terms of individual, group, or situational factors, and what the correlates and processes of accuracy are.

The questions that researchers have asked about accuracy are numerous and many studies have been done. Thus, there is a rich literature. Yet, even though this literature could be – should be – integrated as a coordinated, comprehensive field, this has not happened yet. Researchers pursuing the many different strands of accuracy research have proceeded largely in isolation from each other. Researchers tend to be interested in just one kind of accuracy, for example identifying emotions from facial expressions, and often they have a preferred method of measurement. Sometimes the traditions and habits that grow up are not well rooted in theoretical considerations: for example, in the study of how personal dominance or power correlates with interpersonal accuracy, virtually the entire literature is based on accuracy in detecting emotion – yet rarely does any researcher give a reason why detecting emotion is more relevant to the perceiver's dominance or social power than other kinds of accuracy. In our opinion there is so much fragmentation that most researchers do not think of themselves as belonging to a general field of "interpersonal accuracy research"; rather, they study accuracy in a specific social context and operationalize accuracy in ways that are specific to their questions, and sometimes they choose their measurements out of convenience or in ignorance of what instruments and approaches are available, or theoretically justified. Furthermore, accuracy researchers in different fields or subfields may not communicate with each other or even be aware of each other's work.

The goal of the present book is to summarize numerous large and diverse research traditions, done by many different kinds of researchers and for a wide variety of theoretical and practical purposes. We hope the book serves the research community and any reader who wishes to learn more about interpersonal accuracy. In the remainder of this chapter, we provide background and framework for the rest of the volume. Of course, individual chapters provide much more detail on some of the issues we talk about in general terms here. And, even though the book covers a lot of ground, not all strands of research are represented and certainly not all of the (often fascinating) accuracy topics that have been taken up over the years can be described.

What is accuracy and how is it measured?

For a general definition, we think of interpersonal accuracy as accurate judgment about any verifiable characteristic of a person or about the group that a person belongs to. Mostly in the present volume, this accuracy is based upon people (called perceivers, judges, or decoders) witnessing the behavior and/or appearance of other people (also called encoders or targets) and either making an *inference* based on the behavior and/or appearance (the most common task) or being asked to *remember* aspects of the behavior or appearance.[1] The term "witnessing" means that the perceiver has direct exposure to the target person through some medium, which could be live (physically present, on the telephone, or seen via an electronic interface) or not live (recorded as on videotape, audiotape, photographs, or in a written transcript of the target person's words). In one chapter (on accuracy of knowing others' attitudes), the

[1] Although clearly a kind of accuracy, identity recognition and eyewitness research (being able to say whether a face, voice, or whole person has been seen or heard at an earlier time) is not included in the present volume, except in Chapter 11.

accuracy concept is extended to include success in judgments about whole groups of people (e.g., women).

To measure accuracy, a defensible criterion for determining what is a correct versus incorrect response must be established (Ickes, 1997; West & Kenny, 2011). One common criterion is a state or message the target people were instructed to display (for example, a particular emotion), or a kind of situation they are imagining themselves to be in (for example, acting out talking to a lost child or asking someone for forgiveness); this kind of criterion is often used with tests of judging emotion or affect. Another approach is to gather factual information about the target people. In personality judgment, this is usually the targets' self-ratings on personality scales. As other examples of using documentable information, the criterion for judging the winner of a competition could be the researcher's knowledge of who was the actual winner, and the criterion for judging intelligence could be some kind of cognitive test that the target person has taken. Similarly, the criterion for detecting deception would be the researcher's knowledge of whether the target person was lying or not. Sometimes the criterion is the circumstances that occur at the moment the recordings are made, as in the slide-viewing paradigm of Buck (1979) where the target people's faces are recorded while they watch emotionally evocative photographs or videos. Another commonly used criterion is the target person's retrospective report of what they were thinking or feeling at a particular moment during an earlier interaction that they are watching now in replay (Ickes, 1997). Sometimes the criterion is simply the consensus of a group of observers; if most of them say, for example, that the person appears to be showing pride, then "pride" is declared to be the correct answer. Consensus can be a debatable criterion (Kruglanski, 1989).

Obviously having a defensible criterion is important, and researchers often go to considerable lengths to obtain convergent information to bring the criterion as close to the "truth" as possible (such as personality ratings from friends or family of the targets, not just from the targets themselves; Funder, 1995). In every case, "accuracy" is limited to what is operationally defined by the researcher when establishing the criterion. It is important for researchers to keep in mind that accuracy is an abstract construct that is always, and necessarily, instantiated in an operational definition. Almost every operationally defined criterion has limitations, and it is desirable for researchers to develop measurements that employ different operational definitions. For example, you would like to know whether results from a test that measures emotion recognition accuracy are the same if the criterion is the emotion a target intended to convey versus the emotion a group of viewers consensually says it conveys.

Sometimes one hears statements such as "You aren't measuring accuracy – you are just measuring self–other agreement." Whoever might say this is forgetting that accuracy can only be glimpsed through the lens of operationally defined criteria, of which self–other agreement is one (for example, whether perceivers "see" the same personality traits in the targets as the targets claim to have). Therefore, if the researcher has defined accuracy operationally as self–other agreement, then that researcher is entitled to claim that their test measures accuracy, keeping in mind that it is only one possible operational definition of the concept. It is good if researchers debate the merits of different criteria and then compare them.

The domain of accurate judgment can be anything about a person for which a defensible criterion can be provided, yet there is plenty of room for debate. Sometimes entwined with the choice of criterion is the question of how the stimuli are produced in the first place. The field of emotion recognition has a prevailing paradigm for both – the criterion is the target's intentions, while the production method is deliberate posing – but there is room for debate about intentions per se as a criterion and about posing as a good source of emotional expressions. In yet other domains, there might be even more doubt about what the criterion should be. An example would be pain judgment: should the criterion be the target's self-reported pain, or physiological measurements, or facial displays? Any of these could have significant shortcomings or strengths. One way to overcome the limitations of any one criterion assessment method is to combine different methods. For instance, to determine the criterion for the judgment of facial expressions of emotion, the poser's intention could be combined with a more objective coding of the activated facial muscles (FACS coding; Cohn, Ambadar, & Ekman, 2007) and eventually even with consensus judgments; or as is sometimes done, personality reports from multiple respondents could be combined in the criterion. Finally, researchers might wonder if any measurable criterion can be found – as with judging very transitory mental states during ongoing interaction – because the researcher either cannot "get inside" the heads of target people or cannot do so in a timely way. Researchers have limits on the resources they can expend in getting to the "truth" about people in order to establish acceptable criteria for the kinds of accuracy they wish to measure, and this is one reason why they often fall back on low-cost methods such as instructing target people to pose various emotions.

Terminology

We believe the field, regardless of the details of criterion and measurement, needs a general term, and the one we advocate is *interpersonal* *accuracy*. There is value in having a general term, not only for building a sense of commonality among researchers, but also for the very practical reason that conducting online searches for accuracy studies is extremely difficult when there is no common terminology. Specific terms that are appropriate to a specific accuracy concept or type of measurement are, of course, still useful within the general concept of interpersonal accuracy.

Here we list terms that are used commonly and why we think they are not suitable as a general term.

Interpersonal sensitivity. Though often used to describe interpersonal accuracy (e.g., Hall & Bernieri, 2001), this term is ambiguous. As noted by Bernieri (2001), this term could encompass both accurate perception of others as well as wise, tactful, or otherwise appropriate behavior toward them, as in the observation that someone responded very "sensitively" to her friend's distress. For that reason, we do not advocate this as a synonym for interpersonal accuracy.

Nonverbal sensitivity. This term has the same ambiguity as the preceding one and is, moreover, descriptive only of responses to nonverbal cues (e.g., facial expressions, gestures, postures, voice quality) (Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979). However, being accurate is often based on interpretation of linguistic as well as nonverbal cues. Therefore this is not a useful general term.

Emotion recognition, or emotion recognition ability. This is the most widely used term because emotion recognition is the most commonly studied kind of accuracy (mostly involving photographs of posed facial expressions). Because it focuses solely on emotions, this is not a useful general term.

Decoding ability. This term has two shortcomings as a possible general term. One is that it is too broad; a bibliographic search for "decoding ability" or "decoding accuracy" results in countless references to unrelated topics (how people read, etc.). It is also too narrow, in that it implies only accuracy defined as inference, whereas our definition of accuracy also includes the noticing/recalling process described earlier.

Inferential accuracy (versus recall accuracy). These terms have been used to distinguish between the two kinds of accuracy mentioned earlier: drawing an inference (making a judgment) about a person's states, traits, or other characteristics and noticing/recalling something about a person (Hall, Carter, & Horgan, 2001). Though in a given context they are useful terms, they do not convey the "interpersonal" notion of one person being accurate about another person.

Empathic accuracy. This widely used term was created by Ickes (1997) to describe the method of asking perceivers to guess what target people were thinking and feeling at specific moments during an interaction. This method differs from many other established paradigms because it includes inferences about both affect and cognition, and it is based on spontaneous (not posed or rehearsed) target behavior. Unfortunately, many researchers use the term "empathic accuracy" as a general term for any kind of measured interpersonal accuracy, thus blurring the distinction between the specific methodology for which Ickes chose the term and a wide range of other measurement approaches. We urge researchers to use this term in its correct methodological context and not as a general term.

Mind reading. This term has been used by Ickes (2003) as a synonym for empathic accuracy. However, in the popular imagination the term "mind reading" generally implies psychic powers, which is not what Ickes was suggesting. It is therefore a potentially confusing term.

Mental states attribution. Used by Frith (1997) and others, this term is often used in conjunction with the "theory of mind" concept (Baron-Cohen et al., 1999) to refer to ability to make correct inferences about others' thoughts, knowledge, and intentions. This, like some of the other terms listed above, has relevance for a limited range of interpersonal accuracy tasks.

Accuracy at zero acquaintance, and first impression accuracy. These terms are suitable for judgments made of strangers or of people whom one has just met. These are not suitable as general terms because accuracy can be measured between people who are acquainted.

Judgmental accuracy. To our knowledge this term is used almost exclusively by researchers who study accurate personality judgment (Colvin & Bundick, 2001; Funder, 1995). The shortcoming we see with this term is that the term "judgmental" connotes judgmentalism, that is, being too quick to form moral judgments of others, which is not its intended meaning. Thus the term is not transparent.

As we have said, for different purposes, each of these terms can be appropriate. Our point is that a general term that can subsume all of these is also desirable.

Burgeoning of the accuracy field

The study of interpersonal accuracy is extremely active. A search for "emotion recognition" on PsycINFO found an astonishing trend in entries over the past decades (Table 1.1). Despite likely undercounting in the earlier decades because the exact term "emotion recognition" was not used as consistently as it is now, the explosion of recent research is still amazing, especially considering only half of the current decade has past.

Table 1.1 *Number of citations to "emotion recognition" on PsycINFO, 1950– 2015*

1950s	1
1960s	4
1970s	6
1980s	26
1990s	89
2000s	681
2010–15	1,964

The appearance of meta-analyses within a field is testament to the field's maturation. We located over 50 published meta-analyses on interpersonal accuracy, which are listed at the end of this chapter. Undoubtedly, there are meta-analyses that we did not locate, but even without these it is obvious that there is a great deal of published research on accuracy.

History

The once-popular field of accuracy in personality judgment was derailed for decades in part because of stringent critiques of the measurement methods then used (Cronbach, 1955; Gage & Cronbach, 1955; Funder, 2001). The critiques pointed out that accuracy scores necessarily needed to be decomposed into different components (termed a "componential approach"). Cronbach originally proposed decomposing scores at the level of the perceiver, across targets and judgments (Cronbach, 1955; for a review of modern componential approaches see Kenny, West, Malloy, & Albright, 2006). Not until the 1990s did personality researchers adopt different methods, based on correlations across items or across targets (see Kenny et al., 2006, for a review).

Researchers in adjacent fields, however, continued studying accuracy. Most work was focused on judgments of affect and emotion. Most of this research has used methods that, fortunately, allow researchers to understand better why perceivers were accurate (e.g., they used multiple targets, and often multiple emotions or affective states expressed by the targets, which allows one to test whether perceivers were accurate in reading particular targets or targets in general), and they utilized multiple methods (e.g., multiple choice, rating scales, self-report recall) to assess cross-methodological consistency. Ekman and Friesen (1971) and Izard (1971) were highly invested in research on the correct identification of emotions from facial expressions, asking group-comparison questions such as whether cultures differ in their accuracy. Actually, the path backward for emotion recognition research is long, with studies emerging early in the twentieth century (Adams, 1927; Feleky, 1914; Langfeld, 1918; Ruckmick, 1921) and a review appearing by mid-century (Taft, 1955).

Robert Rosenthal can be credited with starting the tradition of measuring accuracy via standardized, validated instruments – a tradition that is strong to this day though better represented for some content domains (e.g., interpretation of affective cues) than others. His instrument, the Profile of Nonverbal Sensitivity (PONS test; Rosenthal et al., 1979), measures the ability to infer affective states within situational context, based on face, body, and nonverbal vocal cues. Developed originally to explore individual differences in the receptivity to interpersonal expectancy effects (e.g., whether a pupil will pick up and be influenced by a teacher's cues signaling their beliefs about the pupil), the PONS became a staple for researchers measuring individual differences in accuracy. Many other tests have since been validated and adopted widely (e.g., Diagnostic Analysis of Nonverbal Accuracy or DANVA; Nowicki & Duke, 1994), and others continue to be developed (e.g., Geneva Emotion Recognition Test or GERT; Schlegel, Grandjean, & Scherer, 2014). In addition, many stimulus sets, mainly of facial expressions, have been used for measuring accuracy even though they were not specifically developed nor systematically validated as psychometric tests; the most prominent of these is the Pictures of Facial Affect (Ekman & Friesen, 1976). (For comparisons of these and other instruments, see Castro, Cheng, Halberstadt, & Grühn, 2015, and Hall, Bernieri, & Carney, 2005.)

Another major topic in accuracy is lie detection, pursued primarily by researchers in communication sciences. Research on this topic has been going on, though steadily rising, for many decades. As reviewed in the chapter by Burgoon and Dunbar, one major area of focus has been on training perceivers to become more accurate in judging whether targets are lying or speaking the truth. As the authors of that chapter conclude, however, lie detection training is quite difficult, as there are no universal indicators of truth and lie telling that perceivers can rely on; accurate lie detection is a complex interaction between perceiver and target traits, the relationship between the perceiver and target, the social context, and the modality of communication.

Finally, research on accuracy in judging characteristics of people and their social relationships started to blossom only since the 2000s, though in the late 1980s a standard test called the Interpersonal Perception Task (IPT; Costanzo & Archer, 1989) included some item content of that sort (e.g., whether two people were in a romantic relationship or not). Studies of accuracy in judging sexual orientation represent the most recently developed theme in terms of judging social characteristics (e.g., Rule & Ambady, 2008).

Methodologies for studying accuracy

An earlier edited volume covered some methodologies in detail (Hall & Bernieri, 2001), and individual chapters in this volume describe the methodologies used in specific research contexts. Here we offer an overview, first discussing general methodological distinctions and then discussing different ways of calculating accuracy scores.

General methodological distinctions. Most assessments of accuracy are concerned with a single domain to be judged, though there may be much variation within that domain (for example, different numbers and types of emotions in different emotion recognition tasks, or different types and circumstances of lying in different lie detection tasks). The most commonly studied domains are emotion and affect (though what counts under this heading is not agreed on), lie detection, and personality. Other domains are group memberships and social attributes (e.g., sexual orientation, religion/ethnicity), interpersonal relations (e.g., strangers or not), attitudes (e.g., racial attitudes), intelligence, and dominance/status, though this list is certainly not exhaustive. One well-known test, the IPT mentioned earlier, is unusual in encompassing five distinctly different domains of judgment (deception, competition, kinship, intimacy, and status).

There are two basic paradigms for measuring accuracy: the *testing* approach, in which perceivers view, hear, or read a standard set of stimuli and make judgments about them (thus enabling many perceivers to judge the same stimuli), and the *in vivo* approach, in which perceivers make judgments about others with whom they interact or at least have live contact (most commonly, this is done in dyads).

There are many methodological factors that can influence the degree to which perceivers are accurate and what mechanisms can be studied to understand how interpersonal accuracy is achieved. Stimuli can vary on a number of dimensions, such as cue modality (e.g., face, body, paraverbal cues, linguistic cues), whether expression was posed/rehearsed or the expression occurred in a relatively spontaneous manner, and whether the instrument for assessing accuracy has been validated by prior researchers or was developed for a particular study. Researchers may show representative stimuli (for example, all of the instances of lying and truth telling they gathered), or they may show stimuli selected through pretesting to have a desired degree of difficulty or other desired characteristics. For example, in a study of judging the sexual orientation based on a set of target faces, Stern and colleagues (Stern, West, Jost, & Rule, 2013) intentionally chose stimuli that varied in how masculine or feminine the target faces were on a continuum of masculinity. Roughly equal numbers of feminine gay and straight faces, and masculine gay and straight faces, were chosen.

There are many additional factors that are important to consider, such as the age, gender, and ethnicity of target persons, as well as the culture of the target persons and the culture of the individuals making the judgments. All of these factors could potentially influence perceivers' levels of accuracy and the information they utilize in making judgments.

Another consideration is to determine the most appropriate method for gathering perceivers' responses to the stimuli. There are a number of ways of collecting judgments (which will influence how accuracy scores can be calculated), including rating scales, binary decisions (e.g., truth/lie for lie detection), and multiple-choice options. In some cases, such as in the empathic accuracy paradigm (Ickes, 1997), free responses are recorded and then coded by researchers for how well they match the criteria.

Calculating accuracy. An earlier chapter on methodology (Hall et al., 2005) as well as the book mentioned earlier (Hall & Bernieri, 2001) included some discussion on scoring options and their implications, and individual chapters in the present volume give more detail. Here, we highlight some key conceptual issues when it comes to calculating accuracy scores.

As reviewed in a number of places in this chapter, several approaches have been utilized in calculating accuracy scores. One approach is to calculate accuracy scores for each person, either by creating an average across many judgments and the same average across those items for the truth criterion and correlating them, or calculating a mean difference score. For example, participants may report on how much they think certain people (e.g., women, political in-group members) agree with 23 different attitude statements (e.g., guns should be legal in the United States, women and men should have equal pay). Truth criterion data would also be obtained for these same statements. For the correlational approach, each participant receives an accuracy score (perhaps in the form of a Z score) that represents their overall accuracy; for the mean difference approach, each participant receives a mean difference score that represents the extent to which they thought individuals agreed more (or less) with those statements on average. These idiographic scores can then be used as predictors or outcome variables in other models.

Another approach is a nomothetic approach, in which accuracy is estimated not for individual perceivers but for a whole group of participants. For example, West and Kenny's Truth and Bias model estimates accuracy using a regression-based approach in which the judgment is regressed on the truth criterion, and accuracy (the *truth force*) conceptually represents how strongly the judgment is being pulled toward the truth, estimated as a regression coefficient. Benefits of this approach include examining multiple forms of accuracy in one model (e.g., directional bias and the truth force in the Truth and Bias model). When repeated measures data are collected, the random effects of accuracy can be estimated in a multilevel modeling framework, which allows one to examine whether there are within-person correlations between truth and bias (e.g., if I am biased, am I accurate?). With dyadic data, within-dyad correlations can be estimated, which allow one to examine questions such as, "If I am accurate is my partner accurate?"

Another important methodological consideration is how many targets each perceiver judges. When each perceiver judges many targets (and these targets differ across perceivers), a componential analysis can be used to decompose judgments into theoretically relevant sources of variance, such as perceiver, target, and relationship, for both the judgments and the truth criteria. By correlating judgment components with truth criteria components, one can estimate accuracy at different levels. For example, dyadic accuracy assesses how accurate perceivers are at judging particular targets (e.g., is Tom particularly accurate in his judgments of Bob), and generalized accuracy assesses how accurate perceivers are in general (e.g., is Tom accurate in his judgments of everyone in his group; see Kenny & Albright, 1987, for more details).

In some cases, perceivers judge the same set of targets, for example, when making ratings of a set of stimuli, such as 30 different faces. Variance due to target stimuli should be estimated in these models, as it allows scholars to determine if accuracy across a set of stimuli is due to one particular target stimulus. For example, in reading the sexual orientation of 10 target faces, it may be the case that one target face is particularly easy to read, and that target face is driving all of the accuracy effects. Estimating variance due to stimuli is an important methodological step when scholars are initially developing a set of stimuli to insure that particular targets in the stimuli are not driving accuracy effects in the data.

Another important methodological decision is to determine how bias is best operationalized. As many scholars have argued, bias does not necessarily imply error, or inaccuracy (Funder, 1995), and bias might actually contribute to accuracy (West & Kenny, 2011). There are many theoretical and conceptual models that distinguish bias from error, and provide guides of how to best conceptualize bias in a model. Some of these models, such as the Truth and Bias model, provide guidelines of how to estimate how much accuracy is achieved "indirectly" through bias by treating bias as a mediator of the effect of the judgment on the truth criterion. Some investigators use signal detection methods for separating sensitivity from bias.

How different themes and traditions are represented in this book

The first half of the book focuses on specific domains of accuracy, from very traditional ones such as judging emotions (Bänziger), judging truth and lie (Burgoon and Dunbar), and judging personality (Back and Nestler) to more recently pursued ones such as judging thoughts and feelings (Ickes), meta-perception accuracy (Carlson and Barranti), judging attitudes (West), and judging social attributes (Alaei and Rule). The second half focuses on different classes of correlates. Many themes appear in more than one chapter. In general, assessing a social interaction partner accurately is necessary for navigating many types of social relationships (as exemplified in the chapters by Hodges and Wise, and by Schmid Mast and Latu). As we said earlier, there is not much connection between the different research fields because researchers in one field typically stick to one type of interpersonal accuracy assessment, and there has been limited theorizing about how different types of accuracy are related. The chapters by Boone and Schlegel and by Murphy both represent broad views of the field in order to bridge these gaps.

Schmid asks how proximal (cognitive, emotional, situational) circumstances increase or decrease interpersonal accuracy. Other chapters take on various group and individual differences pertaining to gender (Hall, Gunnery, and Horgan), age (Isaacowitz, Vicaria, and Murry), psychopathology (Griffiths and Ashwin), clinicians (Ruben), prosociality (Hodges and Wise), and culture (Luckman and Elfenbein). The effects of short-term training programs to increase interpersonal accuracy are summarized by Blanch-Hartigan, Andrzejewski, and Hill.

Challenges for the study of interpersonal accuracy

In editing this volume, we identified many challenges and future directions for the study of interpersonal accuracy, many of which are discussed in detail in Murphy's concluding chapter. Murphy focuses on the importance of developing new methods to better understand the mechanisms of accuracy, and to work toward a more nuanced understanding of the question "how accurate is accurate?" We have discussed throughout the present chapter that scholars need to develop a cross-cutting theoretical framework of interpersonal accuracy, and one major challenge in doing so is developing conceptual and methodological approaches to interpersonal accuracy that can be universally applied. Such an approach would help scholars develop a thorough understanding of how interpersonal accuracy is achieved and what it in turn predicts. It would also allow scholars to broadly construe categories of outcome variables that fall under different theoretical umbrellas – for example, behaviors required for relationship maintenance, or behaviors that help perceivers achieve dominance. What particular kinds of interpersonal accuracy matter most for these different kinds of outcome? Our hope is that this edited volume is an initial step toward building an integrated approach to studying interpersonal accuracy.

References

Adams, H. F. (1927). The good judge of personality. *Journal of Abnormal Psychology*, 22, 172–181.

Baron-Cohen, S., Ring, H., Williams, S., Wheelwright, S., Bullmore, E., Brammer, M., & Andrew, C. (1999). Social intelligence: The role of the amygdala. *European Journal of Neuroscience*, 11, 1891–1898.

Bernieri, F. J. (2001). Toward a taxonomy of interpersonal sensitivity. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement* (pp. 3–20). Mahwah, NJ: Lawrence Erlbaum Associates.

Buck, R. (1979). Measuring individual differences in the nonverbal communication of affect: The slide viewing paradigm. *Human Communication Research*, 6, 47–57.

Castro, V. L., Cheng, Y., Halberstadt, A. G., & Grühn, D. (2015). EUReKA! A conceptual model of emotion understanding. *Emotion Review*.

Cohn, J. F., Ambadar, Z., & Ekman, P. (2007). Observer-based measurement of facial expression with the Facial Action Coding System. In J. A. Coan & J. J. B. Allen (Eds.), *Handbook of emotion elicitation and measurement* (pp. 203–221). New York: Oxford University Press.

Colvin, C. R., & Bundick, M. J. (2001). In search of the good judge of personality: Some methodological and theoretical concerns. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement* (pp. 47–65). Mahwah, NJ: Lawrence Erlbaum Associates.

Costanzo, M., & Archer, D. (1989). Interpreting the expressive behavior of others: The Interpersonal Perception Task. *Journal of Nonverbal Behavior*, 13, 225–245.

Cronbach, L. J. (1955). Processes affecting scores on “understanding others” and “assumed similarity.” *Psychological Bulletin*, 52, 177–193.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17, 124–129.

Ekman, P., & Friesen, W. V. (1976). *Pictures of facial affect*. Palo Alto: Consulting Psychologists Press.

Feleky, A. M. (1914). The expression of the emotions. *Psychological Review*, 21, 33–41.

Frith, C. D. (1997). Functional brain imaging and the neuropathology of schizophrenia. *Schizophrenia Bulletin*, 23, 525–527.

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, 102, 652–670.

Funder, D. C. (2001). Accuracy in personality judgment: Research and theory concerning an obvious question. In B. W. Roberts & R. Hogan (Eds.), *Personality psychology in the workplace: Decade of behavior* (pp. 121–140). Washington, DC: American Psychological Association.

Gage, N. L., & Cronbach, L. (1955). Conceptual and methodological problems in interpersonal perception. *Psychological Review*, 62, 411–422.

Hall, J. A., & Bernieri, F. J. (Eds.) (2001). *Interpersonal sensitivity: Theory and measurement*. Mahwah, NJ: Lawrence Erlbaum Associates.

Hall, J. A., Bernieri, F. J., & Carney, D. R. (2005). Nonverbal behavior and interpersonal sensitivity. In J. A. Harrigan, R. Rosenthal, & K. R. Scherer (Eds.), *The new handbook of methods in nonverbal behavior research* (pp. 237–281). Oxford: Oxford University Press.

Hall, J. A., Carter, J. D., & Horgan, T. G. (2001). Status roles and recall of nonverbal cues. *Journal of Nonverbal Behavior*, 25, 79–100.

Ickes, W. (1997). *Empathic accuracy*. New York: Guilford Press.

Ickes. W. (2003). *Everyday mind reading: Understanding what other people think and feel*. Amherst, NY: Prometheus Books.

Izard, C. E. (1971). *The face of emotion*. East Norwalk, CT: Appleton-Century-Crofts.

Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology*, 3, 1–24.

Kenny, D. A., & Albright, L. (1987). Accuracy in interpersonal perception: A social relations analysis. *Psychological Bulletin*, 102, 390–402.

Kenny, D. A., West, T. V., Malloy, T., & Albright, L. (2006). Componential analysis of interpersonal perception data. *Personality and Social Psychology Review*, 10, 282–294.

Kruglanski, A. W. (1989). The psychology of being "right": The problem of accuracy in social perception and cognition. *Psychological Bulletin*, 106, 395–409.

Langfeld, H. S. (1918). Judgments of facial expression and suggestion. *Psychological Review*, 25, 488–494.

Nowicki, S., & Duke, M. (1994). Individual differences in the nonverbal communication of affect: The Diagnostic Analysis of Nonverbal Accuracy scale. *Journal of Nonverbal Behavior*, 18, 9–34.

Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., & Archer, D. (1979). *Sensitivity to nonverbal communication: The PONS test*. Baltimore: Johns Hopkins University Press.

Ruckmick, C. A. (1921). A preliminary study of the emotions. *Psychological Monographs*, 30, 30–35.

Rule, N. O., & Ambady, N. (2008). Brief exposures: Male sexual orientation is accurately perceived at 50 ms. *Journal of Experimental Social Psychology*, 44, 1100–1105.

Schlegel, K., Grandjean, D., & Scherer, K. R. (2014). Introducing the Geneva Emotion Recognition Test: An example of Rasch-based test development. *Psychological Assessment*, 26, 666–672.

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.

Taft, R. (1955). The ability to judge people. *Psychological Bulletin*, 52, 1–23.

West, T. V., & Kenny, D. A. (2011). The truth and bias model of judgment. *Psychological Review*, 118, 357–378.

Meta-analyses on interpersonal accuracy

Aamodt, M. G., & Custer, H. (2006). Who can best catch a liar? A meta-analysis of individual differences in detecting deception. *The Forensic Examiner*, 15, 6–11.

Andrzejewski, S. A., Hall, J. A., & Salib, E. R. (2009). Anti-Semitism and identification of Jewish group membership from photographs. *Journal of Nonverbal Behavior*, 33, 47–58.

Babbage, D. R., Yim, J., Zupan, B., Neumann, D., Tomita, M. R., & Willer, B. (2011). Meta-analysis of facial affect recognition difficulties after traumatic brain injury. *Neuropsychology*, 25, 277–285.

Barkl, S. J., Lah, S., Harris, A. W., & Williams, L. M. (2014). Facial emotion identification in early-onset and first-episode psychosis: A systematic review with meta-analysis. *Schizophrenia Research*, 159, 62–69.

Blanch-Hartigan, D., Andrzejewski, S. A., & Hill, K. M. (2012). The effectiveness of training to improve person person accuracy: A meta-analysis. *Basic and Applied Social Psychology*, 34, 483–498.

Bond, C. F., Jr., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, 10, 214–234.

Bond, C. F., Jr., & DePaulo, B. M. (2008). Individual differences in judging deception: Accuracy and bias. *Psychological Bulletin*, 134, 477–492.

Chung, Y. S., Barch, D., & Strube, M. (2014). A meta-analysis of mentalizing impairments in adults with schizophrenia and autism spectrum disorder. *Schizophrenia Bulletin*, 40, 602–616.

Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin*, 136, 1092–1122.

Dalili, M. N., Penton-Voak, I. S., Harmer, C. J., & Munafo, M. R. (2014). Meta-analysis of emotion recognition deficits in major depressive disorder. *Psychological Medicine*, 45, 1135–1144.

Daros, A. R., Zakzanis, K. K., & Rector, N. A. (2014). A quantitative analysis of facial emotion recognition in obsessive-compulsive disorder. *Psychiatry Research*, 215, 514–552.

Davis, M. H., & Kraus, L. A. (1997). Personality and empathic accuracy. In W. J. Ickes (Ed.), *Empathic accuracy* (pp. 144–168). New York: Guilford Press.

Dawel, A., O'Kearney, R., McKone, E., & Palermo, R. (2012). Not just fear and sadness: Meta-analytic evidence of pervasive emotion recognition deficits for facial and vocal expressions in psychopathy. *Neuroscience and Biobehavioral Reviews*, 36, 2288–2304.

Demenescu, L. R., Kortekaas, R., den Boer, J. A., & Aleman, A. (2010). Impaired attribution of emotion to facial expressions in anxiety and major depression. *PLoS ONE*, 5, Article e15058.

Driskell, J. E. (2012). Effectiveness of deception detection training: A meta-analysis. *Psychology, Crime, & Law*, 18, 713–731.

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, 128, 203–235.

Elfenbein, H. A., & Eisenkraft, N. (2010). The relationship between displaying and perceiving nonverbal cues of affect: A meta-analysis to solve an old mystery. *Journal of Personality and Social Psychology*, 98, 301–318.

Elfenbein, H. A., Foo, M. D., White, J., Tan, H. H., & Aik, V. C. (2007). Reading your counterpart: The benefit of emotion recognition accuracy for effectiveness in negotiation. *Journal of Nonverbal Behavior*, 31, 205–223.

Gray, H., & Tickle-Degnen, L. (2010). A meta-analysis of performance on emotion recognition tasks in Parkinson's disease. *Neuropsychology*, 24, 176–191.

Halberstadt, A. G. (1983). Family expressiveness styles and nonverbal communication skills. *Journal of Nonverbal Behavior*, 8, 14–26.

Halberstadt, A. G., & Eaton, K. L. (2003). A meta-analysis of family expressiveness and children's emotion expressiveness and understanding. *Marriage & Family Review*, 34, 35–62.

Hall, J. A. (1978). Gender effects in decoding nonverbal cues. *Psychological Bulletin*, 85, 845–857.

Hall, J. A. (1983). *Nonverbal sex differences: Communication accuracy and expressive style*. Baltimore: Johns Hopkins University Press.

Hall, J. A., Andrzejewski, S. A., Murphy, N. A., Schmid Mast, M., & Feinstein, B. (2008). Accuracy of judging others' traits and states: Comparing mean levels across tests. *Journal of Research in Personality*, 42, 1476–1489.

Hall, J. A., Andrzejewski, S. A., & Yopchick, J. E. (2009). Psychosocial correlates of interpersonal sensitivity: A meta-analysis. *Journal of Nonverbal Behavior*, 33, 149–180.

Hall, J. A., Halberstadt, A. G., & O'Brien, C. E. (1997). "Subordination" and nonverbal sensitivity: A study and synthesis of findings based on trait measures. *Sex Roles*, 37, 295–317.

Hall, J. A., Schmid Mast, M., & Latu, I. (2015). The vertical dimension of social relations and accurate interpersonal perception: A meta-analysis. *Journal of Nonverbal Behavior*, 39, 131–163.

Hartwig, M., & Bond, C. F., Jr. (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychological Bulletin*, 137, 643–659.

Hartwig, M., & Bond, C. F., Jr. (2014). Lie detection from multiple cues: A meta-analysis. *Applied Cognitive Psychology*, 28, 661–676.

Hoekert, M., Kahn, R. S., Pijnenborg, M., & Aleman, A. (2007). Impaired recognition and expression of emotional prosody in schizophrenia: Review and meta-analysis. *Schizophrenia Research*, 96, 135–145.

Kirkland, R. A., Peterson, E., Baker, C. A., Miller, S., & Pulos, S. (2013). Meta-analysis reveals adult female superiority in "Reading the Mind in the Eyes" Test. *North American Journal of Psychology*, 15, 121–146.

Kohler, C. G., Hoffman, L. J., Eastman, L. B., Healey, K., & Moberg, P. J. (2011). Facial emotion perception in depression and bipolar disorder: a quantitative review. *Psychiatry Research*, 188, 303–309.

Kohler, C. G., Walker, J. B., Martin, E. A., Healey, K. M., & Moberg, P. J. (2009). Facial emotion perception in schizophrenia: a meta-analytic review. *Schizophrenia Bulletin*, 36, 1009–1019.

Kok, T. B., Post, W. J., Tucha, O., de Bont, E. S. J. M., Kamps, W. A., & Kingma, A. (2014). Social competence in children with brain disorders: A meta-analytic review. *Neuropsychology Review*, 24, 219–235.

Kurtz, M. M., & Richardson, C. L. (2012). Social cognitive training for schizophrenia: A meta-analytic investigation of controlled research. *Schizophrenia Bulletin*, 38, 1092–1104.

Lavoie, M., Plana, I., Lacroix, J. B., Godmaire-Duhaime, F., Jackson, P. L., & Achim, A. M. (2013). Social cognition in first-degree relatives of people with schizophrenia: A meta-analysis. *Psychiatry Research*, 209, 129–135.

Lozier, L. M., Vanmeter, J. W., & Marsh, A. A. (2014). Impairments in facial affect recognition associated with autism spectrum disorders: A meta-analysis. *Development and Psychopathology*, 26, 933–945.

Luke, N., & Banerjee, R. (2013). Differentiated associations between childhood maltreatment experiences and social understanding: A meta-analysis and systematic review. *Developmental Review*, 33, 1–28.

Marsh, A. A., & Blair, R. J. R. (2008). Deficits in facial affect recognition among antisocial populations: A meta-analysis. *Neuroscience and Biobehavioral Reviews*, 32, 454–465.

McClure, E. B. (2000). A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychological Bulletin*, 126, 424–453.

Merten, J. (2005). Culture, gender and the recognition of the basic emotions. *Psychologia*, 48, 306–316.

Mitchell, A. E., Dickens, G. L., & Picchioni, M. M. (2014). Facial emotion processing in borderline personality disorder: A systematic review and meta-analysis. *Neuropsychology Review*, 24, 166–184.

Mitchell, A. J. (2008). The clinical significance of subjective memory complaints in the diagnosis of mild cognitive impairment and dementia: A meta-analysis. *International Journal of Geriatric Psychiatry*, 23, 1191–1202.

Mitchell, A. J., & Kakkadasam, V. (2011). Ability of nurses to identify depression in primary care, secondary care, and nursing homes – A meta-analysis of routine clinical accuracy. *International Journal of Nursing Studies*, 48, 359–368.

Mitchell, A. J., Meader, N., Bird, V., & Rizzo, M. (2012). Clinical recognition and recording of alcohol disorders by clinicians in primary and secondary care: A meta-analysis. *The British Journal of Psychiatry*, 201, 93–100.

Mitchell, A. J., Meader, N., & Pentzek, M. (2011). Clinical recognition of dementia and cognitive impairment in primary care: A meta-analysis of physician accuracy. *Acta Psychiatrica Scandinavica*, 124, 165–183.

Mitchell, A. J., Rao, S., & Vaze, A. (2010). Do primary care physicians have particular difficulty identifying late-life depression? A meta-analysis stratified by age. *Psychotherapy and Psychosomatics*, 79, 285–294.

Mitchell, A. J., Rao, S., & Vaze, A. (2011). Can general practitioners identify people with distress and mild depression? A meta-analysis of clinical accuracy. *Journal of Affective Disorders*, 130, 26–36.

Mitchell, A. J., Vaze, A., & Rao, S. (2009). Clinical diagnosis of depression in primary care: A meta-analysis. *Lancet*, 374, 609–619.

Murphy, N. A., & Hall, J. A. (2011). Intelligence and nonverbal sensitivity: A meta-analysis. *Intelligence*, 39, 54–63.

Nummenmaa, L., & Calvo, M. G. (2015). Dissociation between recognition and detection advantage for facial expressions: A meta-analysis. *Emotion*, 15, 243–256.

O'Toole, M. S., Hougaard, E., & Mennin, D. S. (2013). Social anxiety and emotion knowledge. *Journal of Anxiety Disorders*, 27, 98–108.

Plana, I., Lavoie, M., Battaglia, M., & Achim, A. M. (2014). A meta-analysis and scoping review of social cognition performance in social phobia, posttraumatic stress disorder and other anxiety disorders. *Journal of Anxiety Disorders*, 28, 169–177.

Ruffman, T., Henry, J. D., Livingstone, V., & Phillips, L. H. (2008). A meta-analytic review of emotion recognition and aging: Implications for neuropsychological models of aging. *Neuroscience and Biobehavioral Reviews*, 32, 863–881.

Samamé, C., Martino, D. J., & Strejilevich, S. A. (2012). Social cognition in euthymic biopolar disorder: Systematic review and meta-analytic approach. *Acta Psychiatrica Scandinavica*, 125, 266–280.

Sayla, G. N., Vella, L., Armstrong, C. C., Penn, D. L., & Twamley, E. W. (2013). Deficits in domains of social cognition in schizophrenia: A meta-analysis of the empirical evidence. *Schizophrenia Bulletin*, 39, 979–992.

Shahrestani, S., Kemp, A. H., & Guastella, A. J. (2013). The impact of a single administration of intranasal oxytocin on the recognition of basic emotions in humans: A meta-analysis. *Neuropsychopharmacology*, 38, 1929–1936.

Thompson, A. E., & Voyer, D. (2014). Sex differences in the ability to recognize non-verbal displays of emotion: A meta-analysis. *Cognition and Emotion*, 28, 1164–1195.

Tskhay, K. O., & Rule, N. O. (2013). Accuracy in categorizing perceptually ambiguous groups: A review and meta-analysis. *Personality and Social Psychology Review*, 17, 72–86.

Uljarevic, M., & Hamilton, A. (2013). Recognition of emotions in autism: A formal meta-analysis. *Journal of Autism and Developmental Disorders*, 43, 1517–1526.

Van Hemert, D. A., Poortinga, Y. H., & van de Vijver, F. J. R. (2007). Emotion and culture: A meta-analysis. *Cognition and Emotion*, 21, 913–943.

Ventura, J., Wood, R. C., Jimenez, A. M., & Hellemann, G. S. (2013). Neurocognition and symptoms identify links between facial recognition and emotion processing in schizophrenia: Meta-analytic findings. *Schizophrenia Research*, 151, 78–84.

Wagner, M. F., Milner, J. S., McCarthy, R. J., Crouch, J. L., McCanne, T. R., & Skowronski, J. J. (2014). Facial emotion recognition accuracy and child physical abuse: An experiment and a meta-analysis. *Psychology of Violence*, 5, 154–162.

2 Accuracy of judging emotions

Tanja Bänziger

Abstract

This chapter proposes an overview of research conducted in recent years on emotion recognition accuracy (ERA). The focus is on outlining the main methods and findings, and also the complexity of the processes involved. Emotion recognition is multimodal (several interpersonal communication channels are involved, such as face, voice, body postures, and gestures) and it is probably as dependent on contextual and social cues as it is on individual skills. The chapter discusses issues related to what accurate emotion recognition is and how it can be estimated. The constructs involved in major studies of emotion recognition and the main processes involved in judging emotions are also addressed.

Judging emotions based on another person's nonverbal behavior is a competence that is crucial for social functioning and has been related to psychological health and well-being. Accurately judging emotional expressions is a component of *interpersonal sensitivity* – a broader construct described by Hall, Andrzejewski, and Yopchick (2009) – and has also been described as an essential component of *emotional intelligence* (EI; see, e.g., the definition by Mayer, Salovey, Caruso, & Sitarenios, 2003). The ability to correctly infer the emotional state of other people has been investigated by many researchers in a large variety of contexts and in relation to numerous research questions.

This chapter outlines a perspective that reflects the most prevalent approaches in research on nonverbal communication of emotion and that is also directly coupled with classical behavioral studies of emotional expression and emotional communication. The most widespread approaches to the assessment of *emotion recognition accuracy* (ERA) are presented first. This description is then followed by an overview of the main findings in this field, which have largely shown that emotion recognition can be fairly accurate, also cross-culturally, but that accuracy may vary depending on the communication channels or the emotions considered. The ensuing and final sections develop various aspects related to the emotion recognition construct and to processes involved in emotion recognition, with the aim to outline the actual complexity of the phenomenon. Those sections also allow questioning possible limitations of the classical approach to the assessment of ERA.

The classical approach, which is presented first, is constrained by a number of central goals. Researchers interested in ERA have emphasized assessment methods that allow quantifying ERA. This in turn allows comparing ERA in various groups, for example, across cultures, gender, or in various clinical groups. ERA is furthermore defined as a nonverbal communication skill. Assessment methods and studies therefore focus on studying ERA independently from verbal communication (no verbal information on the emotional states is involved in the assessments of ERA or, alternatively, the verbal information is strictly controlled and manipulated). ERA is assessed as a competence and is tested most commonly in tasks that require a person to associate an emotional display (e.g., a photograph of another person displaying an emotion) with an emotional label (a verbal category).[1] This observation implies that a number of choices are made regarding the emotion categories used for the assessment (which emotions should be recognized) and also on the communication channels used. A great number of studies have, for example, considered facial expressions alone (often to be recognized in static pictures). In recent years, however, studies involving multiple expressive modalities have become more numerous and instantiate a notable development toward more differentiated assessments of ERA. Those choices (regarding emotion categories or communication channels) depend both on the goals and the theoretical background of the researchers and are obviously related to the construct validity of measures of ERA.

Assessment of emotion recognition accuracy

There are numerous assessment methods for ERA. As mentioned earlier, the most common method is to present a test participant with nonverbal expressions – photographs or video recordings for facial expressions,

[1] This statement simplifies a more complex reality. There are methods for the assessment of "interpersonal skills" that involve a great deal of emotion recognition and are based on completely different paradigms. Such approaches are, however, less specifically interested in ERA, and they tend to focus on a more global "interpersonal competence." This is especially true for approaches that involve actual dyadic or group interactions, in which research participants not only need to perceive and understand the (emotional) reactions of their interaction partner(s), but also need to respond appropriately to those. Further methods used in interpersonal perception studies have been more thoroughly described in other contributions; see for example Hall, Bernieri, and Carney (2005).

postures, or gestures or audio recording for vocal expressions – and to ask the test participant to label those expressions. In most cases, a list of predetermined labels/categories is provided and the participants are asked to select the best matching label. The correct answer (accuracy criterion) can be defined in different ways, but it is often related to the expressive intention or to the self-reported emotion of the sender (the person displaying the expression in the photograph or the recording) or to theoretical expectations of the researchers. Multiple variants of this procedure have been used in different studies. Neuroimaging studies and EEG studies often involve especially large requirements of control over the stimuli that can be presented to the research participants, in order to avoid potential confounds with uncontrolled features of the stimuli.[2] The expressions (facial, postural, or vocal) used in various studies are commonly selected and/or adapted to specific research interests and questions. This implies that – in the absence of a standardized instrument used by different groups of researchers – many studies have published results that are not directly comparable. The same holds true also for clinical research where the instruments for the assessment of emotion recognition are also frequently adjusted for specific studies and specific clinical groups (see Chapter 9).

A limited number of standardized and validated tests have, however, been used and described repeatedly in the literature (see also Hall, 2001; Bänziger, 2014). The Profile of Nonverbal Sensitivity (PONS; Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979), based on 220 2-s video clips (20 recordings, replayed in 11 combinations of expressive modalities), asks the test participants to identify a variety of interpersonal situations and affective states. The PONS allows calculation of separate scores for facial communication, vocal communication, and communication based on gestures and postures. The Japanese and Caucasian Brief Affect Recognition Test (JACBART; Matsumoto et al., 2000) includes only static pictures of faces (test items) corresponding to seven different basic emotion categories.[3] In this test, the emotional expressions are posed according to the facial prototypes theoretically matching the seven basic emotions. The expressive faces are presented very briefly (the longest exposure time is 1/5 s, the shortest 1/15 s, in different test

[2] One example can be that the sound intensity cannot be much larger in some stimuli than in others, or that they need to be of the same length. Faces showing different emotional expressions cannot vary in hairstyle or more simply the identity of the speaker needs to be constant across emotional expressions. This leads the researchers to select and manipulate the stimuli in order to control for other variations that might be confounded with the expressive cues.

[3] Some of the theoretical postulates underlying the different methods (here the inclusion of *basic emotions*) are developed to a larger extent later on in this chapter.

versions) and are framed by the exposure of a neutral expression of the same sender, before and after the brief exposure of the expressive face. The Diagnostic Analysis of NonVerbal Accuracy (DANVA) tests (Nowicki, 2006; Nowicki & Duke, 1994) include several independent instruments; all assess the recognition of four emotions based on nonverbal cues. Some DANVA tests are based on static pictures of facial expressions of adults or of children; others are based on recordings of emotional voices (separate portrayals produced by different senders are used for the recognition of emotional prosody), and some on photographs of postures. The Emotion Recognition Index (ERI; Scherer & Scherer, 2011) also uses static pictures of expressive face (from the Pictures of Facial Affect database, Ekman & Friesen, 1976) and independent vocal expressions (from the International Study of Vocal Emotion Expression; Scherer, Banse, & Wallbott, 2001) and asks the test-takers to choose among five alternative emotion categories.

Further studies used picture material from the collection of Ekman and Friesen (1976) or other picture collections such as the Karolinska Directed Emotional Faces (KDEF; Lundqvist, Flykt, & Öhman, 1998).[4] For the aforementioned collections of photographs, however, the specific form taken by the assessment method can be variable from one study to another, given that different researchers will select different portrayals and/or propose different response alternatives. In recent years, it has been more common also to see studies in which the photographs from various databases were morphed or manipulated to generate new pictures or to create an artificial animation of the static pictures.[5]

Among other recent developments, further audiovisual tests of emotion recognition have been developed. Bänziger, Grandjean, and Scherer (2009) introduced the Multimodal Emotion Recognition Test (MERT), which is based on 30 audio–video recordings. This test requires the test participant to identify 10 emotions and allows computing separate scores for vocal, facial (static and dynamic) expressions, and the combination of facial and vocal expressions. Schlegel, Grandjean, and Scherer (2013a, 2014) developed the Geneva Emotion Recognition Test

[4] The existence of newer photograph collections, for example the Radboud collection (Langner et al., 2010), and also databases of audiovisual expressions, for example the ENTERFACE corpus (Martin, Kotsia, Macq, & Pitas, 2006) or the GEMEP corpus (Geneva Multimodal Emotion Portrayals; Bänziger & Scherer, 2010) can be mentioned as well. A complete list of all the projects conducted in recent years would be difficult to include here. The newer databases have not been used in as many different investigations as the tests and photograph collections mentioned earlier.

[5] An illusion of movement is produced by morphing the neutral facial expression of a given sender onto an expressive picture of the same sender, or also by morphing one expression (e.g., anger) onto another one (e.g., fear).

(GERT) using a selection of the Geneva Multimodal Emotion Portrayals (GEMEP) audio–video recordings (Bänziger & Scherer, 2010), which involves the recognition of 14 emotions based on combined audio–video (vocal and facial) expressions.[6] Ekman (2003c) developed an instrument for facial ERA that is included in the Micro Expression Training Tools (METT) and is similar to JACBART (but based on different photographs); this test has now also been used in several published studies (see, e.g., Russell, Chu, & Phillips, 2006; Matsumoto & Hwang, 2011).

In order to include emotional variation alone and keep the senders and the background/context as homogeneous as possible across emotions, those instruments use acted portrayals or posed facial expressions.[7] The issue of the representativeness of acted or posed expressions (in relation to emotional expressions produced in everyday life) and the ecological validity of the instruments has been questioned in this connection. Kang (2012) proposed to use an emotion recognition test based on spontaneous expressions[8] recorded in interviews. The Spontaneous Expressions Recognition Test (SERT; Kang, 2012) is based on nine audio–video recordings and distinguishes four emotional reactions. The nine test items are scored with respect to the emotions reported by the senders after the recording of the video segments and were selected based on expert opinion and pretests. Baron-Cohen and collaborators (Golan, Baron-Cohen, & Hill, 2006; Golan & Baron-Cohen, 2006) questioned the construct validity of the classical approach and proposed a test

[6] The author of this chapter and several collaborators also work on the development and validation of a multimodal test based on 72 GEMEP recordings. This test allows computing separate scores for video only, audio only, and audio–video emotion portrayals and involves recognition/discrimination of 12 emotions. The validation data are to date unpublished.

[7] For acted expressions, the senders are provided with scenarios or definitions of the emotional states to portray. Sometimes no further techniques are used, but in some cases, the actors (when they are professionals) use "method acting" techniques, which resort to emotional induction mechanisms (the central idea is that the actor can make use of his own emotional reactions and elicit them through the activation of related behaviors/expressions). Posed facial expressions are expressions where the sender is provided with detailed instructions and training regarding the facial muscles he/she is expected to contract in order to achieve an expression that is defined by the researchers based on theoretical expectations.

[8] The question of the definition of "spontaneity" in this domain is a complex issue, which is over simplified here. Portrayals produced by actors although not "spontaneous" can involve genuine emotional expressions, provided that the actors use appropriate techniques. Furthermore, the absence of explicit request to portray an emotion does not guarantee that expressions recorded in the field or in the laboratory are exempt of acting. Some authors (see, e.g., Scherer & Bänziger, 2010, or Scherer, 2013; in line with Goffman's 1959 theory of self-presentation) have argued that most emotional expressions (including expressions that reflect genuine emotional reactions) are regulated to some extent, comply with social demand characteristics and are therefore not entirely "spontaneous."

including a different construct of emotion. The Cambridge Mindreading Face-Voice Battery (CAM; Golan et al., 2006) distinguishes 20 "emotion concepts," which are described as being more subtle and more complex than basic emotions (e.g., "guarded," "insincere," "confronted," or "nostalgic"). Each of the 20 emotion concepts is represented by five items (either vocal or facial expressions produced by actors; the facial expressions are video recorded portrayals, without speech or sound). Another test stemming from the same approach is the Reading the Mind in the Eyes Test (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001). This test also makes use of more varied "emotion concepts" than the classical ERA tests. The participants are shown 36 photographs of eyes (sampled from old British magazines) and are required to choose one of four labels for each photograph (different alternative answers are provided for different pictures).

This overview is not exhaustive but aims at illustrating the variety of instruments used to assess emotion recognition and the main issues in this domain. There is some agreement on the need for standardized and validated instruments, which isolate emotional expressions from other information. Verbal information, situational context, and speaker characteristics have to be controlled in order to assess nonverbal emotion recognition independently of other skills involved in social perception or interpersonal communication. But this entails a cost. Ecological and face validity will probably remain questionable in standardized tests of emotion recognition independently of the stimuli source (e.g., acted or spontaneous) as long as the instruments aim at excluding other aspects of interpersonal communication in order to focus on emotional expressions alone.

A related issue is the so-called accuracy criterion problem, which has been described by Hall, Bernieri, and Carney (2005). In classical test of ERA, this problem is usually solved by defining the accurate answer for any test item by the expressive intention of the sender. This does, however, imply that the accuracy criterion essentially relies on the expressive skills of the senders that are involved in the tests. Expressive skills are generally considered to be as variable as recognition skills. Hence, the tests ultimately make the accuracy of the test-takers dependent upon the variable (and possibly unreliable) expressive skills of the senders (or upon theoretical postulates regarding typical emotional expressions, in the case of posed facial expressions). This issue is regularly discussed in studies reporting data on test development and validation and is a legitimate and central concern in this field. Relying only on the expressive skills of the sender is risky (what if the sender is showing ambiguous signals? or lacks skills for expressing certain emotions?) but using agreement in recognition to identify valid test items automatically leads to include only "easy items" (expressions that most people will recognize, by design). More generally, using consensus (the most frequent emotional attribution) or expert judgments as a criterion for recognition accuracy is obviously rather problematic. Using consensus scoring for ERA is comparable to defining the correct answers in an intelligence test based on the answers of the majority of test-takers. However, most instruments do in fact involve a selection, based on consensual recognition (and/or expert judgments), somewhere along the test construction procedure.

The description of the standardized instruments provided above also aims at showing that there is no agreement regarding the communication channels (voice, face, gestures) to be included, or the number and nature of emotions to be considered for the assessment of emotion recognition. As mentioned earlier, the assessment method will often be adjusted to the specific research question or theoretical background of the researchers. The external and convergent validity of such measures can of course be estimated based on correlates with other phenomena or even between different emotion recognition tests (see Chapter 18). The issue of the construct that is measured (which emotions have to be recognized, based on which communication channel/s) is, however, an important matter that needs to be addressed and developed to a larger extent in future research. It is clear that there is at present great heterogeneity in the methods described in the literature for the assessment of ERA. This issue could also be rephrased as a need to develop a better understanding of the different facets that might be involved in ERA and by extension a better understanding of the construct(s) measured in various ERA tests. Some questions related to the various facets that might be involved in ERA will be addressed later on in this chapter in the section dealing with the construct(s) involved in the definition of ERA.

How accurate is emotion recognition?

Comparison of accuracy across studies presents an obvious challenge that has been addressed by Hall, Andrzejewski, Murphy, Schmid Mast, and Feinstein (2008). Within research on ERA most accuracy scores are expressed as proportions of correct answers. However, the accuracy achieved in tests that differ in number of alternative answers and also in distinctiveness between the possible answers will yield largely different results. To take a specific example, recognition of "happiness" in tests using static pictures with smiling faces (the prototypical "Duchenne smile") with comparably long exposure times and with only one positive emotion category to choose from (all alternative answers are negative labels: e.g., fear, sadness, anger) will tend to lead to 100% correct recognition (i.e., perfect accuracy). On the other hand, asking participants to differentiate between conceptually related states (e.g., rate if an expression displays anger or irritation, satisfaction or pride) will decrease accuracy scores substantially. In other words, accuracy is to a large extent test dependent. Some rating tasks are themselves more difficult than others: the more difficult the task, the lower the average accuracy. This does not mean, however, that comparison is impossible. Accuracy can be compared for different groups or different persons provided the same test (or at least a similar test) is used. It is furthermore possible to tell if test-takers depart significantly from the accuracy level that would be obtained by chance alone. A comparison that is made in most studies and that shows that people do recognize emotional expressions with at least some accuracy even in the most difficult tasks (e.g., when required to make subtle distinctions or with very limited exposure to the sender's nonverbal behavior).

Hall et al. (2008) have reported comparisons between scores for various assessments of interpersonal perception. In this meta-analysis, results from multiple studies are converted to a common metric, the Proportion Index (pi), which expresses accuracy as a proportion for a two-class categorization (i.e., pi = 1 corresponds to perfect accuracy and pi = .50 corresponds to accuracy level that would be reached by guessing; Rosenthal & Rubin, 1989). Hall et al. report average estimates for values of pi larger than .80 for most tests of ERA. For dynamic stimuli (video or audio recordings) the reported average estimates are slightly lower (mostly between .70 and .80). Hall et al. (2008) also report pi values for other interpersonal accuracy measures, and the estimates for other constructs are generally lower (around .60 for accuracy in judging personality traits or intelligence). The results presented by Hall et al. indicate that the research results reported prior to 2008 globally support the notion that emotion recognition is quite accurate, more accurate in fact than most other interpersonal judgments that have been investigated.

Further estimates of accuracy support the notion of accurate emotion recognition and in addition provide some insight into the extent of variation (in accuracy) that might be expected for different communication channels and different emotions. Scherer, Clark-Pollner, and Mortillaro (2011) reviewed accuracy results with a focus on cross-cultural differences. They report few differences in accuracy when comparing Western and non-Western cultures and an in-group advantage that is observed only for Western cultures and static facial expressions. Their description allows to see that ERA varies substantially for different emotions (e.g., an average estimate of 91% accuracy for happiness[9] in static facial expressions and 72% for anger). This review also confirms that dynamic expressions (vocal only but also moving facial expressions in video recordings) yield lower accuracy than static facial expressions. Interactions between emotions and channels are also clearly shown in this review. For example, an average estimate of 54% for happiness and 74% for anger in vocal only expressions indicates that anger is more accurately communicated in vocal expressions, while happiness is more accurately communicated in static facial expressions (see estimates reported above).

In summary, although accuracy is very dependent on the test items selected to estimate it, the accumulated results of past research indicate not only that emotion recognition is accurate but also that it varies depending on the emotions and the communication channels that are considered. The overview of methods and main findings presented above shows the central importance of the definition of the constructs involved in ERA. The following sections develop, first, the theoretical implications and expectations related to the *definition of emotions* in ERA and, later, discuss the *implication of various communication channels* in the definition of ERA.

Emotion constructs in ERA

The theoretical background and conceptualization of emotions and emotional expressions is crucial especially for the definition of the categories (or dimensions) that are expected to be accurately recognized. One dominant school of thought considers emotional reactions as functional biological entities. This perspective can be linked back in time to the ideas of Charles Darwin (1872/1998). Darwin's ideas about emotions were articulated around the issue of emotional expressions and their evolutionary functions. A simplified account of this view is that humans (and many other species) are biologically equipped with a limited number of distinct universal emotional reactions (so-called basic emotions[10]) and

[9] The estimates in the review by Scherer et al. (2011) are not corrected for the number of alternative answers. Given that most ERA tests use 4–7 categories, the accuracy figures in this case should be interpreted accordingly. The values reproduced here are for Western senders and Western perceivers (i.e., within culture) and were chosen because much more studies and results are available in this configuration.

[10] There are unresolved disputes on the number of emotions that should be considered basic. Many studies have considered six or seven basic emotions: happiness, fear, surprise, sadness, anger, disgust (and contempt). But this list has been extended to include, for instance, more than one positive basic emotion. In one account, Ekman (2003b) acknowledged that there might be up to 16 "enjoyable emotions," a theoretical speculation that he considered needed to be further investigated. Other researchers have focused on facial expressions of emotions that may be considered as "compounds" of the above named categories (e.g., Du, Tao, & Martinez, 2014) or on the contrary have reported to

that each of those reactions is characterized by a specific (and universal) pattern of physiological and expressive responses. In this perspective, expressions are seen as functional in several respects, but essentially to signal emotional states to others. The influential work of Ekman and collaborators since the 1960s (see, e.g., Ekman, 2003a; Ekman & Friesen, 1971; Ekman & Keltner, 1997) has repeatedly shown that people can accurately recognize emotional facial displays, also cross-culturally (see Elfenbein & Ambady, 2002 and 2003 for a discussion and a review, as well as Chapter 16).

It is noteworthy that the research carried out in this tradition has been concerned as much with the perception (recognition) of emotional expressions as with expressions per se. In fact, the early studies (e.g., Ekman & Friesen, 1971) and also some of the most recent developments mostly rely on *recognition* of specified facial displays, which are produced following a set of instructions regarding the muscles to be contracted in order to achieve a desired "expressive configuration" (i.e., the resulting facial displays are not spontaneous expressions of felt emotional states, but displays produced on the basis of detailed instructions and training). Several studies have been published showing that such "expressive prototypes," and also emotional expressions more generally, are indeed perceived categorically in experimental settings (see, e.g., Laukka, 2005; Laukka, Audibert, & Aubergé, 2012; Young et al., 1997).

For ERA, the main implication of this theoretical perspective is that accurate recognition is expected to be a part of our species evolutionary heritage, an expectation that has been largely confirmed by studies showing that ERA is indeed very large for selected expressions (photographs of prototypical facial expressions), reaching sometimes near-perfect accuracy (close to 100% correct recognition) in experimental recognition tasks.

Cultural differences in emotional expression and interpretation of expressions displayed by others are not denied by the "Darwinian" perspective outlined above, but are conceived as modulations of otherwise universal processes (see, e.g., Matsumoto & Ekman, 1989). This perspective on emotion has, however, been challenged by other conceptualizations. One main opposing view has been defended by Russell (1994, 2003, 2012), who argued that emotions are psychological constructions. According to this perspective, all emotional states are defined by two underlying continuous dimensions which compose the physiological and universal part of the reaction (the so-called core affect): valence

have found a smaller number of basic categories based on distinctive facial expressions (e.g., Jack, Garrod, & Schyns, 2014).

(pleasure vs. displeasure) and arousal (high vs. low). All other differences between emotional reactions are psychological constructions that are variable across cultures and individuals. This perspective also implies that emotional expressions carry information about valence and arousal reactions only (core affect is expressed, not distinct categories, such as anger, fear, or happiness[11]). Accurate recognition should therefore be possible only for core affect, particularly when observers rely only on nonverbal expressive cues; when other cues are available, for example, verbal information about the situation that elicits an emotional reaction, observers might agree on further emotional attributions, based on culturally shared scripts.

This conceptualization is currently vivaciously defended by Barrett and collaborators under the heading "Conceptual Act Theory" (Barrett, Wilson-Mendenhall, & Barsalou, 2014). Recent publications on emotion recognition in this perspective have focused on demonstrating that the recognition of prototypical expressions, previously used to demonstrate universal recognition of basic emotion categories, is dependent on the availability of linguistic categories and/or is not truly universal (Gendron, Lindquist, Barsalou, & Barrett, 2012; Gendron, Roberson, van der Vyver, & Barrett, 2014a; 2014b; Lindquist, Barrett, Bliss-Moreau, & Russell, 2006).

A further conceptualization of emotion is represented in appraisal theories (cognitive perspectives on emotions), which propose that emotional expressions reflect the cognitive appraisals taking place before the emotional reactions (Sander, Grandjean, Kaiser, Wehrle, & Scherer, 2007; Scherer, 2009; Scherer & Ellgring, 2007). In this view emotion recognition implies a capacity to identify emotion antecedent, abstract, "mental states." Emotional expressions are conceived as having an informative function not only on the sender's internal reactions (emotions) but also on how the sender responded to a (often social) situation. The observer makes inferences, as a part of the emotion recognition process, about the situation and the social relationships involved. For example, an observer might judge that a person displaying a given facial or vocal expression was facing a threat or, in a dimensional perspective, that she witnessed something highly unexpected and very negative, and experienced that she had very little control over the situation (the same expression might be labeled "fearful" in a basic emotion account). A few studies support this assumption

[11] "Facial, vocal, and autonomic changes occur and are accounted for (a) by core affect and (b) as part of, preparation for, or recovery from instrumental action. (There is no nonverbal expressive signal or pattern of autonomic nervous system activity unique to each discrete emotion.)" (Russell, 2003, p. 150).

for facial expressions (Scherer & Grandjean, 2008) as well as vocal expressions (Laukka & Elfenbein, 2012). We can furthermore assume, based on the many findings reported in the literature on interpersonal accuracy (see, e.g., Hall et al., 2009), that people indeed formulate not only judgments on emotions, based on nonverbal behavioral cues, but also a whole range of other judgments on person and situation characteristics. It is therefore likely that an emotional expression observed in an ecological context will indeed trigger multiple attributions on part of the observer. The appraisal perspective on emotion recognition implies, however, that such inferences can and will be drawn also from the nonverbal expressions alone (i.e., when no other, contextual, information is present and when we have no further information on the senders). An important aspect is that this perspective on emotions is intrinsically integrated with other areas of "social perception." To illustrate this claim with a concrete example, we can mention that "anger" (using the basic emotion term) is a reaction that is conceived (in a cognitive perspective) as resulting from situations that are *appraised as* obstructing goals and/or violating social norms; furthermore it has been speculated that anger is associated with dominance (or a sense of control over the circumstances that elicit the emotion). This example illustrates how the processing of emotional expressions may involve (in a cognitive, appraisal, perspective) judgments that are directly related to the sender (dominant, powerful) or the situation (e.g., a social norm has been violated, or something unfair happened). The implications for recognition *accuracy* are rarely discussed but implicitly the postulate is that emotional nonverbal expressions (without further information) allow one to *accurately* assess not only emotional states or feelings but also a range of mental/cognitive states of the sender that are associated with those emotions and expressions.

In summary, different theoretical perspectives support rather different accounts of what constitutes emotion recognition. Which theoretical perspective provides the most adequate description of emotional reactions in order to study ERA is an unresolved empirical question and a matter of debate. The goal of this description is to emphasize that *judging emotions* (emotion recognition) is not in essence limited to matching labels (emotion words: happiness, sadness, fear) to prototypical facial displays. However, the most prevalent method for the assessment of ERA has been and still is discrimination of basic emotions in prototypical facial expressions. Alternative accounts of emotions are less represented in this research field but will hopefully continue to attract more attention in future investigations of ERA. One very concrete evolution in recent years is reflected in the use of ERA assessments that are not restricted to discrimination between basic emotions only. More developments in this direction will allow gaining further knowledge on ERA, beyond the classical approach driven by the basic emotion perspective.

Communication channels in ERA

The descriptions provided in earlier sections of this chapter illustrate that recognition of emotion has been principally studied in connection with facial expressions (and facial perception). There is nevertheless a shared understanding that facial emotional expressions are rarely isolated (in everyday life) from other information and specifically from other behavioral nonverbal information. Gestures and postures ("body language") have been considered as important cues for the nonverbal communication of emotion, already in early accounts of nonverbal communication (early ethological descriptions, e.g., also including Darwin's observations, focused as much on gestures and body postures as on facial expressions). The same is true also for nonverbal vocal emotional expressions, which in addition point to the fact that emotional expressions are necessarily dynamic and that the processing of emotional expression must involve some form of integration of this dynamic dimension (integration of cues that are fluctuating, evolving quickly in time). Emotions can further be communicated by touch (Hertenstein, Holmes, McCullough, & Keltner, 2009) and smells (Chen & Haviland-Jones, 2000) and are also expressed and perceived in all art forms, namely singing, dancing, acting, painting, and writing (Silvia, 2005; Pelowski & Akiba, 2011). Studies on emotion communication via touch or smell are, however, relatively rare and will be left out of the present account. The role of emotion communication in artistic production and performance bears a clear resemblance with nonverbal emotional communication, but falls outside the scope of the present chapter.

In recent years, the interest in the integration of auditory (vocal) and visual (facial, but also postural/gestural) cues in emotion perception has expanded, especially in connection with neurological studies and distributed models of face processing. Several studies have convincingly shown that the processing of facial expressions is modulated by emotions expressed in the other channels (voice and body). Massaro and Egan (1996) showed, for example, experimentally (with manipulated incongruent expressions) that research participants rely differentially on one or the other channel (face or voice) depending on the information presented in the respective channels. De Gelder and collaborators have summarized such findings by stating that emotional prosody and postures affect facial emotion recognition even when the expressions are not processed consciously or when research participants are instructed to ignore one of the modalities. According to these authors, multisensory integration in emotion perception/recognition is "automatic and mandatory" (see Brosch & Grandjean, 2013; De Gelder & Van den Stock, 2011; De Gelder, Stienen, & Van den Stock, 2013; for more developments on this aspect).

Studies have furthermore shown that different communication channels are not merely redundant but most probably also provide complementary information. Since most studies are conducted on emotion recognition in just one channel, there are comparably few results that allow direct comparison of recognition in multiple channels and assess the respective contribution of various channels (but see Massaro & Egan, 1996; Shackman & Pollak, 2005). Reviews of emotion recognition studies in separate channels have repeatedly pointed out that some emotions appear to be better communicated (i.e., better recognized) in specific channels. The most frequently cited example is "disgust," an emotion category that is poorly recognized when actors attempt to express it vocally (nonverbally) in speech, but which is easy to display and to recognize in the face. Other aspects can be better communicated vocally. For instance, emotional arousal is easily perceived in vocal expressions. Arousal is associated with increased vocal effort, which translates into louder, higher pitched voice and also affects the spectral energy distribution of the voice signal. Emotional valence is comparably more difficult to convey vocally (but has been associated with "Duchenne smiles" in facial expressions). Such results have been reported in numerous publications for several decades and have been summarized in review articles (e.g., Scherer et al., 2011; or the results published for the GEMEP database; Bänziger & Scherer, 2010).

A unitary competence or a set of related abilities

The focus on various communication channels almost automatically leads to discussions regarding the broader construct of emotion recognition. The idea that "multisensory integration is automatic and mandatory" supposes that emotional expressions are habitually processed in an "integrated" way (expressive cues from different channels are perceptually integrated before emotion judgments are formulated) and that it is reasonable to consider ERA as a unitary competence (one general competence that would be reflected in ERA for different communication channels, should they be assessed separately). Some authors who compared separate ERA measures for different communication channels reported, however, that those measures are poorly correlated (see, e.g., Scherer & Scherer, 2011). Other accounts furthermore suggest that ERA for positive and negative emotions might reflect fairly independent (uncorrelated or moderately correlated) abilities (Suzuki, Hoshino, & Shigemasu, 2010). ERA for different emotion categories have also been reported to display low inter-correlations (Matsumoto et al., 2000). Such observations are even further reinforced by reports of *emotion-specific* decline in ERA correlated with life span development (Laukka & Juslin, 2007; Ruffman, Henry, Livingstone, & Phillips, 2008; West et al., 2012; also see Chapter 9).

More generally, such observations can also be related to the comparably low internal consistency reported for measures of ERA. Hall (2001) suggested that low internal consistency in ERA assessments can be explained by the fact that various items in the tests measure partly independent abilities. Hall (2001) argued that ERA might be considered as a composite competence, composed of several largely independent (or only mildly correlated) elementary abilities. The overall assessment of ERA would in this case not be considered as a latent factor that should be equally reflected in all of its facets. Rather some persons might obtain very high ERA scores, provided that they succeed very well in all (or in most) facets of the more general ability, yet other people might obtain more average scores because they obtain high scores on some facets and lower scores on others. For example, one person might hypothetically recognize anger very accurately, but not sadness, while another person might recognize sadness accurately, but not anger; neither person would obtain very high general ERA scores. Bänziger et al. (2009) have, for example, reported that despite comparably large correlations across different channel scores, the participants in their sample could be clustered in three groups based on their answers on the MERT (which measures accuracy in four channel combinations). One group of participants in this study was characterized by comparably high scores for vocal ERA (scores for vocal ERA are generally lower than scores that are based on stimuli that involve facial expressions in the MERT, but some participants obtain larger vocal than facial based scores). A more recent study by Schlegel, Grandjean, and Scherer (2013a) confirms that ERA needs to be considered as a multifaceted skill, although the results in this study showed larger differentiation for different emotions than for different channels.

Altogether, the available data strongly point to the need to develop more differentiated measures of ERA in order to be able to better assess the possible existence of multiple facets of ERA. Separate assessments of postulated independent facets will furthermore allow researchers to more systematically examine the extent of correlation between the postulated facets and to test the validity of a postulated general skill (latent factor) in ERA. Chapter 18 addresses these questions further.

Relationships with other interpersonal or social competencies

The issue of the construct of ERA (is it a unitary competence or is it composed of several partly independent abilities) can be further expanded also to the discussion related to the potential association of ERA with further social or interpersonal skills. Does ERA reflect a more broad competence in interpersonal or social perception? Is there, for instance, a connection between explicit emotion recognition (ERA) and empathic behavior?[12] (Chapter 17 in the present volume addresses this specific question in more detail, and Chapter 18 takes up the question of the relations among ERA tests and between ERA tests and other tests of interpersonal accuracy.)

This line of questioning has been linked during the past 15–20 years to the development of research on EI. Some authors made strong claims regarding the existence of a unitary skill (or a set of correlated skills) in emotional information processing (including ERA) and the management or regulation of emotional reactions (including expressivity and expressive regulation). The construct of EI has rapidly expanded to include even further interpersonal competencies and also personality traits/characteristics (e.g., empathy). This expanded conceptualization of EI has been designated as trait-model EI, while the original conceptualization (by Mayer and Salovey) is designated as ability-model EI. The controversies that have resulted from the competing conceptualizations of EI fall well beyond the scope of the current chapter. The reader is referred to an account written by Mayer, Salovey, and Caruso (2008) for further developments. It should be noted, however, that few empirical studies of EI actually have measured ERA (even within the ability-model EI).

Conversely, there are multiple studies showing correlates of ERA with a number of social skills and also with other individual differences. One meta-analysis by Hall and collaborators (Hall et al., 2009) reports small to moderate correlations between "interpersonal sensitivity" (which is to

[12] Venturing into those grounds is complex given the general lack of definitional agreement for constructs such as empathy. The term empathy is used here theoretically to refer to the capacity to understand what other people think or feel and also to the elicitation of a socially appropriate affective response. Empathy is thereby defined as involving implicit or explicit recognition accuracy (for emotions and for further cognitive or affective states signaled by another person) and, importantly, an affective reaction (toward this person). This affective reaction will itself incite the empathic observer to engage in prosocial behaviors (see Batson, 2009, for a more elaborate and informed discussion).

a large extent assessed with the instruments described earlier for the assessment of ERA and some additional measures) and a range of social competencies (e.g., "workplace effectiveness," or "emotional competence"; see Hall et al., 2009, Table 4). The same meta-analysis also reviewed correlates with other individual differences (e.g., empathy,[13] personality traits, locus of control, self-esteem; Hall et al., 2009, Table 3) and reported significant, albeit small, correlations with interpersonal sensitivity for many constructs, more specifically positive correlations for socially desirable traits (e.g., empathy, openness, extraversion, or tolerance) and negative correlations for socially undesirable traits (e.g., neuroticism or shyness).

In a recent contribution, Schlegel, Grandjean, and Scherer (2013b) assessed several individual characteristics that have been described as constitutive of, or at least related with, EI. They termed those characteristics "social and emotional effectiveness constructs" and described an analysis of the correlation patterns between those measures. They found that the scores on ERA tests were grouped on one statistical factor that they termed "emotional ability." They identified three additional factors that were interpreted as corresponding to "expressivity," "sensitivity," and "self-control."[14] The "emotional ability" factor was correlated with the "sensitivity" factor but not the two other factors. There were no correlations between "emotional ability" and Big-Five personality traits, with the exception of a small positive correlation with Neuroticism.

There are multiple prior findings indicating that ERA is weakly to moderately correlated with a variety of social competencies and also with further individual differences. An attempt to formulate an integrative model of emotional competence was made in the framework of EI theorization (see Mayer et al., 2008). However, models of EI have themselves evolved into divergent accounts that do not contribute to clarify the constructs that might constitute a general overarching emotional competence.

[13] Empathy in this case groups measures obtained with eight different instruments and ten different scales. Those measures are in fact relatively heterogeneous, but they revolve around attending to the feelings and needs of others and own affiliative needs. All measures are self-report measures.

[14] The "expressivity factor" groups all measures related to "proactive, expressive, and confident behaviors and traits." The "sensitivity factor" groups measures characterized by self-reported "sensitivity and supportive behavior towards others." This factor includes for instance scales that measure Empathic Concern (from the Interpersonal Reactivity Index, Davis, 1983), or Empathic Support (from the Interpersonal Competence Questionnaire, Buhrmester, Furman, Wittenberg, & Reis, 1988). The "self-control factor" groups the "measures related to emotion regulation skills." Quoted expressions are from Schlegel et al. (2013b, pp. 250–251). All measures except for scores on ERA test are self-reports.

A brief outline of some central processes involved in ERA

In order to thoroughly account for the complex processes involved in emotion recognition, several chapters would be warranted. In the following, the main aspects are briefly outlined in order to draw attention to the most essential aspects and highlight some central issues.

Accurate recognition and automatic processing

Studies of ERA involve explicit assessments which presuppose that conscious judgments are considered. Emotional recognition can, however, also take place automatically and unconsciously. This has been assessed by multiple studies showing that exposure to emotional expressions influence people's reactions and behaviors even when they do not consciously attend to them, are unaware that they have seen or heard them and, consequently, do not formulate any conscious judgments about them (see, e.g., Winkielman, Berridge, & Wilbarger, 2005; or Pessoa, 2005). More generally, there is evidence that different perceptual processes are engaged when emotional expressions are processed explicitly (as in the conventional ERA tests described in this chapter) or implicitly. Much research in this field has been carried out using emotional faces (e.g., Palermo & Rhodes, 2007) and occasionally emotional voices (e.g., Frühholz, Ceravolo, & Grandjean, 2012).

For ERA, this might imply that in situations where explicit recognition is impaired, implicit recognition might still take place. Furthermore, it is also possible that accuracy in explicitly labeling emotional expressions does not translate into similar skills when information on emotions signaled in other people's behavior needs to be processed implicitly (automatically and unconsciously). The distinction between explicit and implicit recognition also has implications for expected correlates of ERA. Is it, for example, necessary to explicitly recognize and label an emotional reaction in order to respond with socially appropriate behaviors? Observations of small children, who are able to display "empathic" reactions long before they can verbally categorize emotions (Eisenberg, Fabes, & Spinrad, 2006), and also other animal species (De Waal, 2008) suggest otherwise. Further evidence on processing of emotional information shows, however, that it is likely, under normal circumstances, that unconscious and automatic processing of emotional signals would tend to (re)direct attention to the emotional signals, which in turn increases the likelihood that conscious (and potentially accurate) judgments may be formed on the emotional cues present in the environment.

The role of mimicry and the mirror neuron system in emotion recognition

Lipps (1907) proposed a model for empathy that is grounded on "emotional contagion." According to this model, people automatically mimic the facial expressions of other persons in social interactions. The motor activation resulting from this automatic imitation allows them to experience (via proprioceptive feedback mechanisms) the mental/psychological state of their interaction partners. This model has received experimental support in many studies (e.g., Dimberg, Thunberg, & Elmehed, 2000). Blairy, Herrera, and Hess (1999) provided a detailed account of the model and of the studies supporting it. Further evidence has also been provided more recently, showing that voluntary suppression of facial mimicry has an adverse effect on ERA (see, e.g., Ponari, Conson, D'Amico, Grossi, & Trojano, 2012; Schneider, Hempel, & Lynch, 2013).

Hess, Philippot, and Blairy (1998) reported, however, that they found no automatic mimicry of facial emotional expressions when the task of the research participants was to rate the spontaneity of the expression, whereas mimicry was present when the task was to recognize emotions. This seems to indicate that mimicry does not follow automatically from exposure to expressive faces, but rather follows from the voluntary engagement in emotion recognition (and/or the allocation of attention to this task). Blairy et al. (1999) found evidence for mimicry in their own studies, but concluded that their data did not support the notion that mimicry mediates recognition accuracy. Moody, McIntosh, Mann, and Weisser (2007) showed that mimicry can be affected by the manipulation of the emotional state of the observer. They conclude from their observations that mimicry is not entirely automatic, but depends instead on the momentary state of the person who observes the expression.

In the past 15 years, mimicry and embodiment have been the subject of renewed interest in the framework of the development of neurological studies focusing on mirror neurons. Several researchers have investigated and described mirror neurons as an essential component of the neurological substrates that account for the link between perception (or active observation) of motor activities – including facial expressions – in others and activation of motor/expressive behavior in the observer (see, e.g., Gallese, Keysers, & Rizzolatti, 2004, or Hadjikhani, 2007). Deficits in the social communication sphere have been related to impairments in emotion recognition and according to some models might be accounted for by deficits in automatic mimicry and in the "mirror neuron system" (Dapretto et al., 2005). Some researchers found that people with Autism spectrum disorders (ASD) did not spontaneously mimic facial emotional expressions but could mimic them when required to do so (McIntosh, Reichmann-Decker, Winkielman, & Wilbarger, 2006). Studies on those issues have multiplied in the past decade. Hamilton (2013) published a review in which she estimates, however, that there is, altogether, not much support for an *overall* deficit in the "mirror neuron system" in ASD.

Such studies nevertheless emphasize that embodiment and mimicry most probably are involved in emotion recognition. To what extent mimicry is necessary for ERA remains, however, a question that is still largely debated. The available evidence suggests that mimicry does contribute to ERA but that it is not fully automatic and rather seems to follow from the fact that people engage in explicit processing of emotional expressions.

Hormonal processes (oxytocin)

Hormonal processes have been shown to affect emotion recognition. In recent years, several studies have described the effect of oxytocin on a number of psychological processes (e.g., on trust and decision making). Oxytocin is thought to be involved in the regulation of interpersonal interactions. It is massively released in women following childbirth and has been associated with mother–child bonding and, more generally, with affiliative behaviors. Findings showing that oxytocin administration (with nasal sprays) increased ERA (see Domes et al., 2010; Leknes et al., 2013; Schulze et al., 2011) raised speculations as to the processes involved in this effect. Some researchers have stated that the administration of oxytocin was related to an increased focus (gaze orientation) on the eye region of human faces (Guastella, Mitchell, & Dadds, 2008), and that this effect might account for the increase in ERA. However, other studies failed to replicate an increased gaze fixation to the eye region of emotionally expressive faces (Domes et al., 2010; Lischke et al., 2012) while replicating the increase of ERA following intranasal oxytocin administration.

Several studies showed furthermore differential effects of oxytocin administration in women and in men (Domes et al., 2010). Questions have been raised also regarding the emotions that account for the observed effects. Di Simplicio, Massey-Chase, Cowen, and Harmer (2009) raised the possibility that oxytocin enhances recognition accuracy specifically for positive emotions.

The study of hormonal (oxytocin) influences on ERA provides an interesting illustration for the need of differentiated measures, which allow to assess several facets of ERA independently (e.g., independent assessment of ERA for positive and negative emotions).

Ecology and the influence of verbal and social context

Classical studies of ERA are essentially concerned with accurate emotion recognition based on *nonverbal cues*. The traditional ERA tests strive to suppress verbal information and contextual (situational) information from the tests. The goal as mentioned in the beginning of this chapter is to assess ERA alone, neutralizing other potential influences. The downside of this procedure is that those tests have fairly low ecological validity. In daily reality, emotional expressions are usually accompanied by multiple other cues (verbal or contextual) that provide valuable information for interpreting the emotions of others.

Social scripts bias the interpretation of nonverbal expressions. This has been shown experimentally, for instance, by Carroll and Russell (1996). Hess and Hareli (2015) provide a thorough account of the effect of social contextual cues on emotion recognition. Among other aspects, they describe how contextual social information, and also personality characteristics, can be inferred from emotional expressions (using "reversed appraisals" – a model that is also described in the present chapter, in Section "Emotion Constructs in ERA"). They furthermore report evidence showing that many social factors influence emotion recognition (social status, group membership, roles, norms, etc.).

Further studies showed that nonverbal expressions can be interpreted differentially depending on the verbal information associated with the nonverbal displays (e.g., Knudsen & Muzekari, 1983). O'Sullivan, Ekman, Friesen, and Scherer (1985) showed that verbal and nonverbal communication (facial, vocal, postural, gestural) all contribute to the perception of the speakers and influence the judgments of speaker characteristics in complex interactions. Another example can be found in Ickes' empathic accuracy method (Ickes, 2001; Chapter 3). Several studies indicate that when research participants are requested to label the covert thoughts and feelings of other people in videotaped interactions, accuracy depends more on the verbal content of the recorded interactions than on the nonverbal signals that are present in the interactions (Gesn & Ickes, 1999; Hall & Schmid Mast, 2007; Zaki, Bolger, & Ochsner, 2009).

Conclusions

This chapter examines the fundamental issues related to the judgment of emotions. Emotion recognition involves complex explicit judgments and also unconscious, automatic processes. A great number of studies have been conducted on emotion recognition, with many different research questions and/or target groups. The most striking observation, when coarsely reviewing this field, concerns the heterogeneity of the assessment methods used. A very large number of studies concentrated initially on just a few emotions and on the recognition of stereotyped facial displays in photographs (as noted also by others, e.g., Hall et al., 2008). Other approaches have, however, begun to expand during the last 10–20 years. This expansion of questions and of assessment methods results both in very rich and interesting confrontations between perspectives and sometimes conflicting results. On the downside, it results also in an overall noisy field, where many results cannot be straightforwardly compared, due to their methodological heterogeneity. For the future, many issues need to be addressed. One very essential issue concerns the role of the various communication channels and their interactions or integration. The definition of the emotion construct is also an issue that needs to be addressed. Discriminating between happiness and fear in prototypical facial expressions is most probably not the same task as recognizing subtle nuances of emotional reactions in complex social interactions in everyday life. One interesting aspect in this respect would be to invest more efforts into investigating the extent to which laboratory assessments of ERA – based for example on prototypical facial displays – are predictive of success in interpersonal communication in more naturalistic social settings.

References

Bänziger, T. (2014). Measuring emotion recognition ability. In A. Michalos (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 3934–3941). Dordrecht: Springer Reference.

Bänziger, T., Grandjean, D., & Scherer, K. R. (2009). Emotion recognition from expressions in face, voice, and body: The Multimodal Emotion Recognition Test (MERT). *Emotion*, 9, 691–704.

Bänziger, T., & Scherer, K. R. (2010). Introducing the Geneva Multimodal Emotion Portrayal (GEMEP) corpus. In K. R. Scherer, T. Bänziger, & E. B. Roesch (Eds.), *Blueprint for affective computing: A sourcebook* (pp. 271–294). Oxford: Oxford University Press.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 42, 241–251.

Barrett, L. F., Wilson-Mendenhall, C. D., & Barsalou, L. W. (2014). The Conceptual Act Theory: A road map. In L. F. Barrett & J. A. Russell (Eds.), *The psychological construction of emotion* (pp. 83–110). New York: Guilford.

Batson, C. D. (2009). These things called empathy: Eight related but distinct phenomena. In J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 3–15). Cambridge, MA: MIT Press.

Blairy, S., Herrera, P., & Hess, U. (1999). Mimicry and the judgment of emotional facial expressions. *Journal of Nonverbal Behavior*, 23, 5–41.

Brosch, T., & Grandjean, D. (2013). Cross-modal modulation of spatial attention by emotion. In P. Belin, S. Campanella, & T. Ethofer (Eds.), *Integrating face and voice in person perception* (pp. 207–224). New York: Springer.

Buhrmester, D., Furman, W., Wittenberg, M. T., & Reis, H. T. (1988). Five domains of interpersonal competence in peer relationships. *Journal of Personality and Social Psychology*, 55, 991–1008.

Carroll J. M., & Russell, J. A. (1996). Do facial expressions signal specific emotions? Judging emotion from the face in context. *Journal of Personality and Social Psychology*. 70, 205–218.

Chen D., & Haviland-Jones, J. (2000). Human olfactory communication of emotion. *Perceptual and Motor Skills*, 91, 771–781.

Dapretto, M., Davies, M. S., Pfeifer, J. H., Scott, A. A., Sigman, M., Bookheimer, S. Y., & Iacoboni, M. (2005). Understanding emotions in others: Mirror neuron dysfunction in children with autism spectrum disorders. *Nature Neuroscience*, 9, 28–30.

Darwin, C. (1872/1998). *The expression of the emotions in man and animals. Introduction, afterword and commentaries by Paul Ekman*. Oxford: Oxford University Press.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44, 113–126.

De Gelder, B., Stienen, B. M. C., & Van den Stock, J. (2013). Emotions by ear and by eye. In P. Belin, S. Campanella, & T. Ethofer (Eds.), *Integrating face and voice in person perception* (pp. 253–270). New York: Springer.

De Gelder, B., & Van den Stock, J. (2011). Real faces, real emotions: Perceiving facial expressions in naturalistic contexts of voices, bodies and scenes. In G. Rhodes, A. Calder, M. Johnson, & J. V. Haxby (Eds.), *The Oxford handbook of face perception* (pp. 535–550). Oxford, UK: Oxford University Press.

De Waal, F. B. (2008). Putting the altruism back into altruism: the evolution of empathy. *Annual Review of Psychology*, 59, 279–300.

Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychological Science*, 11, 86–89.

Di Simplicio, M., Massey-Chase, R., Cowen, P. J., & Harmer, C. J. (2009). Oxytocin enhances processing of positive versus negative emotional information in healthy male volunteers. *Journal of Psychopharmacology*, 23, 241–248.

Domes, G., Lischke, A., Berger, C., Grossmann, A., Hauenstein, K., Heinrichs, M., & Herpertz, S. C. (2010). Effects of intranasal oxytocin on emotional face processing in women. *Psychoneuroendocrinology*, 35, 83–93.

Du, S., Tao, Y., & Martinez, A. M. (2014). Compound facial expressions of emotion. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 1454–1462.

Eisenberg, N., Fabes, R.A., & Spinrad, T. L. (2006). Prosocial development. In W. Damon, R. M. Lerner, & N. Eisenberg (Eds.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (pp. 646–718). New York: Wiley.

Ekman, P. (2003a). Darwin, deception, and facial expression. In P. Ekman, J. J. Campos, R. J. Davidson, & F. B. M. de Waal (Eds.), *Emotions inside out: 130 years after Darwin's The expression of the emotions in man and animals* (pp. 205–221). New York: New York Academy of Sciences.

Ekman, P. (2003b). Sixteen enjoyable emotions. *Emotion Researcher*, 18, 6–7.

Ekman, P. (2003c). *METT: Micro expression training tool*. CD-ROM. Oakland.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17, 124–129.

Ekman, P., & Friesen, W. V. (1976). *Pictures of facial affect*. Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., & Keltner, D. (1997). Universal facial expressions of emotion: An old controversy and new findings. In U. C. Segerstråle & P. Molnár (Eds.), *Nonverbal communication: Where nature meets culture* (pp. 27–46). Mahwah, NJ: Lawrence Erlbaum Associates.

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, 128, 203–235.

Elfenbein, H. A., & Ambady, N. (2003). Universals and cultural differences in recognizing emotions. *Current Directions in Psychological Science*, 12, 159–164.

Frühholz, S., Ceravolo, L., & Grandjean, D. (2012). Specific brain networks during explicit and implicit decoding of emotional prosody. *Cerebral Cortex*, 22, 1107–1117.

Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences*, 8, 396–403.

Gendron, M., Lindquist, K. A., Barsalou, L., & Barrett, L. F. (2012). Emotion words shape emotion percepts. *Emotion*, 12, 314–325.

Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F. (2014a). Perceptions of emotion from facial expressions are not culturally universal: Evidence from a remote culture. *Emotion*, 14, 251–262.

Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F (2014b). Cultural relativity in perceiving emotion from vocalizations. *Psychological Science*, 25, 911–920.

Gesn, P. R., & Ickes, W. (1999). The development of meaning contexts for empathic accuracy: Channel and sequence effects. *Journal of Personality and Social Psychology*, 77, 746–761.

Goffman, E. (1959). *The presentation of self in everyday life*. University of Edinburgh Social Sciences Research Centre.

Golan, O., & Baron-Cohen, S. (2006). Systemizing empathy: Teaching adults with Asperger syndrome or high functioning autism to recognize complex emotions using interactive media. *Development and Psychopathology*, 18, 591–617.

Golan, O., Baron-Cohen, S., & Hill, J. (2006). The Cambridge Mindreading (CAM) Face-Voice Battery: Testing complex emotion recognition in adults with and without Asperger syndrome. *Journal of Autism and Developmental Disorders*, 36, 169–183.

Guastella, A. J., Mitchell, P. B., & Dadds, M. R. (2008). Oxytocin increases gaze to the eye region of human faces. *Biological Psychiatry*, 63, 3–5.

Hadjikhani, N. (2007). Mirror Neuron System and Autism. In P. C. Carlisle (Ed.), *Progress in Autism Research* (pp. 151–166). New York: Nova Science Publishers.

Hall, J. A. (2001). The PONS test and the psychometric approach to measuring interpersonal sensitivity. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement* (pp. 143–161). Mahwah, NJ: Lawrence Erlbaum Associates.

Hall, J. A., Andrzejewski, S. A., Murphy, N. A., Schmid Mast, M., & Feinstein, B. A. (2008). Accuracy of judging others' traits and states: Comparing mean levels across tests. *Journal of Research in Personality*, 42, 1476–1489.

Hall, J. A., Andrzejewski, S. A., & Yopchick, J. E. (2009). Psychosocial correlates of interpersonal sensitivity: A meta-analysis. *Journal of Nonverbal Behavior*, 33, 149–180.

Hall, J. A., Bernieri, F. J., & Carney, D. R. (2005). Nonverbal behavior and interpersonal sensitivity. In J. A. Harrigan, R. Rosenthal, & K. R. Scherer (Eds.), *The new handbook of methods in nonverbal behavior research* (pp. 237–281). Oxford: Oxford University Press.

Hall, J. A., & Schmid Mast, M. (2007). Sources of accuracy in the empathic accuracy paradigm. *Emotion*, 7, 438–446.

Hamilton, A. (2013). Reflecting on the mirror neuron system in autism: A systematic review of current theories. *Developmental Cognitive Neuroscience*, 3, 91–105.

Hertenstein, M. J., Holmes R., McCullough M., & Keltner D. (2009). The communication of emotion via touch. *Emotion*, 9, 566–573.

Hess, U., & Hareli, S. (2015). The role of social context for the interpretation of emotional facial expressions. In M. K. Mandal & A. Awasthi (Eds.), *Understanding facial expressions in communication* (pp. 119–141). Springer India.

Hess, U., Philippot, P., & Blairy, S. (1998). Facial reaction to emotional facial expressions: Affect or cognition? *Cognition and Emotion*, 12, 509–532.

Ickes, W. (2001). Measuring empathic accuracy. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement* (pp. 219–241). Mahwah, NJ: Lawrence Erlbaum Associates.

Jack, R. E., Garrod, O. G. B., & Schyns, P. G. (2014). Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. *Current Biology*, 24, 187–192.

Kang, S. M. (2012). Individual differences in recognizing spontaneous emotional expressions: Their implications for positive interpersonal relationships. *Psychology*, 3, 1183–1188.

Knudsen, H. R., & Muzekari, L. H. (1983). The effects of verbal statements of context on facial expressions of emotion. *Journal of Nonverbal Behavior*, 7, 202–212.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D.H.J., Hawk, S.T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition & Emotion*, 24, 1377–1388.

Laukka, P. (2005). Categorical perception of vocal emotion expressions. *Emotion*, 5, 277–295.

Laukka, P., Audibert, N., & Aubergé, V. (2012). Exploring the determinants of the graded structure of vocal emotion expressions. *Cognition and Emotion*, 26, 710–719.

Laukka, P., & Elfenbein, H. A. (2012). Emotion appraisal dimensions can be inferred from vocal expressions. *Social Psychological and Personality Science*, 3, 529–536.

Laukka, P., & Juslin, P. N. (2007). Similar patterns of age-related differences in emotion recognition from speech and music. *Motivation and Emotion*, 31, 182–191.

Leknes, S., Wessberg, J., Ellingsen, D. M., Chelnokova, O., Olausson, H., & Laeng, B. (2013). Oxytocin enhances pupil dilation and sensitivity to 'hidden' emotional expressions. *Social Cognitive & Affective Neuroscience*, 8, 741–749.

Lindquist, K. A., Barrett, L. F., Bliss-Moreau, E., & Russell, J. A. (2006). Language and the perception of emotion. *Emotion*, 6, 125–138.

Lipps, T. (1907). Das Wissen von fremden Ichen. In T. Lipps (Ed.), *Psychologische Untersuchungen Band 1* (pp. 694–722). Leipzig: Engelmann.

Lischke, A., Berger, C., Prehn, K., Heinrichs, M., Herpertz, S. C., & Domes, G. (2012). Intranasal oxytocin enhances emotion recognition from dynamic facial expressions and leaves eye-gaze unaffected. *Psychoneuroendocrinology*, 37, 475–481.

Lundqvist, D., Flykt, A., & Öhman, A. (1998). *The Karolinska Directed Emotional Faces – KDEF*, CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet.

Martin, O., Kotsia, I., Macq, B., & Pitas, I. (2006). The enterface'05 Audio-Visual Emotion Database. Proc. IEEE Workshop on Multimedia Database Management, Atlanta.

Massaro, D. W., & Egan, P. B. (1996). Perceiving affect from the voice and the face. *Psychonomic Bulletin Review*, 3, 215–221.

Matsumoto, D., & Ekman, P. (1989). American-Japanese cultural differences in intensity ratings of facial expressions of emotion. *Motivation and Emotion*, 13, 143–157.

Matsumoto, D., & Hwang, H. S. (2011). Evidence for training the ability to read microexpressions of emotion. *Motivation and Emotion*, 35, 181–191.

Matsumoto, D., LeRoux, J., Wilson-Cohn, C., Raroque, J., Kooken, K., Ekman, P., Yrizarry, N., Loewinger, S., Uchida, H., Yee, A., Amo, L., & Goh, A. (2000). A new test to measure emotion recognition ability: Matsumoto and Ekman's Japanese and Caucasian Brief Affect Recognition Test (JACBART). *Journal of Nonverbal Behavior*, 24, 179–209.

Mayer, J. D., Salovey, P., & Caruso, D. R. (2008). Emotional intelligence: New ability or eclectic traits? *American Psychologist*, 63, 503–517.

Mayer, J. D., Salovey, P., Caruso, D. R., & Sitarenios, G. (2003). Measuring emotional intelligence with the MSCEIT V2.0. *Emotion*, 3, 97–105.

McIntosh, D. N., Reichmann-Decker, A., Winkielman, P., & Wilbarger, J. L. (2006). When the social mirror breaks: Deficits in automatic, but not voluntary, mimicry of emotional facial expressions in autism. *Developmental Science*, 9, 295–302.

Moody, E. J., McIntosh, D. N., Mann, L. J., & Weisser, K. R. (2007). More than mere mimicry? The influence of emotion on rapid facial reactions to faces. *Emotion*, 7, 447–457.

Nowicki, S. (2006). *Manual for the receptive tests of the Diagnostic Analysis of Nonverbal Accuracy 2 (DANVA2)*. Unpublished manual.

Nowicki, S., Jr., & Duke, M. P. (1994). Individual differences in the nonverbal communication of affect: The Diagnostic Analysis of Nonverbal Accuracy scale. *Journal of Nonverbal Behavior*, 18, 9–35.

O'Sullivan, M., Ekman, P., Friesen, W., & Scherer, K. R. (1985). What you say and how you say it: The contribution of speech content and voice quality to judgments of others. *Journal of Personality and Social Psychology*, 48, 54–62.

Pelowski, M., & Akiba, F. (2011). A model of art perception, evaluation and emotion in transformative aesthetic experience. *New Ideas in Psychology*, 29, 80–97.

Pessoa, L. (2005). To what extent are emotional visual stimuli processed without attention and awareness? *Current Opinion in Neurobiology*, 15, 188–196.

Ponari, M., Conson, M., D'Amico, N. P., Grossi, D., & Trojano, L. (2012). Mapping correspondence between facial mimicry and emotion recognition in healthy subjects. *Emotion*, 12, 1398–1403.

Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., & Archer, D. (1979). *Sensitivity to nonverbal communication: The PONS test*. Baltimore: The Johns Hopkins University Press.

Rosenthal, R., & Rubin, D. B. (1989). Effect size estimation for one-sample multiple-choice-type data: Design, analysis, and meta-analysis. *Psychological Bulletin*, 106, 332–337.

Ruffman, T., Henry, J. D., Livingstone, V., & Phillips, L. H. (2008). A meta-analytic review of emotion recognition and aging: Implications for neuropsychological models of aging. *Neuroscience & Biobehavioral Reviews*, 32, 863–881.

Russell, J. A. (1994). Is there universal recognition of emotion from facial expressions? A review of the cross-cultural studies. *Psychological Bulletin*, 115, 102–141.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, 110, 145–172.

Russell, J. A. (2012). From a psychological constructionist perspective. In P. Zachar & R. Ellis (Eds.), *Categorical versus dimensional models of affect: A seminar on the theories of Panksepp and Russell* (pp. 79–118). Amsterdam: John Benjamins Publishing.

Russell, T. A., Chu, E., & Phillips, M. L. (2006). A pilot study to investigate the effectiveness of emotion recognition remediation in schizophrenia using the micro-expression training tool. *British Journal of Clinical Psychology*, 45, 579–583.

Sander, D., Grandjean, D., Kaiser, S., Wehrle, T., & Scherer K. R. (2007). Interaction effects of perceived gaze direction and dynamic facial expression: Evidence for appraisal theories of emotion. *European Journal of Cognitive Psychology*, 19, 470–480.

Scherer, K. R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition and Emotion*, 23, 1307–1351.

Scherer, K. R. (2013). Vocal markers of emotion: Comparing induction and acting elicitation. *Computer Speech and Language*, 27, 40–58.

Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology*, 32, 76–92.

Scherer, K. R., & Bänziger, T. (2010). On the use of actor portrayals in research on emotional expression. In K. R. Scherer, T. Bänziger, & E. B. Roesch (Eds.), *Blueprint for affective computing: A sourcebook* (pp. 166–176). Oxford, UK: Oxford University Press.

Scherer, K. R., Clark-Polner, E., & Mortillaro, M. (2011). In the eye of the beholder? Universality and cultural specificity in the expression and perception of emotion. *International Journal of Psychology*, 46, 401–435.

Scherer, K. R., & Ellgring, H. (2007). Multimodal expression of emotion: Affect programs or componential appraisal patterns? *Emotion*, 7, 158–171.

Scherer, K. R., & Grandjean, D. (2008) Inferences from facial expressions of emotion have many facets. *Cognition and Emotion*, 22, 789–801.

Scherer, K. R., & Scherer, U. (2011). Assessing the ability to recognize facial and vocal expressions of emotion: Construction and validation of the Emotion Recognition Index (ERI). *Journal of Nonverbal Behavior*, 35, 305–326.

Schlegel, K., Grandjean, D., & Scherer, K. R. (2013a). Emotion recognition: Unidimensional ability or a set of modality- and emotion-specific skills? *Personality and Individual Differences*, 53, 16–21.

Schlegel, K., Grandjean, D., & Scherer, K. R. (2013b). Constructs of social and emotional effectiveness: Different labels, same content? *Journal of Research in Personality*, 47, 249–253.

Schlegel, K., Grandjean, D., & Scherer, K. R. (2014). Introducing the Geneva Emotion Recognition Test: An example of Rasch-based test development. *Psychological Assessment*, 26, 666–672.

Schneider, K. G., Hempel, R. J., & Lynch T. R. (2013). That "poker face" just might lose you the game! The impact of expressive suppression and mimicry on sensitivity to facial expressions of emotion. *Emotion*, 13, 852–866.

Schulze, L., Lischke, A., Greif, J., Herpertz, S. C., Heinrichs, M., & Domes, G. (2011). Oxytocin increases recognition of masked emotional faces. *Psychoneuroendocrinology*, 36, 1378–1382.

Shackman, J. E., & Pollak, S. D. (2005). Experiential influences on multimodal perception of emotion. *Child Development*, 76, 1116–1126.

Silvia, P. J. (2005). Emotional responses to art: From collation and arousal to cognition and emotion. *Review of General Psychology*, 9, 342–357.

Suzuki, A., Hoshino, T., & Shigemasu, K. (2010). Happiness is unique: A latent structure of emotion recognition traits revealed by statistical model comparison. *Personality and Individual Differences*, 48, 196–201.

West, J. T., Horning, S. M., Klebe, K. J., Foster, S. M., Cornwell, R. E., Perrett, D., Burt, D. M., & Davis, H. P. (2012). Age effects on emotion recognition in facial displays: From 20 to 89 years of age. *Experimental Aging Research*, 38, 146–168.

Winkielman, P., Berridge, K. C., & Wilbarger, J. L. (2005). Unconscious affective reactions to masked happy versus angry faces influence consumption behavior and judgments of value. *Personality and Social Psychology Bulletin*, 31, 121–135.

Young, A. W., Rowland, D., Calder A. J., Etcoff, N. L., Seth, A., & Perrett, D. I. (1997). Facial expression megamix: Tests of dimensional and category accounts of emotion recognition. *Cognition*, 63, 271–313.

Zaki, J., Bolger, N., & Ochsner, K. (2009). Unpacking the informational bases of empathic accuracy. *Emotion*, 9, 478–487.

3 Empathic accuracy
Judging thoughts and feelings

William Ickes

Abstract
Key issues in the study of *empathic accuracy* (accuracy in inferring the specific content of other people's thoughts and feelings) are explored through the answers to ten questions. These questions concern (1) how empathic accuracy is measured and studied; (2) how accurate perceivers are, relative to chance and across different types of relationships; (3) which information channels are the most important; (4) why "female motivation" may be more important than "female intuition"; (5) the two most obvious and reliable individual-difference predictors of empathic accuracy; (6) the importance of "target readability" in empathic accuracy; (7) the importance of attention and (8) motivation in empathic accuracy; (9) when empathic accuracy hurts, instead of helps, close relationships; and (10) empathic accuracy's linkage to various clinical problems and populations.

In this chapter, I will ask and answer ten questions about empathic accuracy – one of the many forms of interpersonal accuracy that are discussed in this book. By the end of this question-and-answer session, the reader should have a good sense of how the research on empathic accuracy is relevant to the overall theme of this volume: the social psychology of perceiving other people accurately.

The term *empathic accuracy* refers to one's overall accuracy in inferring the specific content of other people's successive thoughts and feelings. Its theoretical precedent is Carl Rogers's (1957) term *accurate empathy*, which was similarly defined. According to Rogers, this construct refers to the extent to which one can enter into the flow of other people's subjective experience and accurately track the changing contents of their successive thoughts and feelings. Empathic accuracy is therefore a broader construct than emotion detection (i.e., accuracy in identifying

Acknowledgments: The author thanks Anna Park and Vivian Ta for their comments on a previous version of this chapter.

other people's emotional states), because it concerns the accurate identification of thoughts as well as of feelings.

How can we measure accuracy in judging other people's thoughts and feelings?

During the past 25 years of research on empathic accuracy, the construct has been measured and studied using two primary research paradigms: the *unstructured dyadic interaction paradigm* and the *standard stimulus paradigm* (Ickes, 2001; Schmid Mast & Ickes, 2007).[1]

The unstructured dyadic interaction paradigm. In the unstructured dyadic interaction paradigm, two individuals (who may or may not already be acquainted) are escorted to a laboratory "waiting room" by an experimenter who asks them to take a seat on a couch. Consistent with a cover story, the experimenter then leaves them alone together on the pretext of having to retrieve something that is needed for the experiment. At this point, the unstructured dyadic interaction occurs: during the next six minutes, the two participants are unobtrusively video and audio recorded. At the end of this time, the experimenter returns and reveals to the participants that their interaction has been taped.

If both participants then give their written consent to have the video and audio recordings used as data sources, they proceed to the next phase of the study. In this phase, they are seated in separate cubicles, where they each view a copy of the tape that was made while they were interacting. Their task while watching the tape is to stop it whenever they remember having had a thought or a feeling at a particular point during the interaction. Using a supply of *thought/feeling recording forms*, they write down the specific times when their thoughts or feelings occurred, along with the content of each reported thought or feeling.

When both participants have completed this task, the thought/feeling inference phase of the procedure begins. The participants are asked to infer the specific content of each of their partner's thoughts and feelings, using a supply of *thought/feeling inference forms*. In this phase of the procedure, the experimenter (or an assistant) pauses the tape for each participant at the specific points at which the interaction partner reported having had

[1] Other variations have also emerged in recent years. They include the *standard interview paradigm* (Dugosh, Cheng, & Park, 2011); an *emotion rating paradigm* in which perceivers continuously track the changing emotions of the target persons (Levenson & Ruef, 1992, Zaki, Bolger, & Ochsner, 2008); and a *variant of the standard stimulus paradigm* in which perceivers must select the target persons' actual thought or feeling from a set of multiple-choice alternatives (Hall, Ship, et al., 2014). They also include an *interaction diary paradigm* in which the relationship partners report their own feelings and infer their partner's feelings on a daily basis (Wilhelm & Perrez, 2004; Gadassi, Mor, & Rafaeli, 2011).

each of his or her actual thoughts and feelings, and the two perceivers (again working independently) write down each of their thought/feeling inferences.

Later, the empathic accuracy of each perceiver is judged by a group of five to eight raters. The raters assign "accuracy points" to each inference according to how similar it is in content to the thought or feeling that was actually reported by the interaction partner. For example, a reported thought of "This guy is nice" and an inferred thought of "How long will this experiment last?" would be rated with a 0, as they have essentially different content. In contrast, ratings of 1 or 2 denote "similar, but not the same, content" and "essentially the same content," respectively. By aggregating and transforming these "accuracy points" into a percent-correct measure, an overall empathic accuracy score is then calculated for each participant. Higher scores indicate higher levels of empathic accuracy on a percentage scale that has a potential range of 0 (none of the possible accuracy points) to 100 (all of the possible accuracy points). With as many as 6–8 raters, the interrater reliability of this empathic accuracy measure typically averages about .90; with fewer raters, this value will decline.

The unstructured dyadic interaction paradigm has been used to study interactions between strangers (e.g., Garcia, Stinson, Ickes, Bissonnette, & Briggs, 1991; Ickes, Stinson, Bissonnette, & Garcia, 1990) and between individuals who know each other well (e.g., friends, dating partners, and marriage partners). This paradigm can be used to study asymmetries in empathic accuracy within relationships (Clements, Holtzworth-Munroe, Schweinle, & Ickes, 2007) as well as differences in empathic accuracy across relationships that vary in their degree of intimacy/acquaintanceship (Stinson & Ickes, 1992; Thomas & Fletcher, 2003). On the other hand, the major disadvantage of the paradigm is that each perceiver infers a unique partner's thoughts and feelings, making the inference task different for each perceiver and confounding the perceiver's empathic ability with the partner's "readability." For a statistical solution to these problems (one that involves additional measurement and is a bit too complicated to describe here), see Simpson, Oriña, and Ickes (2003) and Flury, Ickes, and Schweinle (2008).

The standard stimulus paradigm. The standard stimulus paradigm was originally developed to assess empathic accuracy in a clinically relevant research setting (Marangoni, Garcia, Ickes, & Teng, 1995). The paradigm employs a standard stimulus video, usually composed of excerpts from a set of previously videotaped interactions, although they can be taken from any tapes for which the target person's thoughts and feelings are known. This compilation videotape is used as a standard stimulus tape in subsequent studies, in which participants are asked to make thought and feeling inferences at those points on the tape where the target person(s) reported having had a thought or feeling. The inferred thoughts and feelings are compared to the actual reported thoughts and feelings (as described in the previous section), and participants are assigned an overall empathic accuracy score.

The major advantage of the standard stimulus paradigm is the fact that the task is the same for all perceivers. This feature allows empathic accuracy scores to be compared across perceivers and to be correlated with relevant perceiver characteristics. Keeping the task the same for all perceivers also obviates the problem of confounding target expressivity with perceiver perceptivity. Researchers can therefore meaningfully assess individual differences in empathic ability and then use these data to explore issues such as cross-target consistency (Marangoni et al., 1995) and the correlates of empathic ability (Gleason, Jensen-Campbell, & Ickes, 2009).

How accurate are perceivers (relative to chance, and across different types of relationships)?

David Kenny (personal communication) proposed a useful way to assess a perceiver's level of "chance accuracy" in inferring a target person's thoughts and feelings. First, *randomly pair* the target's actual thoughts and feelings with the perceiver's inferences. Second, have a group of raters judge how similar each inferred thought or feeling is to its randomly paired actual thought or feeling. Third, compute the average percentage empathic accuracy score across all thought/feeling inferences for that group of raters.

In the studies in which this random pairing procedure has been used, the average "chance accuracy" score has been about 5% on our scale of 0% to 100% (Ickes, Stinson, et al., 1990; Stinson & Ickes, 1992). Relative to this 5% baseline level of chance accuracy, strangers infer each other's actual (non-randomly paired) thoughts and feelings with an average accuracy score of about 20%; close friends make these inferences with an average accuracy score of about 30%; and married couples achieve average accuracy scores that usually range no higher than 35% (Graham, 1994; Stinson & Ickes, 1992; Verhofstadt, Buysse, & Ickes, 2007; Verhofstadt, Buysse, Ickes, Davis, & Devoldre, 2008).[2] All of these groups perform significantly better than the 5% level that would be expected by chance, but there is obviously a lot of room for improvement between 35% accuracy and the theoretical maximum of 100% accuracy.

[2] See also Thomas and Fletcher (2003), who used slightly different coding rules to assess the average level of empathic accuracy achieved by strangers, friends, and dating partners.

Why are empathic accuracy scores limited to the lower portion of their potential range? In a previous commentary, I speculated that "evolutionary pressures operated over countless generations to eventually optimize the effective range of empathic accuracy in humans so that it was high enough to enable us to deal effectively with others but was not so high that we put our genetic futures at grave risk by weighting everyone else's interests as heavily as our own" (Ickes, 2011, p. 201). It will be an interesting challenge for researchers to see if they can figure out ways to test this speculation.

What information channels contribute the most to empathic accuracy?

When the dramatic wave of interest in nonverbal behavior swept through the fields of anthropology, communication studies, and social and clinical psychology in the 1970s, it led a few enthusiasts to make some overhyped but under-supported claims. Here are two examples. "In some way that we don't understand completely yet, the Nonverbal Act seems to be more important in interpersonal relationships than language itself" (Key, 1975, p. 20). And "nonverbal, not verbal, factors are the major determinants of meaning in the interpersonal context" (Leathers, 1978, p. 4).

With regard to perceivers' empathic accuracy, these claims are now known to be false. In the first study to challenge these claims, Gesn and Ickes (1999) systematically varied the information channels that were available to perceivers: *video plus audio, video plus filtered audio*, or *audio only*. (In the *video plus filtered audio* condition, the target persons' speech was passed through a series of electronic filters, with the result that no words were intelligible but all of the paralinguistic cues were preserved.)

The results of this study revealed that eliminating all of the visual cues in the *audio only* condition reduced the perceivers' empathic accuracy by an average of only 4.1%, when compared to the empathic accuracy of the perceivers in the *video plus audio* condition. This small impairment contrasted dramatically with the 59.2% reduction in empathic accuracy that was observed when the researchers compared the performance of the perceivers in the *video plus filtered audio* condition with that of the perceivers in the *video plus audio* condition. In other words, the perceiver's empathic accuracy was greatly impaired when the target person's words were rendered unintelligible.

Similarly, in a replication-and-extension study using a different set of videotaped target persons and employing an expanded design that also included a *written transcript only* condition, Hall and Schmid Mast (2007) found that removing the visual information from their *audio only* condition reduced the perceivers' empathic accuracy by only a negligible amount (an average reduction of only 2.2% in comparison to their *video plus audio* condition). In contrast, deleting the spoken words from their *video only* condition reduced perceivers' empathic accuracy by an average of 54.2% when compared to the level observed in the *video plus audio* condition.

More recently, Zaki, Bolger, and Ochsner (2009) reported a study in which empathic accuracy was assessed by correlating perceivers' ratings of the target person's affect with the target person's self-rated affect. The perceivers were randomly assigned to one of three information channel conditions: *video plus audio, video only*, or *audio only*. As in the previous studies, the results obtained by Zaki et al. suggested that "auditory, and especially verbal information, is critical to EA [empathic accuracy]" (p. 478).

In my attempt to integrate and summarize these converging findings (Ickes, 2006), I concluded that, in the usual case when all of the information channels are available, the target's verbal information (i.e., the words themselves) contributes the most to the perceiver's empathic accuracy, with an estimated weight of 50–60%; followed next by the target's paralinguistic cues (e.g., volume, pitch, and inflection), which collectively have an estimated weight of about 30%; and followed last – and decidedly least – by the target's visible nonverbal cues, with an estimated weight of about 10–20%.

One can, of course, force the visible nonverbal cues to be given more weight by eliminating or minimizing the verbal and paralinguistic ones (in cases, for example, when the perceiver is interacting with a mime, watching a silent film, making love with a vocally inhibited partner, or interacting with a person who has no language in common with the perceiver). In the overwhelming majority of everyday interactions, however, our empathic accuracy appears to rely most upon *what the other person says* (the words themselves), next most upon *how he or she says it* (the paralinguistic cues), and least upon *the accompanying visible nonverbal cues that the other person displays* (Ickes, 2006).[3]

[3] An obvious mismatch between the target person's words and his or her nonverbal behavior is the single most important exception here. When, for example, a marriage counselor observes a client say "I love you" while displaying a facial expression of contempt toward his or her spouse, the nonverbal cue usually trumps the verbal one in the counselor's assessment of the client's actual sentiment (Ickes, 2006). For evidence of a similar mismatch in interracial interactions, see Dovidio, Kawakami, and Gaertner (2002).

Are female perceivers generally more accurate than male perceivers?

According to the social stereotype of "women's intuition," the average woman might be expected to display greater empathic accuracy (as we measure it) than the average man. Surprisingly, however, Graham and Ickes (1997) found no evidence for this predicted gender difference in the first seven studies in which the empathic accuracy of male and female perceivers could be compared. After these first seven studies, the empathic accuracy data collection procedure was changed so that perceivers rated the accuracy of their empathic inferences immediately after making each of them. Remarkably, following this procedural change, evidence for the expected gender difference now emerged in the form of greater empathic accuracy scores for women than for men. So what is going on?

The results of a more recent and inclusive meta-analytic study (Ickes, Gesn, & Graham, 2000) have clarified these findings. First, this study revealed that gender differences in empathic accuracy that favor female perceivers tend to be the exception rather than the rule, in contrast to the research on nonverbal decoding accuracy, where women display a more consistent advantage (see Rosip & Hall, 2004). Second, this study suggested that when such differences do occur, they reflect gender-based differences in empathic *motivation* rather than empathic *ability* (i.e., when women are "reminded" that they are supposed to excel on empathy-relevant tasks, it motivates them to outperform men). For articles that support this motivational interpretation, see Graham and Ickes (1997), Ickes, Gesn, and Graham (2000), Klein and Hodges (2001), and Thomas and Maio (2008). For additional meta-analytic studies that are also consistent with this interpretation, see Berman (1980) and Eisenberg and Lennon (1983).

In summary, women usually *aren't* more empathically accurate than men, and on those exceptional occasions when they *are* more accurate, the difference appears to reflect women's greater motivation rather than women's greater empathic ability. (See Chapter 15 for a summary of gender differences on other interpersonal accuracy tasks.)

What are the characteristics of good versus poor "everyday mind readers"?

In contrast to other interpersonal accuracy measures, for which reliable individual difference correlates have been identified (Hall, Andrzejewski, & Yopchik, 2009; Hall, Schmid Mast, & Latu, 2014; Marsh & Blair, 2008; Murphy & Hall, 2011), reliable individual difference correlates of empathic accuracy have been remarkably difficult to find – apart from two fairly obvious ones (Ickes, 2003; Ickes & Hodges, 2013). And what are the two fairly obvious ones? The first is the perceiver's age, and the second is the perceiver's location on what has been termed the "Autism Spectrum" (Baron-Cohen, 1995).

Age-related changes in empathic ability. Infants are notoriously bad everyday mind readers, as evidenced by their rude tendency to scream their demands at us at any hour of the day or night. The maturational change associated with age – at least during the period from infancy through middle to late childhood – is a good predictor of empathic accuracy. In a recent EEG/ERP study, Cheng, Chen, and Decety (2014) found evidence which suggests that empathic accuracy may increase as young children shift from having a mostly emotional reaction to other people's experiences to having a mostly cognitive reaction to them (i.e., a shift from empathic arousal to cognitive appraisal). It remains to be seen, however, whether this shift involves the acquisition of domain-specific "theory of mind" modules in the brain (e.g., Gopnik & Wellman, 1994; Perner & Wimmer, 1985) or whether it instead results from children learning to apply to social problems computational operations of increasing complexity – from nonrepresentational to representational to meta-representational (Ickes & Decety, 2009). Finally, it is important to note that beyond the dramatic increase in everyday mind-reading ability that occurs during the relatively brief period between infancy through middle childhood, age-related change during adolescence and adulthood is not clearly associated with improved empathic accuracy (Ickes & Hodges, 2013). (See Chapter 10 for a summary of age-related performances on other interpersonal accuracy tasks.)

Autism-related deficits in empathic ability. As Hodges, Lewis, and Ickes (2014) have noted, people's location on the Autism Spectrum has also been associated with their empathic accuracy:

> People with severe autism often have severe language deficits and other behavioral disturbances that make testing their empathic accuracy essentially impossible. However, within samples who are high enough functioning to be tested using the Ickes paradigm, people diagnosed with ASD show worse empathic accuracy (Demurie, De Corel, & Roeyers, 2011; Ponnet, Buysse, Roeyers, & De Clercq, 2008). The lack of predictable structure in a social interaction appears to be an important moderator of this effect (Ponnet et al., 2008). High functioning autistics may be able to understand the "gist" of what the other person is thinking or feeling by learning to apply the relevant social schemas or scripts (see Grandin, Barron, & Zysk, 2005; Hirschfeld, Bartmess, White, & Frith, 2007; White, Hill, Winston, & Frith, 2006).

The reader is also referred to Chapter 9, which summarizes other measures of interpersonal accuracy in relation to various categories of psychopathology.

Are some target persons' thoughts and feelings harder to "read" than others'?

Yes. Target persons vary in the overall "readability" of their thoughts and feelings.[4] In fact, in studies using the unstructured dyadic interaction paradigm, substantially more variance in perceivers' empathic accuracy scores is attributable to differences in the "readability" of the target persons than to differences in the empathic ability of the perceivers themselves (Ickes, Buysse et al., 2000). Not surprisingly, this variable has emerged as an important moderator variable in empathic accuracy research. As an example, Thomas and Maio (2008) found that increasing the perceivers' motivation facilitated their empathic accuracy for more readable strangers but not for less readable ones. As another example, Thomas and Fletcher (2003) found that the target's overall readability was more predictive of the perceiver's empathic accuracy when the target was a friend versus a stranger, presumably because the perceivers had already learned how to read the thoughts and feelings of their more readable friends, but not those of their less readable ones.

How important is the amount of attention the perceiver pays to the target person?

Paying more attention to the target person is clearly important to achieving a higher level of empathic accuracy, as illustrated by a converging set of findings reported by Ickes et al. (1990). In their study of 38 pairs of opposite-sex strangers who were interacting for the first time, they found that the perceivers' empathic accuracy scores were positively correlated with (1) the percentage of partner-focused (i.e., target-focused) thoughts and feelings the perceiver had reported; (2) the percentage of partner-focused attributions the perceiver had reported (e.g., "She's nice," "He seems really nervous"); (3) the perceiver's self-monitoring score (which assesses the strength of the tendency to use other people's behavior as a

[4] Using a precedent established by Marangoni et al. (1995), the "readability" of each of the target's thoughts and feelings is assessed by trained judges who rate how easy or difficult it should be to infer that target-reported thought or feeling based on the events that immediately preceded it on the videotape record. These item-level readability ratings are then aggregated to assess how readable the target person was overall.

guide to one's own, see Ickes & Barnes, 1977); and (4) how physically attractive the perceiver's opposite-sex partner was.[5]

Paying more attention to the partner appears to be the common theme that underlies these four findings and provides the most parsimonious interpretation of them. Quantifying the exact amount of attention that the perceiver pays to the target person is a daunting challenge, however. For the present, we can say only that the perceiver's degree of attention to the partner appears to be an important predictor variable, but that a nuanced view of its role in empathic accuracy has yet to emerge.

How important is the perceiver's motivation to be accurate?[6]

It is very important – far more important than I thought it would be when my colleagues and I started studying empathic accuracy over 25 years ago. In fact, the first hint that the perceiver's motivation might be important emerged in our very first empathic accuracy study (Ickes et al., 1990). In this study of 38 mixed-sex dyads, perceivers were more accurate at "reading" the thoughts and feelings of opposite-sex strangers who were physically attractive than those who were less attractive. Still, it wasn't until several years had passed and the accumulating data had basically beat us over the head with the importance of the perceiver's motivation that we began to give it the theoretical and research attention it deserved (e.g., Ickes & Simpson, 1997, 2001; Simpson, Ickes, & Blackstone, 1995; Simpson, Ickes, & Grich, 1999; Simpson et al., 2011; Simpson, Oriña, & Ickes, 2003).

Some motives are externally induced (e.g., by a stranger's physical attractiveness) and can therefore be addressed, as Zaki and Ochsner (2011) have proposed, under the general rubric of "context effects." However, other motives are carried around inside the perceiver, sometimes interacting with relevant context variables and sometimes operating independent of them. For example, there is accumulating evidence that individuals who score high in *anxious attachment* become "hypervigilant" (and therefore more accurate) in regard to their romantic partner's thoughts and feelings in contexts in which their relationship appears to be threatened (Dugosh et al., 2011; Simpson et al., 1999; Simpson et al., 2011). This is a case in which the perceiver's mind-reading motive interacts with a relevant context variable (perceived relationship threat) to

[5] The relevant correlations were .24, .31, .24, and .24, respectively (see Ickes, 2009, p. 169).
[6] My answer to this question is adapted from an earlier commentary by Ickes (2011).

affect the perceiver's inferential accuracy. On the other hand, there is also accumulating evidence that individuals who score high in *avoidant attachment* are generally less motivated to learn what their partners are thinking and feeling (Rholes, Simpson, Tran, McLeish, & Friedman, 2007; Simpson et al., 1999; Simpson et al., 2011), an effect that appears to be relatively independent of context.

In most of the laboratory research on empathic accuracy, the researcher tries to minimize individual differences in motivation by encouraging all participants to make the most accurate inferences they can. Different means are deployed in the attempt to accomplish this goal: using motivating task instructions, presenting the task as a kind of "test," or even offering a financial incentive for good performance (Klein & Hodges, 2001). However successful or unsuccessful these procedures might be in "homogenizing" the perceivers' motivation levels, the fact remains that, in people's everyday lives, their motivation to be accurate or not is both complicated and highly variable, stemming both from people's long-standing motives and from the more transient, situationally induced motives that can quickly appear and then, just as quickly, vanish (see Chapter 11 for a review of situational influences on interpersonal accuracy). Moreover, there is a remarkable body of recent evidence showing that perceivers can "dial down" or "dial up" their inferential accuracy skills virtually at will, in the service of their current motivations (for a recent summary of this work, see Smith, Ickes, Hall, & Hodges, 2011).

If this last assertion sounds hard to believe, consider the following research examples. First, maritally aggressive men are not only more likely to disattend (i.e., "tune out") a woman's complaints (Schweinle & Ickes, 2007), but the resulting empathic *in*accuracy predicts their likelihood of abusing their wives (Schweinle & Ickes, 2007; Schweinle, Ickes, & Bernstein, 2002). Second, anxiously attached women who listen to their male dating partners being interviewed by an attractive female interviewer are more likely than less anxious women to closely monitor (i.e., "tune in" to) their partner's behavior and correctly predict his answers to the interview questions (Dugosh et al., 2011). Third, men who expect to get paid for greater empathic accuracy appear to exert more effort to (successfully) achieve it than men who don't expect to get paid (Klein & Hodges, 2001). Fourth, studies by Thomas and Maio (2008) and by Smith and Lewis (2009) have shown that men dial up or dial down their interpersonal sensitivity depending on whether they believe that being sensitive or being insensitive is associated with a socially-desirable male gender role.

After resisting the importance of mind-reading motivation for several years and eventually being forced by the data to acknowledge its pervasive effects, I can't regard any theoretical model of the inferential accuracy process as complete unless it explicitly addresses the perceiver's motivational concerns. In fact, it was this grudging admission that eventually led Jeffry Simpson and me to propose our *empathic accuracy model*, in which the perceiver's motivation to be accurate or inaccurate plays a central theoretical role (Ickes & Simpson, 1997, 2001). As the reader will see, this model is useful in helping us to answer the next question we consider.

When does empathic accuracy help relationships? When does it hurt them?

Ickes and Simpson's (1997, 2001) empathic accuracy model proposed that empathic accuracy can either help or hurt relationships. In situations that are unlikely to evoke evidence that one's partner's thoughts and feelings would threaten the relationship, greater empathic accuracy should typically help the relationship by fostering greater understanding of, and closeness to, the partner. However, in situations that are likely to evoke evidence that one's partner's thoughts and feelings are relationship threatening, greater empathic accuracy should typically hurt the relationship by increasing the perceiver's personal distress, reducing the perceiver's feelings of closeness to the partner, and destabilizing the relationship.

Consistent with these general predictions, Simpson, Oriña, and Ickes (2003) studied nearly 100 marital conflict interactions and found that greater empathic accuracy for the partner's nonthreatening thoughts and feelings increased the perceiver's feeling of closeness to the partner, whereas greater empathic accuracy for the partner's threatening thoughts and feelings decreased the perceiver's feelings of closeness to the partner. Looking at these predictions from a different perspective, Simpson, Ickes, and Blackstone (1995) predicted that the *avoidance* of empathic accuracy about a dating partner's relationship-threatening thoughts and feelings should help to keep the relationship from becoming unstable in the face of this situational threat. The authors confirmed this prediction and further reported that the couples who displayed the greatest avoidance of empathic accuracy in this situation were significantly more likely to still be dating five months later.

What kinds of situations are especially likely to evoke evidence that one's partner's thoughts and feelings are relationship threatening (Ickes, 2014)? In general, greater empathic accuracy appears to hurt relationships when perceivers:

1. uncover blunt, unpleasant truths about each other's private thoughts and feelings that could undermine their views of each other and of

their relationship (Aldous, 1977; Rausch, Barry, Hertel, & Swain, 1974; Sillars, 1985; Watzlawick, Weakland, & Fisch, 1974)
2. discover that their differences are greater than they previously believed, or are apparently irreconcilable, so that extended discussion and clarification of their respective viewpoints will not improve the relationship, but only make things worse (Aldous, 1977; Kursh, 1971; Sillars, 1985)
3. discover that certain "benign misconceptions" they have previously held about each other are false and can no longer be sustained (Levinger & Breedlove, 1966; Sillars, 1985)
4. use their empathic insights to torture each other and "push each other's buttons," like the characters George and Martha in Edward Albee's play *Who's Afraid of Virginia Woolf?* (Chapter 11 in Ickes, 2003)
5. insist on knowing virtually everything that their partners are thinking or feeling, to the point that their partners feel violated, intruded upon, and lacking any sense of privacy within their own minds (Chapter 1 in Ickes, 2003)

Is empathic accuracy relevant to clinical problems and populations? If so, which ones?

After more than 25 years of empathic accuracy research, the answer to this question is a resounding yes. Many published studies have linked empathic accuracy to a wide range of clinical problems and populations (Ickes, 2011). These problems and populations are summarized below:

1. autism and Asperger syndrome (Demurie et al., 2011; Ponnet et al., 2008; Ponnet, Buysse, Roeyers, & De Corte, 2005; Ponnet, Roeyers, Buysse, De Clercq, & Van Der Heyden, 2004; Royers, Buysse, Ponnet, & Pichal, 2001)
2. attention-deficit/hyperactivity disorder (Demurie et al., 2011)
3. borderline personality disorder (Flury, Ickes, & Schweinle, 2008)
4. deficits in social proficiency (Bartz et al., 2010)
5. marital conflict interactions (Simpson et al., 2011, 2003)
6. the social cognition of maritally aggressive/abusive men (Clements et al., 2007; Robillard & Noller, 2011; Schweinle & Ickes, 2007; Schweinle et al., 2002; Schweinle, Ickes, Rollings, & Jacquot, 2010)
7. schizophrenia (Lee, Zaki, Harvey, Ochsner, & Green, 2011; Harvey, Zaki, Ochsner, & Green, 2013)
8. schizotypal personality disorder (Ripoll et al., 2013)
9. the training of student nurses (Hall, Ship, et al., 2014)
10. the training of student psychotherapists (Barone et al., 2005).

Although space does not permit even a cursory overview of these clinically relevant studies, it should be evident from the length and diversity of the above list that empathic accuracy may prove to be a phenomenon of major importance to a wide range of clinical problems and populations. And I hope it is also evident from everything I have said in this chapter that the research on empathic accuracy is highly relevant to the social psychology of perceiving other people accurately.

References

Aldous, J. (1977). Family interaction patterns. *Annual Review of Sociology*, 3, 105–135.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind.* Cambridge: MIT Press.

Barone, D. F., Hutchings, P. S., Kimmel, H. J., Traub, H. L., Cooper, J. T., & Marshall, C. M. (2005). Increasing empathic accuracy through practice and feedback in a clinical interviewing course. *Journal of Social and Clinical Psychology*, 24, 156–171.

Bartz, J., Zaki, J., Bolger, N., Hollander, E., Ludwig, N., Kolevzon, A., & Ochsner, K. (2010). Oxytocin selectively improves empathic accuracy in less socially proficient individuals. *Psychological Science*, 21, 1426–1428.

Berman, P. T. (1980). Are women more responsive than men to the young? A review of developmental and situational variables. *Psychological Bulletin*, 88, 668–695.

Cheng, Y., Chen, C., & Decety, J. (2014). An EEG/ERP investigation of the development of empathy in early and middle childhood. *Developmental Cognitive Neuroscience*, 10, 160–169.

Clements, K., Holtzworth-Munroe, A., Schweinle, W., & Ickes, W. (2007). Empathic accuracy of intimate partners in violent versus nonviolent relationships. *Personal Relationships*, 14, 369–388.

Demurie, E., De Corel, M., & Roeyers, H. (2011). Empathic accuracy in adolescents with autistic spectrum disorders and adolescents with attention-deficit/hyperactivity disorder. *Research in Autism Spectrum Disorders*, 5, 126–134.

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.

Dugosh, J. W., Cheng, W., & Park, A. E. (2011). Adult attachment styles and motivated accuracy. In J. L. Smith, W. Ickes, J. A. Hall, & S. D. Hodges (Eds.), *Managing interpersonal sensitivity: Knowing when—and when not—to understand others* (pp. 125–142). New York: Nova Science Publishers.

Eisenberg, N., & Lennon, R. (1983). Sex differences in empathy and related capacities. *Psychological Bulletin*, 94, 100–131.

Flury, J., Ickes, W., & Schweinle, W. (2008). The borderline empathy effect: Do high BPD individuals have greater empathic ability? Or are they just more difficult to "read"? *Journal of Research in Personality*, 42, 312–332.

Gadassi, R., Mor, N., & Rafaeli, E. (2011). Depression and empathic accuracy in couples: An interpersonal model of gender differences in depression. *Psychological Science*, 22, 1033–1041.

Garcia, S., Stinson, L., Ickes, W., Bissonnette, V., & Briggs, S. R. (1991). Shyness and physical attractiveness in mixed-sex dyads. *Journal of Personality and Social Psychology*, 61, 35–49.

Gesn, P. R., & Ickes, W. (1999). The development of meaning contexts for empathic accuracy: Channel and sequence effects. *Journal of Personality and Social Psychology*, 77, 746–761.

Gleason, K. A., Jensen-Campbell, L. A., & Ickes, W. (2009). The role of empathic accuracy in adolescents' peer relations and adjustment. *Personality and Social Psychology Bulletin*, 35, 997–1011.

Gopnik, A., & Wellman, H. M. (1994). The "theory" theory. In L. A. Hirschfield & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 257–294). Cambridge, UK: Cambridge University Press.

Graham, T. (1994). *Gender, relationship, and target differences in empathic accuracy*. Unpublished master's thesis, University of Texas at Arlington.

Graham, T., & Ickes, W. (1997). When women's intuition isn't greater than men's. In W. Ickes (Ed.), *Empathic accuracy* (pp. 117–143). New York: Guilford Press.

Grandin, T., Barron, S., & Zysk, V. (2005). *The unwritten rules of social relationships*. Arlington, TX: Future Horizons.

Hall, J. A., Andrzejewski, S. A., & Yopchick, J. E. (2009). Psychosocial correlates of interpersonal sensitivity: A meta-analysis. *Journal of Nonverbal Behavior*, 33, 149–180.

Hall, J. A., & Schmid Mast, M. (2007). Sources of accuracy in the empathic accuracy paradigm. *Emotion*, 7, 438–446.

Hall, J. A., Schmid Mast, M., & Latu, I. (2014). The vertical dimension of social relations and accurate interpersonal perception: A meta-analysis. *Journal of Nonverbal Behavior*, 39, 131–163.

Hall, J. A., Ship, A. N., Ruben, M. A., Curtin, E. M., Roter, D. L., Clever, S. L. Smith, C. C., & Pounds, K. (2014). The Test of Accurate Perception of Patients' Affect (TAPPA): An ecologically valid tool for assessing interpersonal perception accuracy in clinicians. *Patient Education and Counseling*, 94, 218–223.

Harvey, P-O., Zaki, J., Lee, J., Ochsner, K., & Green, M. F. (2013). Neural substrates of empathic accuracy in people with schizophrenia. *Schizophrenia Bulletin*, 39, 617–628.

Hirschfield, L., Bartmess, E., White, S., & Frith, U. (2007). Can autistic children predict behavior by social stereotypes? *Current Biology*, 17, 451–452.

Hodges, S. D., Laurent, S. M., & Lewis, K. L. (2011). Specially motivated, feminine, or just female: Do women have an empathic accuracy advantage? In J. L. Smith, W. Ickes, J. A. Hall, & S. D. Hodges (Eds.), *Managing interpersonal sensitivity: Knowing when—and when not—to understand others* (pp. 59–74). New York: Nova Science Publishers.

Hodges, S. D., Lewis, K. L., & Ickes, W. (2014). The matter of other minds: Empathic accuracy and the factors that influence it. In M. Mikulincer, & P. R. Shaver (Eds.), J. A. Simpson, & J. F. Dovidio (Assoc. Eds.) (2015), *APA handbook of personality and social psychology, Volume 3: Interpersonal relations* (pp. 319–348). Washington, DC: American Psychological Association.

Ickes, W. (2001). Measuring empathic accuracy. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement* (pp. 219–241). Mahwah, NJ: Erlbaum.

Ickes, W. (2003). *Everyday mind reading: Understanding what other people think and feel*. Amherst, NY: Prometheus.

Ickes, W (2006). *Inferring other people's thoughts and feelings: The relative importance of verbal versus nonverbal cues*. Invited talk given at the annual conference of the National Communication Association, San Antonio, Texas, November 18, 2006.

Ickes, W. (2011). Everyday mind reading is driven by motives and goals. *Psychological Inquiry*, 22, 200–206.

Ickes, W. (2014). *Empathic accuracy in close relationships*. Invited talk given at the Interdisciplinary Symposium on Empathy at Utrecht University, Utrecht, Netherlands, May 1, 2014.

Ickes, W., & Barnes, R. D. (1977). The role of sex and self-monitoring in unstructured dyadic interactions. *Journal of Personality and Social Psychology*, 35, 315–330.

Ickes, W., Buysse, A., Pham, H., Rivers, K., Erickson, J. R., Hancock, M., Kelleher, J., & Gesn, P. R. (2000). On the difficulty of distinguishing "good" and "poor" perceivers: A social relations analysis of empathic accuracy data. *Personal Relationships*, 7, 219–234.

Ickes, W., & Decety, J. (2009). Introduction: Seeking to understand the minds (and brains) of people who are seeking to understand other people's minds. In J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. vii–ix). Cambridge, MA: The MIT Press.

Ickes, W., Gesn, P. R., & Graham, T. (2000). Gender differences in empathic accuracy: Differential ability or differential motivation? *Personal Relationships*, 7, 95–109.

Ickes, W., & Hodges, S. (2013). Empathic accuracy in close relationships. In J. A. Simpson & L. Campbell (Eds.), *Handbook of close relationships* (pp. 348–373). New York: Oxford University Press.

Ickes, W., & Simpson, J. (1997). Managing empathic accuracy in close relationships. In W. Ickes (Ed.), *Empathic accuracy* (pp. 218–250). New York: Guilford Press.

Ickes, W., & Simpson, J. (2001). Motivational aspects of empathic accuracy. In G. J. O. Fletcher & M. S. Clark (Eds.), *Interpersonal processes: Blackwell Handbook in Social Psychology* (pp. 229–249). Oxford, UK: Blackwell.

Ickes, W., Stinson, L., Bissonnette, V., & Garcia, S. (1990). Naturalistic social cognition: Empathic accuracy in mixed-sex dyads. *Journal of Personality and Social Psychology*, 59, 730–742.

Key, M. R. (1975). *Paralanguage and kinesics*. Metuchen, NJ: The Scarecrow Press.

Klein, K. J. K., & Hodges, S. (2001). Gender differences, motivation, and empathic accuracy: When it pays to understand. *Personality and Social Psychology Bulletin*, 27, 720–730.

Kursh, C. O. (1971). The benefits of poor communication. *Psychoanalytic Review*, 58, 189–208.

Leathers, D. (1978). *Nonverbal communications systems*. Boston: Allyn & Bacon.

Lee, J., Zaki, J., Harvey, P., Ochsner, K., & Green, M. (2011). Schizophrenia patients are impaired in empathic accuracy. *Psychological Medicine*, 41, 2297–2304.

Levenson, R. W., & Ruef, A. M. (1992). Empathy: A physiological substrate. *Journal of Personality and Social Psychology*, 63, 234–246.

Levinger, G., & Breedlove, J. (1966). Interpersonal attraction and agreement. *Journal of Personality and Social Psychology*, 3, 367–372.

Marangoni, C., Garcia, S., Ickes, W., & Teng, G. (1995). Empathic accuracy in a clinically relevant setting. *Journal of Personality and Social Psychology*, 68, 854–869.

Marsh, A. A., & Blair, R. J. R. (2008). Deficits in facial affect recognition among antisocial populations: A meta-analysis. *Neuroscience and Biobehavioral Reviews*, 32, 454–465.

Murphy, N. A., & Hall, J. A. (2011). Intelligence and nonverbal sensitivity: A meta-analysis. *Intelligence*, 39, 54–63.

Perner, J., & Wimmer, H. (1985). "John thinks that Mary thinks that ..." Attribution of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology*, 39, 437–471.

Ponnet, K., Buysse, A., Roeyers, H., & De Clercq, A. (2008). Mind-reading in young adults with ASD: Does structure matter? *Journal of Autism and Developmental Disorders*, 38, 905–918.

Ponnet, K., Buysse, A., Roeyers, H., & De Corte, K. (2005). Empathic accuracy in adults with a pervasive developmental disorder during an unstructured conversation with a typically developing stranger. *Journal of Autism and Developmental Disorders*, 35, 585–600.

Ponnet, K., S., Roeyers, H., Buysse, A., De Clercq, A., & Van Der Heyden, E. (2004). Advanced mind-reading in adults with Asperger syndrome. *Autism*, 8, 249–266.

Rausch, H. L., Barry, W. A., Hertel, R. K., & Swain, M. A. (1974). *Communication conflict and marriage*. San Francisco, CA: Jossey-Bass.

Rholes, W. S., Simpson, J. A., Tran, S., McLeish, M. A., & Friedman, M. (2007). Attachment and information seeking in romantic relationships. *Personality and Social Psychology Bulletin*, 33, 422–438.

Ripoll, L. H., Zaki, J., Perez-Rodriguez, M. M., Snyder, R., Strike, K. S, Boussi, A., Bartz, J. A., Ochsner, K. N., Siever, L. J., & New, A. S. (2013). Empathic accuracy and cognition in schizotypal personality disorder. *Psychiatry Research*, 210, 232–241.

Robillard, L., & Noller, P. (2011). Rejection sensitivity, violence and decoding deficits in married men. In J. L. Smith, W. Ickes, J. A. Hall, & S. D. Hodges (Eds.), *Managing interpersonal sensitivity: Knowing when and when not to understand others* (pp. 154–167). New York: Nova Science Publishers.

Roeyers, H., Buysse, A., Ponnet, K., & Pichal, B. (2001). Advancing advanced mind-reading tests: Empathic accuracy in adults with a pervasive developmental disorder. *Journal of Child Psychology and Psychiatry*, 42, 271–278.

Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology*, 21, 95–103.

Rosip, J. C., & Hall, J. A. (2004). Knowledge of nonverbal cues, gender, and nonverbal decoding accuracy. *Journal of Nonverbal Behavior*, 28, 267–286.

Schmid Mast, M., & Ickes, W. (2007). Empathic accuracy: Measurement and potential clinical applications. In T. F. D. Farrow & P. W. R. Woodruff (Eds.), *Empathy in mental illness* (pp. 408–427). Cambridge, UK: Cambridge University Press.

Schweinle, W., & Ickes, W. (2007). The role of men's critical/rejecting overattribution bias, affect, and attentional disengagement in marital aggression. *Journal of Social and Clinical Psychology*, 26, 173–198.

Schweinle, W. E., Ickes, W., & Bernstein, I. H. (2002). Empathic inaccuracy in husband to wife aggression: The overattribution bias. *Personal Relationships*, 9, 141–159.

Schweinle, W. E., Ickes, W., Rollings, K., Jacquot, C. (2010). Maritally aggressive men: Angry, egocentric, impulsive, and/or biased. *Journal of Language and Social Psychology*, 29, 399–424.

Sillars, A. L. (1985). Interpersonal perception in relationships. In W. Ickes (Ed.), *Compatible and incompatible relationships* (pp. 277–305). New York: Springer-Verlag.

Simpson, J., Ickes, W., & Blackstone, T. (1995). When the head protects the heart: Empathic accuracy in dating relationships. *Journal of Personality and Social Psychology*, 69, 629–641.

Simpson, J. A., Ickes, W., & Grich, J. (1999). When accuracy hurts: Reactions of anxious-uncertain individuals to a relationship-threatening situation. *Journal of Personality and Social Psychology*, 76, 754–769.

Simpson, J. A., Kim, J. S., Fillo, J., Ickes, W., Rholes, S., Oriña, M. M., & Winterheld, H. A. (2011). Attachment and the management of empathic accuracy in relationship threatening situations. *Personality and Social Psychology Bulletin*, 37, 242–254.

Simpson, J. A., Oriña, M. M., & Ickes, W. (2003). When accuracy hurts, and when it helps: A test of the empathic accuracy model in marital interactions. *Journal of Personality and Social Psychology*, 85, 881–893.

Smith, J. L., Ickes, W., Hall, J. A., & Hodges, S. D. (Eds.). (2011). *Managing interpersonal sensitivity: Knowing when and when not to understand others*. New York: Nova Science Publishers.

Smith, J. L., & Lewis, K. L. (2009). Men's interpersonal (mis)perception: Fitting in with gender norms following social rejection. *Sex Roles*, 61, 252–264.

Stinson, L., & Ickes, W. (1992). Empathic accuracy in the interactions of male friends versus male strangers. *Journal of Personality and Social Psychology*, 62, 787–797.

Thomas, G., & Fletcher, G. J. O. (2003). Mind-reading accuracy in intimate relationships: Assessing the roles of the relationship, the target, and the judge. *Journal of Personality and Social Psychology*, 85, 1079–1094.

Thomas, G., & Maio, G. R. (2008). Man, I feel like a woman: When and how gender-role motivation helps mind-reading. *Journal of Personality and Social Psychology*, 95, 1165–1179.

Verhofstadt, L. L., Buysse, A., & Ickes, W. (2007). Social support in couples: An examination of gender differences using self-report and observational methods. *Sex Roles*, 57, 267–282.

Verhofstadt, L. L., Buysse, A., Ickes, W., Davis, M., & Devoldre, I. (2008). Support provision in marriage: The role of emotional linkage and empathic accuracy. *Emotion*, 8, 792–802.

Watzlawick, P., Weakland, J., & Fisch, R. (1974). *Principles of problem formation and problem resolution*. New York: Norton.

White, S., Hill, E. L., Winston, J., & Frith, U. (2006). An islet of social ability in Asperger Syndrome: Judging social attributes from faces. *Brain and Cognition*, 61, 69–77.

Wilhelm, P., & Perrez, M. (2004). How is my partner feeling in different daily life settings? Accuracy of spouses' judgments about their partner's feelings at work and at home. *Social Indicators Research*, 67, 183–246.

Zaki, J., Bolger, N., & Ochsner, K. (2008). It takes two: The interpersonal nature of empathic accuracy. *Psychological Science*, 19, 399–404.

Zaki, J., Bolger, N., & Ochsner, K. (2009). Unpacking the informational bases of empathic accuracy. *Emotion*, 9, 478–487.

Zaki, J., & Ochsner, K. (2011). Re-integrating the study of accuracy into social cognition research. *Psychological Inquiry*, 22, 159–182.

4 Accuracy of distinguishing truth from lie

Judee K. Burgoon and Norah E. Dunbar

Abstract

Judging another's veracity is a complex and difficult, albeit second-nature, undertaking. Humans ascertaining the truthfulness of another's communication without the benefit of any special training or instrumentation average only 54% overall accuracy. Experts often perform no better than laypeople when judging the brief, decontextualized, low-stakes video and audio clips that dominate the meta-analytic literature. However, judges attain higher accuracy when they are trained and experienced, engage in longer interactions, have access to context information or baseline behavior, and adopt strategic questioning strategies. Judgment accuracy is a function not only of characteristics of the judge but also the sender, sender–receiver relationship, signal, communication genre (e.g., interview), modality (e.g., face-to-face, audio), and context. Behavioral observation techniques used by trained human coders, software for automating behavioral observation, and instruments for sensing and tracking behavior are being developed at a rapid rate, enabling improved accuracy by both experts and lay people.

A fundamental canon of human communication, tracing back to the Greek philosopher Aristotle, is that people are judged intrinsically based on their ethos or credibility, with veracity being central to that judgment (McCroskey, 1966). Indeed, a stable social system depends on truth and honesty, a principle formalized by Grice (1989) as the cooperative principle (CP). The CP maxim of *quality* holds that people should/will speak the truth, avoid falsehoods, and eschew claims for which they lack evidence. Yet paradoxically, survival of the species has relied on just the opposite, with the "fittest" often being the most successful liars and deceivers (Knapp, 2008). In fact, deceit is not an anomaly. Many empirical estimates point to deceit being quite prevalent in contemporary discourse, yet humans – both lay people and professionals (e.g., law enforcement, fraud investigators, intelligence analysts, security screening officers) – are said to be poor at detecting such dissembling (Hartwig & Bond, 2011). Another paradox: If humans have frequent exposure to and practice committing deception, why do they fail to detect it in others? This chapter is intended to explain these paradoxes.

Deception, detection, and related constructs

Defining deception

A perusal of deception literature over the last 50 years shows almost as many definitions of deception as there are authors on the subject. Some scholars have suggested *deception* is an umbrella term for the full gamut of fraudulent, tricky, or misleading behaviors, with *lies* being but one type (Knapp, 2008). Other scholars have explicitly stated they use the terms *lies* and *deception* interchangeably (Masip, Garrido, & Herrero, 2004; Vrij, 2008); for the sake of economy we follow that tradition here. Based on a review of several definitions offered by researchers in psychology and communication, Masip et al. (2004) settled on this comprehensive definition:

> Deception can be understood as the deliberate attempt, whether successful or not, to conceal, fabricate, and/or manipulate in any other way factual and/or emotional information, by verbal and/or nonverbal means, in order to create or maintain in another or in others a belief that the communicator himself or herself considers false. *(p. 148)*

Embedded within this definition are some important concepts with implications for detecting deceit. First, deception is intentional. So-called honest errors should not be called deception, which involves deliberate manipulation of verbal information and nonverbal displays. Second, while many definitions omit emotional or nonverbal information and focus on factual verbal statements, the hundreds of studies on the impact and perceptions of nonverbal behavior reinforce its importance (Sporer & Schwandt, 2007). Third, deception serves an instrumental purpose and is rarely an end in itself. Motives matter.

Veracity versus credibility

Implicit in the "considers false" part of the definition is the notion that the actual veracity of a claim is different from the credibility of a claim. The former refers to the actual truth or falsity of information, whereas the latter refers to its believability. Put differently, veracity is a property of sender communication, whereas credibility is a judgment held by receivers. Deceivers are perfectly *credible* if they are invariably perceived to be truthful – whether lying or telling the truth (Bond & DePaulo, 2008).

Dichotomous versus gradient conceptualizations

Although truth/lie has often been conceptualized as a dichotomy, deception comprises a panoply of tactics for misleading others: omission, equivocation, evasion, ambiguity, hyperbole, white lies, half-truths, fabrications, falsehoods, and dissembling, to name a few. These varied forms of deception should make clear that veracity forms a "gradient phenomenon" (Coleman & Kay, 1981) ranging from perfectly truthful to entirely deceptive. The variety of characteristics distinguishing forms of deceit contribute to the challenges of detecting it. Moreover, Vrij (2008) notes that lies are often embedded within truths. Liars can tell very plausible stories by lying about only one detail and peppering their stories with many other truthful elements, making lies further difficult to detect.

Detection accuracy and bias

Conceptually, accuracy in the context of judging truths and lies refers to how well judges (interlocutors or observers) recognize, comprehend, recall, and interpret the messages produced by other actors (hereafter, senders). In practice, accuracy refers to an objectively quantifiable metric by which the communication that is received can be compared against some objective standard, or what is often referred to as "ground truth." Detection accuracy can be measured dichotomously as lying/not lying or as a continuous measure of honesty or sincerity (e.g., DePaulo & Kirkendol, 1989). In our coauthored experiments (e.g., Burgoon, Blair, & Strom, 2008; Burgoon & Qin, 2006; Dunbar et al., 2015), we have employed rating scales ranging from 0 (*not at all*) to 10 (*completely truthful* or *completely honest*) by which participants self-report their truthfulness, and judges report their assessment of sender veracity. We calculated accuracy as the difference between these two values such that smaller values reflect closer alignment between senders' and judges' reports. In many cases, even participants who are being deceptive may self-report values above the mean (i.e., toward the truthful end of the continuum) but self-report even higher truthfulness ratings, indicating differentiation between deception and truth.

Bias, by contrast, is the degree to which the judgment by the perceiver systematically deviates from what the sender said. The goal for detection is to increase the accuracy of judgments and reduce bias.

Several biases have been demonstrated to interfere with accuracy, the most robust among them being the truth bias (a systematic overestimate of truthfulness, irrespective of actual truth) and its opposite, the lie bias (the systematic overestimate of deceptiveness irrespective of actual degree of deception). The "veracity effect" (Levin, Park, & McCornack, 1999) offers a heuristic explanation for the truth bias. Because lies are rare relative to truths, and people typically do not want to accuse others of deception without proof, judges are more apt to choose truth when guessing; thus truth accuracy is usually higher than deception accuracy. However, guessing correctly that a sender is truthful is not the same as making a considered judgment of truthfulness; it just means that the truth bias has served the receiver well in that case. We prefer to reserve the term "accuracy" for instances where judgments are systematically correct, not the result of merely random but lucky guesses.

Signal detection theory

Signal detection theory (SDT) is not a theory per se but rather a way of decomposing accuracy into forms of correct and incorrect judgments. When detection accuracy is reported, it is often unclear whether what is being reported is only accurate detection of *deception* or accurate detection of *truth and deception combined*. It has become common practice to follow practitioners in applying SDT (Green & Swets, 1966), which parses accuracy into hits, misses, false alarms, and false negatives. *Hits* (also known as the true positive rate (TPR) or the genuine accept ratio (GAR) or sensitivity) are the percentage of true deceptive cases that are classified accurately as deceptive. *Misses* (also known as the false negative rate (FNR)) are the percentage of actual deceptive cases that are erroneously misclassified as truthful. *False alarms* (also called false positive rate (FPR) or false accept rate (FAR)) are the percentage of truthful cases that are erroneously identified as deceptive. *True negatives* (or true negative rate (TNR)) are the percentage of truthful cases that are correctly identified as truthful. Hits and true negatives together form the total detection accuracy rate. Best overall deception detection accuracy occurs when true deception detection is coupled with a low rate of false alarms. This can be displayed graphically as a receiver operating characteristic (ROC) curve, which shows alternative trade-offs between how sensitive the judge or instrument is in detecting true deception and how specific the judge or instrument is by avoiding false alarms. The area under the curve (AUC) provides an objective measure of how well humans or instruments are doing detecting target cases while excluding nontarget cases, where target cases are the ones a judge is trying to detect, i.e., deceit and nontarget cases are the ones to be excluded, i.e., truths. One could, for example, judge all cases as deceptive and attain 100% deception detection accuracy, but the corresponding truth detection accuracy would then be 0% (all truth judgments would be false alarms). The greater the AUC, the more successful the classification model in simultaneously detecting lies and avoiding false alarms.

The status of detection accuracy

One of the major streams of deception detection has compared the ability of laypeople and professionals to judge veracity. The typical study shows judges a limited set of stimuli such as video clips, audio segments, or transcripts; judges either make a dichotomous truth/lie judgment for each sample or rate each on an interval metric of truthfulness, honesty, or sincerity. Four separate meta-analyses of thousands of data points from hundreds of samples and thousands of judges (Aamodt & Custer, 2006; Bond & DePaulo, 2006, 2008; Hartwig & Bond, 2011; Sporer & Schwandt, 2007) have arrived at remarkably consistent estimates, showing that unaided veracity detection by humans averages 54%, which though statistically different than chance (50%), is only marginally so and not considered a meaningful improvement. When that estimate is split into separate truth and deception detection accuracy rates, an important distinction emerges: truth detection accuracy approximates 61%, whereas deception detection accuracy approximates 47% (i.e., below chance). The higher truth detection accuracy demonstrates the truth bias at work.

The conclusion that professionals do not differ substantially from laypeople in their detection accuracy has been challenged due to the aforementioned methods by which judges are tested. In the bulk of deception detection studies that populate the meta-analyses, the samples that judges watch, listen to, or read are very brief, decontextualized, often inconsequential lies; judges make passive truth–lie judgments after each snippet. This is a far cry from what happens in real, high-stakes circumstances where judges have access to contextual and background information and can engage in active, prolonged, and strategic questioning such as occurs in interviews and interrogations (Vrij & Granhag, 2012). When experts render judgments under the latter kinds of conditions, their detection accuracy is much higher (e.g., Blair, Levine, & Shaw, 2010; O'Sullivan, Frank, Hurley, & Tiwana, 2009; Vrij, Mann, Robbins, & Robinson, 2006). Although professionals overestimate their own detection accuracy and skill, experts' confidence is actually statistically uncorrelated with accuracy (DePaulo, Charlton, Cooper, Lindsay, & Muhlenbruck, 1997; Vrij & Baxter, 1999). Interestingly, self-reported accuracy is also weak (though not quite nonexistent) when judging affective states (Hall, Andrzejewski, & Yopchick, 2009). Thus, self-reported accuracy or confidence is no substitute for objective measures.

Theoretical perspectives on deception detection

Some perspectives on deception detection draw upon general human judgment models to understand the mechanisms behind veracity judgments. A prominent example is Chaiken's heuristic-systematic processing model (Chaiken, 1980; Chaiken, Liberman, & Eagly, 1989), which argues that humans rely on mental shortcuts (or heuristics) when under time pressure or when not highly motivated to think systematically about decisions. The dangerous decisions theory (DDT; Porter, ten Brinke, & Gustaw, 2010) applies the concept of heuristics specifically to deception detection. Grounded in the judicial context, DDT proposes that jurors make rapid intuitive judgments based on stereotypes and other heuristics that are influenced by facial appearance. Like the "what is beautiful is good" stereotype, impressions from the face unduly influence credibility judgments (e.g., not taking time to decide a criminal suspect's guilt when the suspect's face connotes trustworthiness). The result is often poor and irrational judgments that impair the ability to discern deceit.

Of theories and models that center explicitly on receiver detection, those developed by psychologists tend to be narrow in scope and to concern single causal mechanisms. For example, the motivational impairment effect (MIE; DePaulo & Kirkendol, 1989) considers how motivation of the sender (operationally defined by DePaulo variously as interacting with the opposite versus same gender, lying to an attractive or unattractive other, being incentivized to succeed, fear of detection, or having high versus low achievement motivation; see Burgoon, 2005, for a discussion) affects judgment accuracy. The MIE is a two-part hypothesis stating that judges do better detecting deception by motivated senders than unmotivated ones when judges have access to nonverbal channels (ones with visual and/or auditory nonverbal signals) but do worse with motivated deceivers when judgments are based on the verbal channel (i.e., words) (DePaulo, Lanier, & Davis, 1983). Frequently, the MIE is mischaracterized in the literature as only concerned with the effects of motivation on nonverbal performance, but the original hypothesis is actually about the judgment process and modalities – ones presenting nonverbal or verbal signals – under which judgments are made.[1]

Theoretical models from communication scholars tend to be more macroscopic in scale and to foreground aspects of discourse production or the communication process. McCornack's (1992) information manipulation theory (IMT) assumes that deception involves *covert* violations of Grice's (1989) conversational maxims or rules (quantity, quality, manner, and relation) in which messages provide incomplete, misleading, ambiguous, or irrelevant information. Because these violations are covert, with intentions hidden, and discourse manipulation easily creates the perception that senders are fulfilling their conversational obligations, judges fail to detect them. In a recent radical shift from the receiver focus of IMT to the sender focus of the new "IMT2," McCornack, Morrison, Paik, Wisner, and Zhu (2014) shift their focus from the characteristics of deceptive messages or the ways in which violations of the CP are manifested in discourse[2] to the cognitive processing that underpins the production of deceptive messages. Their premise is that the biggest determinant of whether someone will deceive is the nature of the information held in working or long-term memory, and how problematic it is.

Interpersonal deception theory (IDT; Buller & Burgoon, 1996; Burgoon & Buller, 2015) is the broadest in scope in that it considers not only message production and dyadic interaction factors but also receiver judgment factors. The theory lays out both the assumptions upon which it is founded and a series of empirically testable propositions from which hypotheses can be generated. IDT contends that although deceivers may experience emotional and cognitive factors that might cause them to be more detectable, they also engage in strategic management of their performance, which should make them less detectable.

[1] For an analysis of the original tests of MIE and for the effects of motivation on verbal and nonverbal performance, see Burgoon (2005) and Burgoon and Floyd (2000). They conclude that extreme motivation and arousal can impair aspects of both verbal and nonverbal performance by both truth tellers and deceivers, but moderate motivation and arousal can facilitate both for both kinds of senders. The effects of modality of judgment are discussed in the "Partitioning Variance" section.

[2] Because McCornack et al. (2014) argue that lies are not necessarily more cognitively demanding than truths, they dispute the common contention that cognitive difficulty adds to the diagnosticity of deceptive messages. Vrij and colleagues, among others, although arguing that most lies should be more cognitively taxing than truths, concede "that the differences [in cognitive load] between liars and truth tellers may be relatively small and not readily discernible by observers" (Vrij, Leal, Mann, & Fisher, 2012, p. 581). They recommend using questioning techniques that magnify the cognitive load for deceivers and make the differences from truth tellers more apparent (Vrij, 2004; Vrij, Granhag, Mann, & Leal, 2011).

One central factor that affects strategic activity is context interactivity (i.e., when sender and receiver participate in a face-to-face dialogue in real time, and utterances are contingent upon one another). Interactivity gives deceivers ongoing access to receiver feedback and facilitates adapting to any perceived suspicion that receivers inadvertently display. Comparatively, when deceivers give a monologue to receiver(s), or receivers passively observe deceivers rather than interacting with them, deception is more readily detected (Buller, Strzyzewski, & Hunsaker, 1991; Burgoon, Buller, & Floyd, 2001; Dunbar, Ramirez, & Burgoon, 2003). Interactivity also implies longer exchanges and dynamic changes in deception displays that accompany them. Many interactive studies (e.g., Burgoon, Buller, Floyd, & Grandpre, 1996; Burgoon, Buller, White, Afifi, & Buslig, 1999; Burgoon & Qin, 2006; White & Burgoon, 2001) have demonstrated that deception becomes more difficult to detect over the course of a lengthy interaction, because deceivers adapt their communication over time to more closely approximate the demeanor and discourse of truth tellers.

Other factors that IDT posits are most likely to influence deception detection and credibility assessment include sender communication skill, expectedness of sender communication, receiver truth bias, and familiarity (Burgoon & Buller, 2015). Receivers are more likely to judge senders as credible and less likely to detect deception when senders are skilled communicators (Bond & DePaulo, 2006) and receivers are truth biased (Levine, Park, & McCornack, 1999). Skillful communication and ability to conform to expected communication patterns reduce the detectability of sender lies, aided and abetted by the receiver's truth biases.

Familiarity affects detection in conflicting ways. Relational familiarity, which refers to the degree of acquaintanceship between sender and receiver, may impair deception detection because receivers give friends and family the benefit of the doubt. Behavioral familiarity, which refers to how much the receiver knows about deception displays through experience or training, and informational familiarity, which refers to the amount of autobiographical and contextual information the receiver has about the sender, may supply receivers with the resources to be more discerning about a sender's duplicity (Buller & Burgoon, 1996). Stiff et al. (1989) introduced a fourth form of familiarity, situational familiarity, proposing that judges rely more on verbal than nonverbal details when senders' verbal accounts relate more to familiar than unfamiliar situations. Reinhard, Sporer, Scharmach, and Marksteiner (2011) found support for situational familiarity leading to more reliance on verbal cues and improved detection accuracy.

Two recent additions to the theoretical landscape include the probability model (PM; Levine, Clare, Green, Serota, & Park, 2014) and truth-default theory (TDT; Levine, 2014). The PM is not a theory per se but rather an observation that receiver judgments are influenced significantly by the base rate of truths and lies to which they are exposed. If judges are exposed to 20% lies and 80% truths, for example, their judgments should tend toward mostly truths and few lies, as compared to a base rate of 50% lies and 50% truths, which should yield a more balanced set of judgments. Affecting the influence of base rate on judgments is the truth bias. Because of it, even in the case of balanced truth and lie stimuli, judgments are partially calibrated to the judge's truth bias, yielding more truthful judgments than deceptive ones. The interpretation the authors give to their data is that the judgments are random and proportionate to the actual distribution of stimuli to which they are exposed. An alternative explanation is that judgments are a function of the actual amount of deceptive versus truthful verbal and nonverbal signals that judges process. These competing alternatives have yet to be unpacked and tested.

The TDT is a theory that expands on the "veracity effect" (Levine et al., 1999). The TDT follows the Spinozan view of information processing laid out by Gilbert, Krull, and Malone (1990) that presumes truth is the default judgment because it is more efficient to initially tag all incoming information as truth and there is usually insufficient evidence to pique suspicion. Unlike previous stances that saw the truth bias as a flaw in reasoning, the TDT suggests that truth default and truth bias are "functional, adaptive, and facilitate accuracy in non-research settings" because people are truthful more often than not (Levine, 2014, p. 381). That said, receivers may still judge a message as deceptive if given a reason to believe a person is lying based on the person's behavior, the receiver's assessment of sender motives, or other known facts.

Finally, other deception models that focus on the sending rather than receiving side of deception have tangential implications for detection. The leakage hypothesis (Ekman & Friesen, 1969), which proposes that deceivers give off inadvertent, uncontrolled autonomic "tells," and the four-factor model of deception (Zuckerman, DePaulo, & Rosenthal, 1981), which identifies physiological arousal, felt emotions, cognitive load, and attempted control as etiologies of deception displays, point to signals that could affect the detectability of deceit. To the extent that deception produces these inadvertent and uncontrolled signals, observers should be more successful in discerning duplicity. However, the lack of specific cues that are highly reliable and consistent across different communicators, and the potential of different causes to generate the same nonverbal signals, suggests that overemphasis on these indicators is not the key to detection accuracy.

Partitioning the detection accuracy variance

Why are humans so poor at detecting deception? Researchers have offered many explanations including the cognitive difficulty of the interactive deception task, the use of poor questioning techniques or focusing on the "wrong" cues, improper training with a lack of immediate feedback on judgment accuracies, the biases of the detectors, and others (Dunbar & Jensen, 2011; Vrij, 2008). Traditionally, most explanations of veracity judgments and deception detection accuracy have centered on factors affecting receiver judgments, often studied under the rubric of decoding accuracy. However, a fuller picture of lie/truth judgments can be gained by partitioning judgments into the array of factors that contribute to belief in another's veracity. Here we briefly examine some key factors according to whether they relate to the signal, receiver, sender, interpersonal relationship, or context (see also Porter, Campbell, Stapleton, & Birt, 2002).

Signal factors

Detection accuracy is profoundly influenced by the nature of the signal itself. Just as some emotions are more "readable" than others, so are some clues to deception more valid and diagnostic than others. One controversy in the literature has been whether judgment inaccuracy is due to judges relying on the wrong deception cues or there being few strong cues upon which to rely. Hartwig and Bond (2011) investigated this question using a Brunswikian lens model and four separate meta-analyses to estimate each link in the lens model. Their results and those of other meta-analyses (e.g., Malone, 2001; Vrij, 2008; Zuckerman & Driver, 1985) revealed that judges believed deception is best revealed by gaze aversion, signs of nervousness such as fidgeting, verbal statements lacking contextual details, speech errors (e.g., dysfluencies, long response latencies, pauses), abnormal behaviors, apparent indifference, or incompetence. Many of these indicators turn out to correlate with the ones that have been shown to be valid, although other valid indicators, especially verbal and auditory ones, escape judges' awareness (Burgoon, Guerrero, & Floyd, 2010). Hartwig and Bond (2011) concluded that inaccuracies are generally due to the lack of strong, individually diagnostic signals. In a subsequent meta-analysis, however, they did find that multiple cue combinations (rather than single cues) improve detection accuracy (Hartwig

& Bond, 2014), a position long held by other researchers and many practitioners. In addition, some potentially reliable indicators of deception have been omitted from recent meta-analyses either because there were too few studies of them or they are better measured with instrumentation than human judgment.

Another aspect of the judgmental accuracy problem is that signals themselves also vary in their degree of consistent and exclusive association with duplicity. It is a canon of deception detection that there is no "Pinocchio's nose" – no single surefire sign of truthfulness or deception that is universally present (Vrij, 2008). Moreover, when detection takes place in an interactive context, senders strategically manage their self-presentations to appear honest and truthful; they actively attempt to modify, minimize, and mask signals they believe will give them away (Hurley & Frank, 2011; Okubo, Kobayashi, & Ishikawa, 2012; Porter, ten Brinke, & Wallace, 2012). Deception signals are also dynamic rather than static, so that the complexion of deception fluctuates as interaction progresses (Burgoon & Buller, 2015). To the extent that deceiver and target are engaged in more than brief encounters, the iterative nature of conversation produces a complex and ever-changing set of signals that can overtax the judgment process, resulting in degraded detection accuracy.

Receiver factors

The receiver factors attracting the most attention are cognitive processes, individual differences, and the potential to improve accuracy through training or experience. As already noted, receivers often rely on mental shortcuts. Heuristic processing can sometimes lead to efficient, automatized, and effective decision making, but it can also result in biased, inaccurate judgments. Beyond the aforementioned truth bias, there is evidence that some judges, especially professional lie catchers, hold a lie bias – chronic skepticism and suspicion that result in an overestimate of dissembling (Meissner & Kassin, 2002). Other biases that have been linked to deception detection are a visual bias, in which over-attentiveness to visual cues results in overlooking more diagnostic auditory cues; a demeanor bias, in which judges are more inclined to attribute truthfulness to certain senders; expectancy violation bias, in which unexpected and anomalous behavior is viewed as a sign of deception; and a probing heuristic, in which the mere act of asking a question makes the questioner more inclined to believe the answers received (see Burgoon et al., 2008, and Vrij, 2006, for reviews). These biases and heuristic processing are more pronounced when judges have very brief exposures and make assessments at the beginning of interactions (Masip, Garrido, & Herrero, 2009). Notwithstanding, Hartwig and Bond (2011) concluded from their series of meta-analyses on why lie-catchers fail that "lie-catchers' intuitive notions about cues to deception are reasonably accurate" (p. 656). Judgmental failures must be partially attributed to causes other than misguided intuitions.

Although research related to receiver factors has focused largely on sources of inaccuracy, there has also been interest in sources of success. Early on, researchers believed that receivers vary in their detection skills, to the point of some qualifying as "wizards" who have exceptional detection ability (O'Sullivan & Ekman, 2004). This idea of superior judges sparked a lively debate about what constitutes acceptable evidence to confirm "exceptionalism" (see, e.g., Bond & Uysal, 2007; O'Sullivan, 2008). To investigate this matter more definitively, Bond and DePaulo (2008) conducted another meta-analysis of 247 statistical estimates related to individual differences in detection skill. The results showed that variability in objectively measured receiver detection ability is minuscule relative to receivers' subjective tendency to believe senders are truthful. Bond and DePaulo concluded that "the outcome of a deception judgment depends more on the liar's credibility than any other individual difference" (p. 477).

Another cognitive variable that has generated mixed empirical findings is suspicion. Suspicion is usually thought of as a state of uncertainty about another's veracity that falls in the intermediate ranges of a continuum from complete confidence in the other's truthfulness to a strong belief that the other is deceitful. Suspicion can be chronic and general in nature or situational and specific (see McCornack & Levine, 1990), the former being what contributes to a lie bias. It can be externally aroused, such as with an experimental induction, or arise naturally based on the sender's demeanor or statements. In the former case, it alters the distribution of truth-lie judgments and the cues upon which judges rely (Granhag & Strömwall, 2000). In the latter case, it can vary in degree, and those variations can alter the receiver's own demeanor (Burgoon, Buller, Ebesu, Rockwell, & White, 1996), which in turn can affect the communication patterns between the sender and receiver (Burgoon, Stern, & Dillman, 1995) and the relative credibility of the sender.

The empirical research on the interrelated factors of training, experience, and expertise has produced mixed findings. Many early investigations reported that these factors confer little benefit and boost confidence rather than detection acuity (Vrij, 2008). However, a number of criticisms have been lodged against the investigations that formed the basis for those conclusions. A meta-analysis found a small to medium training effect for detecting deception (Hauch, Sporer, Michael, & Meissner, 2014). Training was most beneficial when it emphasized verbal and content cues. Type of training, duration, and mode of instruction were additional moderators. Researchers have advanced numerous recommendations for achieving more effective training and have presented stronger evidence for positive training effects when those recommendations were followed (e.g., Bull, 2004; Crews, Cao, Lin, Burgoon, & Nunamaker, 2007; Frank & Feeley, 2003; ten Brinke & Porter, 2010).

Sender factors

As noted, the literature on sender demeanor and skill has shown definitively that these factors account for more variance in detection accuracy than receiver factors. Some senders are more skillful than others in their ability to adapt their presentations, control their behavior, and put forth an honest appearing demeanor (Burgoon, Buller, & Guerrero, 1995; Strömwall, Hartwig, & Granhag, 2006). Less skilled communicators are more "transparent" in their dissembling (Bond & DePaulo, 2008). Some of the factors that contribute to senders being more believable and evading detection when lying are nonverbal behaviors: kinesics, proxemics, vocal expressivity, dominance, immediacy, composure, and fluency (Burgoon et al., 2010). Verbal expressivity and control of the verbal stream are also part of being a socially skilled communicator (Riggio, 1986). On the basis of their five experiments, Levine et al. (2011) endorsed "the strong impact of sender effects in deception detection" (p. 377).

Relationship factors

Limited research and theorizing has considered how the interpersonal relationship between sender and receiver (e.g., familiar or unfamiliar, trusting or distrusting) affects detection accuracy (e.g., Burgoon & Buller, 2015; Burgoon, Buller, Ebesu, & Rockwell, 1994; Millar & Millar, 1995). For example, greater familiarity can hinder the ability to recognize deceit, with intimate partners being inclined to see a loved one through rose colored glasses and discount signs of deceit (Stiff, Kim, & Ramesh, 1992). Intimate partners might also discount the lies of their partners even when they are detected for the sake of saving the relationship (Knapp, 2006). Relational familiarity also affects the modalities on which detectors rely when forming judgments. If romantic partners rely more on one another's face to make judgments, then they may be less accurate because facial expressions are not particularly diagnostic.

Relational familiarity might lead to a greater truth bias as well, which also negatively impacts accuracy.

Other relationship factors that may affect detection accuracy but have received little empirical attention are interpersonal attractiveness, power differentials and the gender combination of judge and target (Porter et al., 2002). For example, the power relationship between the two partners might change the types of deceptions used as well as the acceptability of calling one's interaction partner a liar. In the doctor–patient relationship, patients and physicians have different reasons to lie to one another but patients might lie to preserve their autonomy and restore lost power (Fainzang, 2002). In the workplace, managers might feel more entitled to lie to their subordinates than vice versa, which might make those lies more difficult to detect (Lindsey, Dunbar, & Russell, 2011).

Communication context factors

Other factors related to the interaction between sender and receiver that have substantial influence on detection accuracy include the communication context itself (e.g., interactive or noninteractive, prepared or unprepared discourse, structured or unstructured type of discourse, dyadic or group interaction), communication medium (e.g., face-to-face, text-based, audio, videoconferencing), and topic (see, e.g., Bond & DePaulo, 2006, 2008; Hartwig & Bond, 2011; Vrij, 2008). For example, the medium through which communication takes place affects how many channels a deceiver must monitor and control. Ekman and Friesen (1969) identified controllability as one of the key considerations in what kinds of telltale signs would become evident during deceit. Hocking and Leathers (1980) developed the controllability hierarchy into a theory of nonverbal behavioral monitoring and controllability, such that more easily controlled and monitored communication channels such as the face should show fewer deception signals than less controlled and monitored channels such as the voice. Their test of the theory supported the voice as a more telltale channel. In face-to-face contexts, deceivers must manage the full gamut of possible telltale signs, whereas with mediated communication, fewer channels must be regulated. Text affords deceivers the greatest control of all, which may partially account for the poor ability of receivers to detect deceit from written texts and transcripts of oral communication.

In recent years, more attention has turned to the interviewing context and the questioning that occurs, including question types (e.g., closed-ended or open-ended ones that provoke greater cognitive load or elicit potentially diagnostic behaviors), expectedness of the questions, interviewer style (e.g., rapport building or accusatory), interviewer strategy (e.g., strategic use of evidence, active questioning rather than passive observation, requesting retelling of narratives from multiple perspectives), and the availability of other contextual information for interpreting responses (e.g., Burgoon et al., 1994; Dunbar et al., 2003; Geiselman, 2012; Hartwig et al., 2011; Levine, Blair, & Clare, 2014). For example, participation in dialogue rather than mere observation leads judges to tilt toward leniency rather than skepticism. Not surprisingly, judgments made in a vacuum are less accurate than ones that take advantage of contextual knowledge or respondents' baseline behaviors. When known evidence is disclosed to suspects, question wording and many other subtle facets of questions can also produce powerful effects (see Vrij & Granhag, 2012, for a discussion of these factors).

In short, truth/lie judgments are a function of many factors beyond those associated just with the receiver. Signal, sender, relationship, and context factors all exert substantial influence.

Detection methods and associated accuracies

There are essentially two methods of assessing the truthfulness of another's statements: unassisted human judgments and judgments assisted by instrumentation. That humans are prone to biases and have limits on their perceptions and interpretations makes it difficult for them to detect deception in an interactive context when they are already cognitively busy with the interaction itself. Although unaided deception detection accuracy is low, promising new technologies offer hope that humans can improve their accuracy with assistance when it is feasible. The fact remains that in most of the situations, using technological devices is neither practical nor efficient for everyday interactions. This section reviews some methods to improve unaided detection and instruments available to aid human judgment.

Unaided human judgments

One standard approach to improving human judgment is through training in the use of valid deception indicators and avoidance of stereotypical but unreliable ones. Training has been shown to have modest effects overall (in the range of 5–10% improvement). The most effective training offers immediate feedback on accuracy, focuses on the verbal content cues rather than nonverbal cues, includes both lecture and practice components, and lasts longer than 20 minutes per session (Hauch et al., 2014). One promising avenue of research for unaided humans is to improve the questioning techniques that are used. The strategic use of evidence (SUE) technique uses open-ended questions and asks about possibly incriminating evidence (Vrij, Granhag, Mann, & Leal, 2011) while the controlled cognitive engagement (CCE) interview protocol (Ormerod & Dando, 2015) builds on the cognitive approach proffered by Vrij and colleagues (e.g., Vrij et al., 2011) and uses rapport-building techniques and unanticipated questions to catch liars off guard.

Another way that may improve detection accuracy is the indirect method. Instead of judges being directly asked to determine if someone is lying, they are asked to assess the person's cognitive load, apparent nervousness, or other factors from which deceptiveness can be inferred indirectly. For example, Vrij, Edward, and Bull (2001) found that police officers were better able to distinguish between truths and lies when using an indirect detection method rather than direct veracity judgments. Other studies in which comparisons were made between direct and indirect measures have similarly found greater accuracy using indirect methods. One such indirect method developed by Vrij and colleagues is the "rapid judgments" (RJ) approach. The method borrows cues from two well-known coding systems, criteria-based content analysis (CBCA) (Steller & Köhnken, 1989) and reality monitoring (RM) (Johnson & Raye, 1981). It trains observers to look for cues that are related to the fact that lying is more cognitively taxing than telling the truth, resulting in liars exhibiting a lack of detail in their verbal statements and more nonverbal "leakage" (i.e., inadvertently revealed signs of deceit) in the form of longer latency periods between question and answer, fewer illustrator gestures that accompany speech, fewer hand/finger movements, and more speech hesitations (Dunbar, Harvell, Jensen, Burgoon, & Kelley, 2012; Griesel, Ternes, Schraml, Cooper, & Yuille, 2013; Masip, Sporer, Garrido, & Herrero, 2005). Perceivers who make a direct judgment after coding with the RJ method attain higher accuracy than when they use the direct method alone (Dunbar et al., 2012; Vrij, Evans, Akehurst, & Mann, 2004).

Instrument-aided assessment

Devices, software, and sophisticated data analysis methods to detect deception are becoming more commonplace. The best known of these is the polygraph, which monitors physiological indicators of stress and arousal through the use of sensors that measure respiration, heart rate, blood pressure, and electrodermal conductance during questioning. The polygraph is used primarily in three main areas: event-specific investigations such as after a crime, in periodic employee reviews, and in preemployment screening to determine employment suitability. The bulk of evidence for the effectiveness of the polygraph comes primarily from event-specific uses. A comprehensive review of the effectiveness of the polygraph conducted by the National Research Council (2003) concluded that both laboratory and field tests of the polygraph have serious flaws but discriminate truth from deception at above chance levels in event-specific testing. A meta-analysis of 38 event-specific studies conducted by the American Polygraph Association (Gougler et al., 2011) determined that the comparison question technique used in such testing produced an aggregated decision accuracy rate of 89% and that the various validated types of polygraph tests had accuracies ranging from 82% to 99% (excluding outliers). However, the NRC warned that the polygraph is susceptible to countermeasures that render the test invalid by a motivated deceiver. In security screening, the NRC concluded that the polygraph was an "unacceptable choice" because of the risk that loyal employees would be falsely judged deceptive, while major security threats would go undetected. Vrij (2008) and others have argued that the polygraph fails the scientific standard for admissibility in American courts. Concerns about the reliability of the polygraph have prompted several other technologies to emerge.

Additional contact devices such as functional magnetic resonance imaging (fMRI), which uses MRI technology to examine blood flow in the brain, and electroencephalography (EEG), which records electrical activity along the scalp, have also been used to detect deception. Both methods offer promise because brain activity should be less vulnerable to active manipulation (known as countermeasures) than the physiological cues of the polygraph but are difficult to implement and require specialized knowledge and expensive equipment. In a review of the fMRI literature, Langleben (2008) found that increased activation of the prefrontal-parietal region of the brain was observed in most studies across experimental paradigms involving deception. EEG studies that utilize *event-related potentials* (ERPs; waveforms averaged over numerous trials time-locked to an event) have found that ERPs are different when participants are truthful versus when they are deceptive (Sun & Luo, 2008). One of the most common uses of ERPs in deception detection is the P300-based methodology, which analyzes a particular waveform that is responsive to familiarity with crime-relevant images and information. It finds that guilty respondents show greater P300 amplitudes than innocent respondents (see, e.g., Hahm et al., 2009; Rosenfeld, 2010). Others have critiqued the P300 because the neurophysiological response occurs prior to rather than after the conscious experiencing of lying (and thus may have a questionable link to deception) and have advocated other methods (Thornton, 2005). One significant drawback to the EEG/ERP methods is they rely on keyboard rather than oral responses, which may not be equivalent to typical uses for deception detection techniques.

Other tools such as voice stress analysis do not apply sensors to the body. Although commercial voice stress tools have been criticized as unreliable (Damphousse, 2008), the voice can be used to detect deception effectively when the correct features are examined (Elkins, 2011; Sporer & Schwandt, 2006). Among those features are fundamental frequency (known as pitch), pitch slope, pitch variance, amplitude (perceived as loudness), variance in amplitude, tempo, hesitations, and dysfluencies.

In the area of oculometrics, eye trackers are used to measure pupil dilation, blinks, eye gaze, eye fixations, and eye saccades (movements between fixation points). Blinks and latencies in blinks, following startle-inducing noises, distinguish truthful from deceptive responses (Leal & Vrij, 2008; Verschuere, Crombez, Koster, Van Bockstaele, & De Clercq, 2007) because they are linked to arousal, defensive, and orientation responses. Pupillary responses have been shown to differentiate truth from deception because pupil dilation is related to both arousal and emotional state. Guilty suspects who lie show greater pupil dilation than innocent suspects who tell the truth (Lubow & Fein, 1996).

Thermal imaging of the face, which detects greater flushing and temperature changes in guilty suspects, was touted as a rival to the polygraph in an often-cited *Nature* article (Pavlidis, Eberhardt, & Levine, 2002) but has been found to be less effective than trained interviewers in other studies (Warmelink et al., 2011). None of these new technologies to measure physiological changes, still in their infancy, has proven to be a rival to the polygraph as of yet; the polygraph currently is the most trusted tool despite its drawbacks (see the report on the state of the polygraph published by the National Research Council in 2003).

One promising development is software to automate linguistic analysis. Linguistic Inquiry and Word Count (LIWC) is a dictionary-based text analysis software program that counts categories of words in speech such as pronouns, perceptions, emotion states, and thought processes (Pennebaker, Francis, & Booth, 2001). Early research produced a mix of findings (Jensen, Bessarabova, Adame, Burgoon, & Slowik, 2011; Newman, Pennebaker, Berry, & Richards, 2003; Vrij, Mann, Kristen, & Fisher, 2007), leading one review (Ali & Levine, 2007) to conclude that no linguistic cue consistently comes out in the same direction across studies, and opposite-direction findings are commonplace. Nevertheless, LIWC has gained widespread adoption, and a more recent meta-analysis found that linguistic cues successfully discriminate truth from deception with LIWC categories such as exclusive words, pronouns, time- and space-related words, emotion words, motion verbs, and negations (see the LIWC homepage at www.liwc.net for a thorough description of these terms).

Other methods for automated linguistic analysis combine parsers that tag parts of speech, dictionaries, and algorithms for established indices. For example, Structured Programming for Linguistic Cue Extraction (SPLICE; Moffitt, Giboney, Ehrhardt, Burgoon, & Nunamaker, 2012), which originated as the software program called Agent99, uses a computer-based tool to parse and extract parts of speech. These, along with several dictionaries and formulas for established indices such as readability, are then combined into indices. SPLICE has validated a wide variety of linguistic dimensions such as quantity, complexity, diversity, specificity, personalism, immediacy–nonimmediacy, and hedging/uncertainty that distinguish truth from deception (Burgoon, Twitchell, Nunamaker, & Zhou, 2004; Fuller, Biros, Burgoon, & Nunamaker, 2013; Hancock, Curry, Goorha, & Woodworth, 2008). Both LIWC and SPLICE are relatively recent innovations, and different studies of them have used different tests, questioning styles, and populations; therefore it is too early to draw definitive conclusions about the directionality and diagnosticity of linguistic markers for deception.

Summary

It should be apparent from this review that deception detection is a complicated endeavor, not unlike a Rubik's cube in the complex assembly and alignment of all the relevant variables. It is a process better described as opaque than transparent, elusive rather than straightforward. Nevertheless, research has successfully uncovered numerous diagnostic indicators and the conditions under which those indicators are valid. This last point is worth underscoring – there are no universal signals nor completely context-independent indicators. The task of detecting deception is thus contingent on knowing the factors that are operative during a given detection episode. That said, if the context in which deception indicators are observed is properly specified, it should be possible to say which indicators will effectively discriminate and which will not *under those conditions*. For example, when deceivers have opportunities to prepare or edit their messages, judgments of veracity will be more challenging than when deceivers produce messages on the spot. The former condition lends itself to persuasive deception, in which deceivers can go on the offense, be more loquacious and dominant, and polish their communication so as to appear normal and credible. The latter circumstance lends itself to opting for a reticent, submissive, and obsequious demeanor that deflects blame and suspicion.

Detection success does not rely exclusively on the innate or trained abilities of the judge but rather also on signal, sender, relationship, and communication context factors. The likelihood of detecting deceit is lower, for example, with skillful senders, a trusted acquaintance, reliance on facial cues, and longer interactions. It is therefore unsurprising that more often than not the advantage in evading detection goes to the deceptive sender rather than the receiver.

In recent decades, great advances have been made in the methods for behavioral observation of the verbal, nonverbal, and physiological indicators of veracity, and technologies have come online for automating measurement and aiding human judgment. The accuracy of judgments of truths and lies is therefore likely to improve with the adoption of these methods and tools. In their summary of the research on the polygraph, the National Research Council (2003) concluded that "the research program should be open to supporting alternative ways of looking at the problems of deterrence and detection because there is no single research approach that clearly has the most promise for meeting national security objectives" (p. 9). More broadly, just as there is no Pinocchio's nose for detecting deception, we agree that there is no one valid approach to the study of truth/lie accuracy. A combination of detection methods rather than a single method will likely prove the most fruitful in the future.

References

Aamodt, M. G., & Custer, H. (2006). Who can best catch a liar? A meta-analysis of individual differences in detecting deception. *Forensic Examiner*, 15, 6–11.

Ali, M., & Levine, T. (2007). The language of truthful and deceptive denials and confessions. *Communication Reports*, 20, 82–91.

Blair, J. P., Levine, T. R., & Shaw, A. S. (2010). Content in context improves deception detection accuracy. *Human Communication Research*, 36, 423–442.

Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, 10, 214–234.

Bond, C. F., & DePaulo, B. M. (2008). Individual differences in judging deception: Accuracy and bias. *Psychological Bulletin*, 134, 477–492.

Bond, C. F., & Uysal, A. (2007). On lie detection "wizards." *Law and Human Behavior*, 31, 109–115.

Bull, R. (2004). Training to detect deception from behavioral cues: Attempts and problems. In P. A. Granhag & L. A. Strömwall (Eds.), *The detection of deception in forensic contexts* (pp. 251–268). Cambridge, UK: Cambridge University Press.

Buller, D. B., & Burgoon, J. K. (1996). Interpersonal deception theory. *Communication Theory*, 6, 203–242.
Buller, D. B., Strzyzewski, K. D., & Hunsaker, F. G. (1991). Interpersonal deception: II. The inferiority of conversational participants as deception detectors. *Communication Monographs*, 58, 25–40.
Burgoon, J. K. (2005). The future of motivated deception detection. In P. Kalbfleisch (Ed.), *Communication Yearbook 29* (pp. 49–95). Mahway, NJ: Erlbaum.
Burgoon, J. K., Blair, J. P., & Strom, R. E. (2008). Cognitive biases and nonverbal cue availability in detecting deception. *Human Communication Research*, 34, 572–599.
Burgoon, J. K., & Buller, D. B. (2015). Interpersonal deception theory: Purposive and interdependent behavior during deceptive interpersonal interactions. In D. O. Braithwaite & P. Schrodt (Eds.), *Engaging theories in interpersonal communication*, 2nd ed. (pp. 349–362). Los Angeles, CA: Sage.
Burgoon, J. K., Buller, D. B., Ebesu, A., & Rockwell, P. (1994). Interpersonal deception: V. Accuracy in deception detection. *Communication Monographs*, 61, 303–325.
Burgoon, J. K., Buller, D. B., & Floyd, K. (2001). Does participation affect deception success? A test of the inter-activity effect. *Human Communication Research*, 27, 503–534.
Burgoon, J. K., Buller, D. B., Floyd, K., & Grandpre, J. (1996). Deceptive realities: Sender, receiver, and observer perspectives in deceptive conversations. *Communication Research*, 23, 724–748.
Burgoon, J. K., Buller, D. B., & Guerrero, L. K. (1995). Interpersonal deception: IX. Effects of social skill and nonverbal communication on deception success and detection accuracy. *Journal of Language and Social Psychology*, 14, 289–311.
Burgoon, J. K., Buller, D. B., Ebesu, A., Rockwell, P., & White, C. (1996). Testing interpersonal deception theory: Effects of suspicion on nonverbal behavior and relational messages. *Communication Theory*, 6, 243–267.
Burgoon, J. K., Buller, D. B., White, C. H., Afifi, W., & Buslig, A. L. S. (1999). The role of conversational involvement in deceptive interpersonal interactions. *Personality and Social Psychology Bulletin*, 25, 669–685.
Burgoon, J. K., & Floyd, K. (2000). Testing for the motivation impairment effect during deceptive and truthful interaction. *Western Journal of Communication*, 64, 243–267.
Burgoon, J. K., Guerrero, L., & Floyd, K. (2010). *Nonverbal communication*. Boston, MA: Allyn & Bacon.
Burgoon, J. K., & Qin, T. (2006). The dynamic nature of deceptive verbal communication. *Journal of Language and Social Psychology*, 25, 76–96.
Burgoon, J. K., Stern, L. A., & Dillman, L. (1995). *Interpersonal adaptation: Dyadic interaction patterns*. New York: Cambridge University Press.
Burgoon, J. K., Twitchell, D. P., Nunamaker, J. F., & Zhou, L. (2004). Automating linguistics-based cues for detecting deception in text-based

asynchronous computer-mediated communication. *Group Decision and Negotiation*, 13, 81–106.
Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, 39, 752–766.
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.
Coleman, L., & Kay, P. (1981). Prototype semantics: The English word lie. *Language*, 57, 26–44.
Crews, J. M., Cao, J., Lin, M., Burgoon, J. K., & Nunamaker, J. F. (2007). A comparison of instructor-led vs. web-based training for deception detection. *Journal of STEM (Science, Technology, Engineering, and Math) Education: Innovations and Research (JSTEM)*, 8, 1–10.
Damphousse, K. R. (2008). Voice stress analysis: Only 15 percent of lies about drug use detected in field test. *NIJ Journal*, 259, 8–12.
DePaulo, B. M., Charlton, K., Cooper, H., Lindsay, J. J., & Muhlenbruck, L. (1997). The accuracy-confidence correlation in the detection of deception. *Personality and Social Psychology Review*, 1, 346–357.
DePaulo, B. M., & Kirkendol, S. E. (1989). The motivational impairment effect in the communication of deception. In J. C. Yuille (Ed.), *Credibility assessment* (pp. 51–70). New York: Kluwer Academic/Plenum Publishers.
DePaulo, B.M., Lanier, K., & Davis, T. (1983). Detecting the deceit of the motivated liar. *Journal of Personality and Social Psychology*, 45, 1096–1103.
Dunbar, N. E., Harvell, L. A., Jensen, M. L., Burgoon, J. K., & Kelley, K. M. (2012). *The viability of using rapid judgments as a method of deception detection*. Paper presented at the 45th Hawaiian International Conference on System Sciences, Maui, HI.
Dunbar, N. E., & Jensen, M. L. (2011). Digital deception in personal relationships. In K. B. Wright & L. M. Webb (Eds.), *Computer-mediated communication in personal relationships* (pp. 324–343). New York: Peter Lang.
Dunbar, N. E., Jensen, M. L., Burgoon, J. K., Kelley, K. M., Harrison, K. J., Adame, B. J., & Bernard, D. R. (2015). Effects of veracity, modality, and sanctioning on credibility assessment during mediated and unmediated interviews. *Communication Research*. Vol. 42(5) 649–674. DOI: 10.1177/0093650213480175.
Dunbar, N. E., Ramirez, A., & Burgoon, J. K. (2003). Interactive deception: Effects of participation on participant-receiver and observer judgments. *Communication Reports*, 16, 23–33.
Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, usage, and coding. In T. A. Sebeok (Ed.), *Semiotica* (pp. 49–98). London: Mouton.
Elkins, A. C. (2011). *Vocalic markers of deception and cognitive dissonance for automated emotion detection systems*. Unpublished doctoral dissertation, University of Arizona, Tucson, AZ.

Fainzang, S. (2002). Lying, secrecy and power within the doctor-patient relationship. *Anthropology & Medicine*, 9, 117–133.

Frank, M. G., & Feeley, T. H. (2003). To catch a liar: Challenges for research in lie detection training. *Journal of Applied Communication Research*, 31, 58–75.

Fuller, C. M., Biros, D. P., Burgoon, J. K., & Nunamaker, J. F. (2013). An examination and validation of linguistic constructs for studying high-stakes deception. *Group Decision and Negotiation*, 22, 117–134.

Geiselman, R. (2012). The cognitive interview for suspects (CIS). *American Journal of Forensic Psychology*, 30, 1–17.

Gilbert, D. T., Krull, D. S., & Malone, P. S. (1990). Unbelieving the unbelievable: Some problems in the rejection of false information. *Journal of Personality and Social Psychology*, 59, 601–613.

Gougler, M., Nelson, R., Handler, M., Krapohl, D., Shaw, P., & Bierman, L. (2011). Meta-analytic survey of criterion accuracy of validated polygraph techniques. *Polygraph*, 40, 194–305.

Granhag, P. A., & Strömwall, L. A. (2000). Effects of preconceptions on deception detection and new answers to why lie-catchers often fail. *Psychology, Crime and Law*, 6, 197–218.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Oxford: John Wiley.

Griesel, D., Ternes, M., Schraml, D., Cooper, B. S., & Yuille, J. C. (2013). The ABC's of CBCA: Verbal credibility assessment in practice. In B. S. Cooper, D. Griesel, & M. Ternes (Eds.), *Applied issues in investigative interviewing, eyewitness memory, and credibility assessment* (pp. 293–323). New York: Springer.

Grice, H. P. (1989). *Studies in the ways of words*. Cambridge, MA: Harvard University Press.

Hahm, J., Ji, H. K., Jeong, J. Y., Oh, D. H., Kim, S. H., Sim, K. B., & Lee, J. H. (2009). Detection of concealed information: Combining a virtual mock crime with a P300-based Guilty Knowledge Test. *CyberPsychology & Behavior*, 12, 269–275.

Hall, J. A., Andrzejewski, S. A., & Yopchick, J. E. (2009). Psychosocial correlates of interpersonal sensitivity: A meta-analysis. *Journal of Nonverbal Behavior*, 33, 149–180.

Hancock, J. T., Curry, L., Goorha, S., & Woodworth, M. T. (2008). On lying and being lied to: An automated linguistic analysis of deception. *Discourse Processes*, 46, 1–23.

Hartwig, M., & Bond, C. F. (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychological Bulletin*, 137, 643–659.

Hartwig, M., & Bond, C. F. (2014). Lie detection from multiple cues: A meta-analysis. *Applied Cognitive Psychology*, 28, 617–813.

Hartwig, M., Granhag, P.-A., Stromwall, L., Wolf, A. G., Vrij, A., & Hjelmsäter, E. R. (2011). Detecting deception in suspects: Verbal cues as a function of interview strategy. *Psychology, Crime & Law*, 17, 643–656.

Hauch, V., Sporer, S. L., Michael, S. W., & Meissner, C. A. (2014). Does training improve the detection of deception? A meta-analysis. *Communication Research*. Advance online publication.

Hocking, J. E., & Leathers, D. G. (1980). Nonverbal indicators of deception: A new theoretical perspective. *Communications Monographs*, 47, 119–131.

Hurley, C. M., & Frank, M. G. (2011). Executing facial control during deceptive situations. *Journal of Nonverbal Behavior*, 35, 119–131.

Jensen, M. L., Bessarabova, E., Adame, B., Burgoon, J. K., & Slowik, S. M. (2011). Deceptive language by innocent and guilty criminal suspects: The influence of dominance, question, and guilt on interview responses. *Journal of Language and Social Psychology*, 30, 357–375.

Johnson, M. K. & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, 88, 67–85.

Knapp, M. L. (2006). Lying and deception in close relationships. In A. L. Vangelisti & D. Perlman (Eds.), *The Cambridge handbook of personal relationships* (pp. 517–532). New York: Cambridge University Press.

Knapp, M. L. (2008). *Lying and deception in human interaction*. Boston, MA: Pearson.

Langleben, D. D. (2008). Detection of deception with fMRI: Are we there yet? *Legal and Criminological Psychology*, 13, 1–9.

Leal, S., & Vrij, A. (2008). Blinking during and after lying. *Journal of Nonverbal Behavior*, 32, 187–194.

Levine, T. R. (2014). Truth-Default Theory (TDT): A theory of human deception and deception detection. *Journal of Language and Social Psychology*, 33, 378–392.

Levine, T. R., Blair, J. P., & Clare, D. D. (2014). Diagnostic utility: Experimental demonstrations and replications of powerful question effects in high-stakes deception detection. *Human Communication Research*, 40, 262–289.

Levine, T. R., Clare, D. D., Green, T., Serota, K. B., & Park, H. S. (2014). The effects of truth–lie base rate on interactive deception detection accuracy. *Human Communication Research*, 40, 350–372.

Levine, T. R., Park, H. S., & McCornack, S. A. (1999). Accuracy in detecting truths and lies: Documenting the "veracity effect." *Communications Monographs*, 66, 125–144.

Levine, T. R., Serota, K. B., Shulman, H., Clare, D. D., Park, H. S., Shaw, A. S., Shim, J. C., & Lee, J. H. (2011). Sender demeanor: Individual differences in sender believability have a powerful impact on deception detection judgments. *Human Communication Research*, 37, 377–403.

Lindsey, L. L. M., Dunbar, N. E., & Russell, J. (2011). Risky business or managed event? Perceptions of power and deception in the workplace. *Journal of Organizational Culture, Communications and Conflict*, 15, 55–79.

Lubow, R., & Fein, O. (1996). Pupillary size in response to a visual guilty knowledge test: New technique for the detection of deception. *Journal of Experimental Psychology: Applied*, 2, 164.

Malone, B. E. (2001). *Perceived cues to deception: A meta-analytic review (Unpublished master's thesis)*. University of Virginia, Charlottesville, VA.

Masip, J., Garrido, E., & Herrero, C. (2004). Defining deception. *Anales de Psicología*, 20, 147–171.

Masip, J., Garrido, E., & Herrero, C. (2009). Heuristic versus systematic processing of information in detecting deception: Questioning the truth bias. *Psychological Reports*, 105, 11–36.

Masip, J., Sporer, S. L., Garrido, E., & Herrero, C. (2005). The detection of deception with the reality monitoring approach: A review of the empirical evidence. *Psychology, Crime & Law*, 11, 99–122.

McCornack, S. A. (1992). Information manipulation theory. *Communication Monographs*, 59, 203–242.

McCornack, S. A., & Levine, T. R. (1990). When lovers become leery: The relationship between suspicion and accuracy in detecting deception. *Communications Monographs*, 57, 219–230.

McCornack, S. A., Morrison, K., Paik, J. E., Wisner, A. M., & Zhu, X. (2014). Information manipulation theory 2: A propositional theory of discourse production. *Journal of Language and Social Psychology*, 33, 348–377.

McCroskey, J. C. (1966). Scales for measurement of ethos. *Speech Monographs*, 33, 65–72.

Meissner, C. A., & Kassin, S. M. (2002). "He's guilty!": Investigator bias in judgments of truth and deception. *Law and Human Behavior*, 5, 469–480.

Millar, M., & Millar, K. (1995). Detection of deception in familiar and unfamiliar persons: The effects of information restriction. *Journal of Nonverbal Behavior*, 19, 69–84.

Moffitt, K., Giboney, J., Ehrhardt, E., Burgoon, J. K., & Nunamaker, J. F. (2012). *Structured programming for linguistic cue extraction*. Proceedings of the 45th Hawaii International Conference on System Sciences, Maui, January.

National Research Council. (2003). *The polygraph and lie detection*. Washington, DC: National Academies Press.

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, 29, 665–675.

Okubo, M., Kobayashi, A., & Ishikawa, K. (2012). A fake smile thwarts cheater detection. *Journal of Nonverbal Behavior*, 36, 217–225.

Ormerod, T. C., & Dando, C. J. (2015). Finding a needle in a haystack: Toward a psychologically informed method for aviation security screening. *Journal of Experimental Psychology: General*, 144, 76–84.

O'Sullivan, M. (2008). Home runs and humbugs: Comment on Bond and DePaulo (2008). *Psychological Bulletin*, 134, 493–497.

O'Sullivan, M., & Ekman, P. (2004). The wizards of deception detection. In P. A. Granhag & L. A. Strömwall (Eds.), *Deception detection in forensic contexts* (pp. 269–286). Cambridge, UK: Cambridge University Press.

O'Sullivan, M., Frank, M. G., Hurley, C. M., & Tiwana, J. (2009). Police lie detection accuracy: The effect of lie scenario. *Law and Human Behavior*, 33, 530–538.

Pavlidis, I., Eberhardt, N. L., & Levine, J. A. (2002). Human behaviour: Seeing through the face of deception. *Nature*, 415, 35.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic Inquiry and Word Count (LIWC): LIWC2001*. Mahwah, NJ: Lawrence Erlbaum Associates.

Porter, S., Campbell, M. A., Stapleton, J., & Birt, A. R. (2002). The influence of judge, target, and stimulus characteristics on the accuracy of detecting deceit. *Canadian Journal of Behavioural Science*, 34, 172–185.

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, 477–491.

Porter, S., ten Brinke, L., & Wallace, B. (2012). Secrets and lies: Involuntary leakage in deceptive facial expressions as a function of emotional intensity. *Journal of Nonverbal Behavior*, 36, 23–37.

Reinhard, M.-A., Sporer, S. L., Scharmach, M., & Marksteiner, T. (2011). Listening, not watching: Situational familiarity and the ability to detect deception. *Journal of Personality and Social Psychology*, 101, 467–484.

Riggio, R. E. (1986). Assessment of basic social skills. *Journal of Personality and Social Psychology*, 51, 649.

Rosenfeld, J. P. (2010). P300 in detecting concealed information. In B. Verschuere, G. B. Shakhar, & E. Meijer (Eds.), *Memory detection: Theory and application of the Concealed Information Test* (pp. 63–89). Cambridge, UK: Cambridge University Press.

Sporer, S. L., & Schwandt, B. (2006). Paraverbal indicators of deception: A meta-analytic synthesis. *Applied Cognitive Psychology*, 20, 421–446.

Sporer, S. L., & Schwandt, B. (2007). Moderators of nonverbal indicators of deception: A meta-analytic synthesis. *Psychology, Public Policy, and Law*, 13, 1–34.

Steller, M., & Köhnken, G. (1989). Criteria-based content analysis. In D. C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence* (pp. 217–245). New York: Springer Verlag.

Stiff, J. B., Kim, H. J., & Ramesh, C. N. (1992). Truth biases and aroused suspicion in relational deception. *Communication Research*, 19, 326–345.

Stiff, J. B., Miller, G. R., Sleight, C., Mongeau, P., Garlick, R., & Rogan, R. (1989). Explanations for visual cue primacy in judgments of honesty and deceit. *Journal of Personality and Social Psychology*, 56, 555–564.

Sun, S.-Y., & Luo, Y.-J. (2008). Feedback-related negativity in outcome evaluation with a deception task. *Acta Psychologica Sinica*, 40, 693–700.

ten Brinke, L., & Porter, S. (2010).The truth about lies: What works in detecting high-stakes deception? *Legal and Criminological Psychology*, 15, 57–75.

Thornton, K. E. (2005). The qEEG in the lie detection problem: The localization of guilt? *Journal of Neurotherapy*, 9, 31–43.

Verschuere, B., Crombez, G, Koster, E.H., Van Bockstaele, B., & De Clercq, A. (2007). Startling secrets: Startle eye blink modulation by concealed crime information. *Journal of Biological Psychology*, 76, 52–60.

Vrij, A. (2004). Why professionals fail to catch liars and how they can improve. *Legal and Criminological Psychology*, 9, 159–181.

Vrij, A. (2006). Nonverbal communication and deception. In V. Manusov & M. L. Patterson (Eds.), *The Sage handbook of nonverbal communication* (pp. 341–359). Thousand Oaks, CA: Sage.

Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities* (2nd ed.). Chichester, UK: Wiley.

Vrij, A., & Baxter, M. (1999). Accuracy and confidence in detecting truths and lies in elaborations and denials: Truth bias, lie bias and individual differences. *Expert Evidence*, 7, 25–36.

Vrij, A., Edward, K., & Bull, R. (2001). Police officers' ability to detect deceit: The benefits of indirect deception detection. *Legal and Criminological Psychology*, 6, 185–196.

Vrij, A., Evans, H., Akehurst, L., & Mann, S. (2004). Rapid judgements in assessing verbal and nonverbal cues: Their potential for deception researchers and lie detection. *Applied Cognitive Psychology*, 18, 283–296.

Vrij, A., & Granhag, P. A. (2012). Eliciting cues to deception and truth: What matters are the questions asked. *Journal of Applied Research in Memory and Cognition*, 1, 110–117.

Vrij, A., Granhag, P. A., Mann, S., & Leal, S. (2011). Outsmarting the liars: Toward a cognitive lie detection approach. *Current Directions in Psychological Science*, 20, 28–32.

Vrij, A., Leal, S., Mann, S., & Fisher, R. (2012). Imposing cognitive load to elicit cues to deceit: Inducing the reverse order technique naturally. *Psychology, Crime & Law*, 18(6), 579–594.

Vrij, A., Mann, S., Kristen, S., & Fisher, R. P. (2007). Cues to deception and ability to detect lies as a function of police interview styles. *Law and Human Behavior*, 31, 499–518.

Vrij, A., Mann, S., Robbins, E., & Robinson, M. (2006). Police officers' ability to detect deception in high stakes situations and in repeated lie detection tests. *Applied Cognitive Psychology*, 20, 741–755.

Warmelink, L., Vrij, A., Mann, S., Leal, S., Forrester, D., & Fisher, R. P. (2011). Thermal imaging as a lie detection tool at airports. *Law and Human Behavior*, 35, 40.

White, C. H., & Burgoon, J. K. (2001). Adaptation and communicative design: Patterns of interaction in truthful and deceptive conversations. *Human Communication Research*, 27, 9–37.

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 14, pp. 1–57). New York: Academic Press.

Zuckerman, M., & Driver, R. (1985). Telling lies: Verbal and nonverbal correlates of deception. In A. W. Siegman & S. Feldstein (Eds.), *Multichannel integrations of nonverbal behavior* (pp. 129–148). Hillsdale, NJ: Lawrence Erlbaum Associates.

5 Accuracy of judging personality

Mitja D. Back and Steffen Nestler

Abstract

Judging other people's personality is a widespread social phenomenon early on in the acquaintance process. The accuracy of these interpersonal impressions colors the way we select, shape, and maintain our social environments. In this chapter, we give an overview of the state of the art in research on the accuracy of personality judgments. First, we describe and discuss existing methodological alternatives (variable- vs. person-centered approaches, choice of accuracy criteria, individual vs. aggregated perceiver approaches). Second, we tackle the question of how well humans can judge the personality of unknown others, summarizing the wealth of existing studies across a large variety of contexts. Third, following a lens model approach, we discuss the cue-expression and cue-perception processes that mediate the amount of judgmental accuracy and summarize initial empirical process insight. Fourth, based on a process understanding we describe domains of moderators that influence how well perceivers can judge others' personalities (e.g., good trait, good judge, good target, good information). Finally, we highlight a set of issues we deem as important challenges for future research on the accuracy of personality judgments.

Laypeople as well as professionals such as clinicians, therapists, and personnel recruiters regularly form judgments about other people's personality traits. These personality judgments inform one's everyday social experiences and decisions. Judging an applicant as low in conscientiousness will reduce the chances that he or she will be given a job. Depending on whether strangers are evaluated as aggressive or not, one will tend to approach or avoid them. Mating decisions might be informed by how trustworthy or neurotic a potential partner is judged to be. In all of these cases, personality judgments tend to be quickly apparent and relatively stable (Harris & Garris, 2008; Kenny, Horner, Kashy, & Chu, 1992; Willis & Todorov, 2006).

Acknowledgment: We thank Joscha Stecker and Niklas Stein for comments on an earlier version of this chapter.

Here, we will present an overview of the current state of research on this ubiquitous and consequential social phenomenon. For the sake of conciseness and given that personality judgments between close social partners are driven by distinct processes, we will focus on the first judgmental snapshots of who our social counterparts are. That is, we include research on personality judgments at zero acquaintance (i.e., before any interaction has taken place) and short-term acquaintance (after initial interactions). We will first present methodological approaches to investigating the accuracy of these judgments. Afterward, a brief summary is given on how well lay perceivers judge people's personality traits. Focusing on a Brunswikian lens model approach, we will then describe how the amount of accuracy can be understood by close investigations of the expression and utilization of cues as mediators implied in the judgment process. Based on this, we discuss research on moderators of accuracy and try to explain why certain traits can be judged more accurately, certain contexts yield more accurate judgments, certain individuals are better judges, and certain targets are judged more accurately. In a final section, we will briefly discuss issues we regard as important for future research on the accuracy of personality judgments.

Methodological approaches to assessing accuracy

The accuracy of personality judgments can be defined as the *correspondence* between how *perceivers* judge the *personality characteristics* of targets and how targets "really" are concerning these personality characteristics ("*true personality*"). Empirical approaches to this question agree that *personality characteristics* pertain to relatively stable mental and behavioral interindividual differences, that is to *how people think, feel, and behave in comparison to others*. They differ, however, with respect to the operationalization of three key aspects of this lay definition: (1) whether *correspondence* is defined based on a variable-centered or person-centered approach, (2) how "*true personality*" is defined, and (3) whether the accuracy of personality perceptions is captured on the level of the *individual perceiver* or the *aggregated perceiver*.[1]

Definition of correspondence: accuracy in judging traits versus trait profiles

According to a general distinction in differential psychology (Cattell, 1946; Stern, 1911), an individual's uniqueness can be conceptualized in two

[1] For the sake of brevity we do not cover further componential approaches to accuracy here (cf., Cronbach, 1955; Judd & Park, 1993; Levesque & Kenny, 1993).

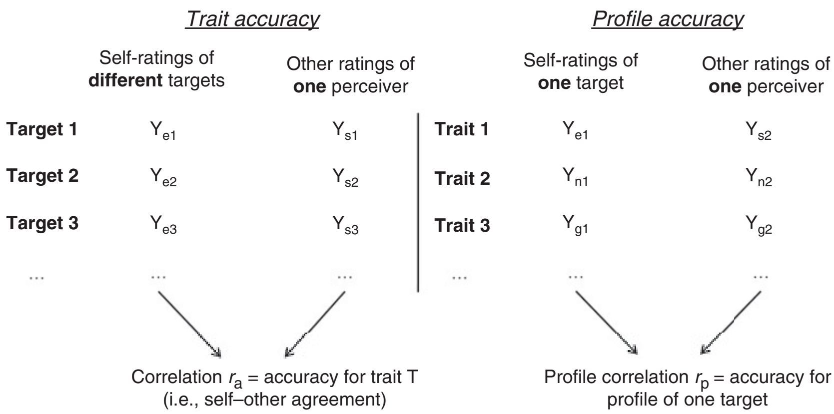

Figure 5.1 Two definitions of accuracy: trait accuracy and profile accuracy.

fundamentally different ways. First, following a variable-centered or trait-based approach, each individual's standing on a given trait can be compared to the trait standing of several other individuals on this trait (rank order of individuals for each trait). Second, following a person-centered or profile-based approach, for each individual the profile of several trait values is analyzed (rank order of traits within each individual). This classic distinction is mirrored by two complementary methodological approaches to the accuracy of personality judgments (see Figure 5.1).

A *trait-based approach* to assessing the accuracy of personality judgments focuses on perceivers' ability to judge targets' trait standing(s) relative to others. Accuracy is computed across targets for each specific trait: Perceivers' judgments of a given trait are compared with the targets' actual values on this trait. To determine perceivers' trait accuracy across a number of traits this is done separately for each trait. This approach allows the researcher to assess whether individuals can evaluate who is, for example, more or less trustworthy, extraverted, or neurotic, and whether there are certain perceivers who are particularly good at judging trustworthiness, extraversion, or neuroticism (cf. Funder, 1999; Nestler & Back, 2013, in press).

A *profile-based approach* focuses on a perceiver's ability to detect each target's idiosyncratic ordering of traits relative to each other. Profile accuracy is determined by comparing, for each perceiver–target combination separately, the profile of trait judgments (e.g., concerning the target's trustworthiness, extraversion, and neuroticism) with the corresponding profile of the target's trait criterion values. The higher the agreement between the two profiles, the higher the profile accuracy of the perceiver (for this target). With the profile-based approach, one can, for example, examine how much an individual can evaluate whether targets are more extraverted than they are trustworthy, and whether there are certain perceivers who are particularly good at judging a target's idiosyncratic personality profile. Most comprehensively, the profile approach is covered in the social accuracy model (SAM; Biesanz, 2010). The SAM allows one to empirically disentangle the accuracy in judging the average personality profile of others (*normative profile accuracy*) from the accuracy in judging targets' unique personality profiles (*distinctive profile accuracy*), both of which contribute to the convergence of a given target's personality profile and a given perceiver's judgments profile (see Furr, 2008, for a related approach to profile similarity).

Often, the distinction between trait-based and profile-based approaches is not made explicitly and they are rather applied interchangeably as mere methodological variants of analyzing the accuracy of personality judgments. We regard it as important that these approaches are different on a conceptual level. Trait accuracy examines whether a perceiver can correctly judge the rank order of the personality trait values of different targets. Profile accuracy analyses assess whether a perceiver knows which of a target's personality traits are individually more prominent than other traits. Profile accuracy and trait accuracy thus refer to different psychological phenomena. This also follows from the observation that accuracy at the profile level does not entail accuracy at the trait level: Perceivers might have an accurate conception of how a target's traits relate to each other (whether a target is more extraverted than trustworthy) but may be wrong with regard to how the target is distinct compared with other people (whether the target is extraverted or whether s/he is trustworthy). One should, thus, apply these approaches (one of them or both) depending on the research question(s) at hand (i.e., whether one is interested in the accuracy of judging targets' trait standings or the accuracy of judging targets' idiosyncratic trait profiles). In addition, more research is needed that directly compares these approaches with the same sets of data (e.g., Bernieri, Zuckerman, Koestner, & Rosenthal, 1994).

Definition of "true" personality: accuracy criteria

Another key issue for accuracy research is to define what "true" personality is. Although, there is, obviously, not a definite answer to this question, researchers might nevertheless gather as much information as possible for the best approximation to "true personality." Typically, three kinds of criterion measures are applied (with descending frequency and ease of assessment): self-reports, acquaintance reports, and behavioral assessments (specifically created observation situations or tests are used to observe and code behaviors theoretically related to the personality traits in question). As there is not a single best criterion of "true personality" and each kind of criterion measure not only captures valid aspects but also entails blind spots, the assessment of several valid measures can be regarded as a "gold standard" (cf. Back & Vazire, 2012; Funder, 1999). When only one of these possible criteria is used, it is advisable to explicitly mention the specific operationalization of accuracy applied (e.g., *self–other agreement* when self-reports are used as criterion).

Single versus aggregate perceiver accuracy

When analyzing the accuracy of perceivers one can calculate the variable-centered trait correlations or person-centered profile correlations between criterion measures and judgments for (a) an aggregate of the judgments of several perceivers yielding an overall *aggregate perceiver accuracy*, or (b) for each perceiver separately and then compute a mean *single perceiver accuracy* across perceivers. Typically, the former aggregate perceiver approach is applied in accuracy research. One should, however, be aware that the aggregate approach does not yield accuracy scores for individual perceivers. Furthermore, results using the aggregate method are dependent on the number of perceivers in a given study, with more perceivers leading to higher accuracy estimates for purely statistical reasons (cf. Hall & Bernieri, 2001; Hirschmüller, Egloff, Nestler, & Back, 2013; Nestler, Egloff, Küfner, & Back, 2012). Analyzing accuracies at the level of the individual perceiver has the additional advantage that one can report on the variability of accuracy across individuals and, based on this, analyze the determinants, moderators, and consequences of individual differences in accuracy (see Nestler & Back, in press; and Biesanz, 2010; for unified statistical approaches). For reasons of comparability we recommend that researchers apply and consistently report single as well as aggregate perceiver accuracies. (It should also be noted that aggregate perceiver accuracy will, again for statistical reasons, typically be higher than the average accuracy across individual perceivers; Hall, Bernieri, & Carney, 2005.)

Level of accuracy: are people accurate when they judge the personality of others?

How well can lay perceivers judge the personality of others? The accuracy of trait judgments was a lively field of research up to the 1950s (Estes, 1938; Taft, 1955). Following a complex methodological critique on some of the common statistical analyses (Cronbach, 1955), the research field experienced a "sudden death." Henceforward, the question of accuracy was only indirectly captured in research on social cognition biases (Gilovich, 1991; Nisbett & Ross, 1980). This research focused on the kinds of errors humans commit when judging others, and, unfortunately, came along with the labeling of the phenomenon of personality judgments per se as bias (i.e., fundamental attribution error) and the explicit dismissing of any value of accuracy research (Fiske & Taylor, 1991; Jones, 1990; see Funder, 1999; Jussim, 2012 for detailed historical overviews and pointed critiques of this approach). With the exception of a study by Norman and Goldberg (1966) who had students rate each other's Big Five personality traits on the first day of class and found stranger ratings to be related to the student's self-ratings, it was not until the late 1980s that researchers started to systematically reinvestigate the accuracy of lay judgments of personality.

Kenny and colleagues as well as Watson, replicating the early findings by Norman and Goldberg, had groups of strangers interact with and judge each other (e.g., Albright, Kenny, & Malloy, 1988; Watson, 1989). Borkenau and colleagues presented lay perceivers videos in which targets read out a weather report (Borkenau & Liebler, 1992, 1993). Gangestad, Simpson, and colleagues used videotaped interviews of individual targets (Gangestad, Simpson, DiGeronimo, & Biek, 1992), while Funder and colleagues used videotaped dyadic interactions (Funder & Colvin, 1988), and Gifford looked at judgments based on videotaped interactions of triads (Gifford, 1994). Ambady and colleagues (Ambady, Hallahan, & Rosenthal, 1995) examined personality judgments in groups without any direct interaction. Finally, research groups around Zebrowitz and Berry obtained personality ratings based on facial photographs only (e.g., Berry, 1990; Zebrowitz, Voinescu, & Collins, 1996). In all of these initial studies, a certain amount of correspondence between strangers' ratings and personality criteria was observed.

These groundbreaking studies have led to an increasing amount of further research showing above-chance levels of accuracy in judging personality in a wide variety of contexts (see Ambady, Bernieri, & Richeson, 2000; Funder, 2012; Kenny & West, 2008, Nestler & Back, 2013, for overviews) including personality judgments based on direct interactions (Human & Biesanz, 2011a, 2011b; Kenny et al., 1992; Kenny & West, 2010), photographs (Borkenau, Brecke, Möttig, & Paelecke, 2009; Naumann, Vazire, Rentfrow, & Gosling, 2009), video and audio materials (Borkenau, Mauer, Riemann, Spinath, & Angleitner, 2004; Funder & Sneed, 1993), natural conversations (Holleran, Mehl, & Levitt, 2009), streams of thought (Holleran & Mehl, 2008), writing styles (Küfner, Back, Nestler, & Egloff, 2010), online (networking) profiles (Back et al., 2010; Tskhay & Rule, 2014; Vazire & Gosling, 2004), microblogs (Qiu, Lin, Ramsay, & Yang, 2012), vlogs (Biel & Gatica-Perez, 2013), e-mail addresses (Back, Schmukle, & Egloff, 2008), music preferences (Rentfrow & Gosling, 2006), and offices and bedrooms (Gosling, Ko, Mannarelli, & Morris, 2002).

Most of these studies investigated the Big Five personality dimensions (e.g., John, Naumann, & Soto, 2008), but accurate judgments have also been found for other traits such as intelligence (Borkenau & Liebler, 1993; Murphy, Hall, & Colvin, 2003; Reynolds & Gifford, 2001; Zebrowitz, Hall, Murphy, & Rhodes, 2002), self-esteem (Kilianski, 2008), narcissism (Buffardi & Campbell, 2008), aggressiveness (Kenny et al., 2007), sociosexuality (Gangestad et al., 1992), and attitudes (Beer & Watson, 2008; Paunonen & Kam, 2014).

For trait-wise accuracies, the amount of accuracy tends to lie between .20 and .40 (aggregated perceivers) and between .10 and .30 (single perceiver analyses) across traits and contexts (Connelly & Ones, 2010; Connolly, Kavanagh, & Viswesvaran, 2007; Funder, 1999; Hall, Andrzejewski, Murphy, Schmid Mast, & Feinstein, 2008; Kenny, 1994; Kenny & West, 2010). When translating these correlation coefficients into a more intuitive metric, the proportion of correct judgments (Hall et al., 2008; Rosenthal & Rubin, 1989), lay judgments of personality typically reach hit rates of about 55% to 70%. Accuracies in judging the personality profile of others are typically in similar ranges, but it has to be noted that the accuracy in judging the typical profile of others (normative profile accuracy) is substantially higher than the accuracy in judging another's unique personality profile (distinctive profile accuracy; Biesanz, 2010).

Are these numbers high or low, impressive or disappointing? When evaluating the amount of accuracy, one has to bear in mind that judging the personality of strangers is a nontrivial process. Multiple consecutive steps have to be successfully passed through for accurate judgments to occur: Target personality needs to be expressed and observable in a given context, and these expressions need to be detected and correctly utilized by perceivers (see Funder, 1999, and the following section on *Processes*). In sum, although judging others is a complex process, humans seem to have developed an above chance ability to distinguish other people's trait levels and trait profiles, a finding that also makes sense from an evolutionary perspective (Haselton & Funder, 2006).

Based on this, accuracy research has moved further, trying to understand *how* (by means of which processes) and *when* (under which personal and situational circumstances) perceivers render more or less accurate personality judgments.

Processes: why are perceivers (in)accurate?

How do lay perceivers manage to render above chance accurate personality judgments of strangers? And why are they not better? While there are now a large number of studies showing that zero-acquaintance personality judgments show a small but significant amount of accuracy, less is known about the mediating processes.

Here, we will focus on the so-called lens model (Brunswik, 1956; Hammond, 1996; Karelaia & Hogarth, 2008; Nestler & Back, 2013), as a general conceptual framework and analytical tool to understand and analyze the accuracy of personality judgments following a trait-based approach – and to gain insights into the idiosyncratic judgment processes involved. Grounded in Egon Brunswik's functional probabilism (Brunswik, 1956), the lens model assumes that in order to navigate in fundamentally uncertain environments, humans built impressions of latent, not directly observable, characteristics of their environment by focusing on and utilizing sets of observable cues to these characteristics. In order to achieve accurate judgments, perceivers need to be sensitive to available (i.e., observable) and valid information (i.e., cues that are indeed indicative of the latent characteristics of interest).

Applied to the context of judging personality characteristics (Borkenau & Liebler, 1992; Gangestad et al., 1992; Gifford, 1994; Nestler & Back, 2013; see Figure 5.2), for a perceiver to be accurate, targets' latent trait values need to be expressed in observable cues (e.g., extraverted individuals show more expressive gestures), and the perceiver needs to use valid cues to render his/her judgments (e.g., s/he judges targets who show expressive gestures as more extraverted).[2] Each cue in a given context can be characterized by how much it is a valid indicator of a trait (cue validity) and how much it is used by perceivers (cue utilization). By focusing on observable cues, the lens model framework allows the researcher to explain the amount of accuracy by the predictability of a trait in a given context (to what extent valid cues exist), as well as the response consistency (how much perceivers apply cue information

[2] Beyond behavioral cues, there are other sources of personality judgments such as perceivers' self-perceptions (Kenny, 1994) and stereotypes (Bodenhausen & Macrae, 1998), typically not covered in lens model research. Importantly, however, both important aspects of the judgment process can be analyzed within lens model frameworks (e.g., Gosling et al., 2002; Stopfer et al., 2014).

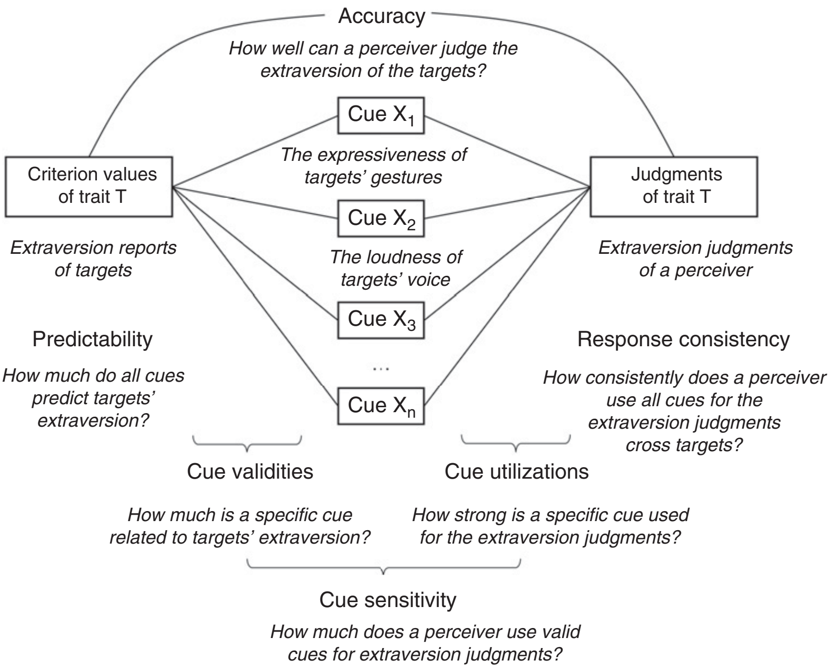

Figure 5.2 A lens model describing the processes underlying accurate personality judgments.

consistently across targets) and sensitivity of perceivers (how much perceivers make stronger use of valid as compared to invalid cues).

A variant of the lens model is the Realistic Accuracy Model (RAM; Funder, 1999, 2012). Here, the cue validity and cue utilization stages of the lens model are each separated into two independent process steps. According to the RAM, accurate judgments of personality can occur if (a) the trait value of a target is related to the degree of some sort of behavioral expression (*relevance*), (b) these expressions are observable, and, therefore, available to the perceiver in a given context (*availability*), (c) perceivers detect them (*detection*), and (d) correctly utilize them for their judgment (*utilization*). The RAM allows for a very detailed conceptual analysis of the accuracy processes and nicely disentangles theoretically the multitude of stages involved in the accuracy process. One challenge of the RAM to date is that it is relatively difficult to distinguish the cue relevance and availability stages as well as the detection and utilization stages of this process chain empirically. When collapsing the first two processes and the last two processes (i.e., in its more classic variant of the lens model), things are easier to grasp empirically. Specifically, by distinguishing between the three classes of variables in the lens model (personality criteria, cues, personality judgments), one is left with two process steps: (1) personality expression (relations between personality criteria and observable cues, i.e., cue validities) and (2) impression formation (relations between observable cues and personality judgments, i.e., cue utilizations).

In fact, an important feature of the lens model is that the key explanatory concepts (cue validity, predictability, cue utilization, response consistency, and sensitivity) can be computed in a straightforward way (see Figure 5.3 for a schematic illustration; see Karelaia & Hogarth, 2008). Regarding the left side of the lens model, the validity of each cue is given by its correlation with the trait criterion or alternatively its unique validity is given by its regression weight, when regressing the trait criterion on all available cues. Predictability, then, is assessed by the multiple regression coefficient of this regression, indicating how much a personality trait is expressed in (and, therefore, is predictable by) observable cues in a given context. Regarding the right side of the lens model, cue utilizations pertain to each cue's correlation with the trait judgment or alternatively its regression weight, when regressing the trait judgment on all available cues (unique utilization). Response consistency is calculated as the multiple correlation of this regression. It indicates how much each perceiver consistently applies his/her judgment model across targets (e.g., always judges target as extraverted if s/he shows expressive gestures). Finally, sensitivity is given by the correlation between the predicted values of the aforementioned two kinds of regressions. It thus reflects how much perceivers make stronger use of valid as compared to invalid cues – or, in other words, whether perceivers are sensitive to the validity of observable cues.[3] Importantly, all perceptual lens model indices (cue utilization, cue sensitivity, response consistency) can be calculated on the level of the individual perceiver in a given situation.

[3] This is the classic regression-based method of calculating judgment sensitivity in lens model research (e.g., Hursch, Hammond, & Hursch, 1964; Tucker, 1964). In the domain of personality judgments, another common method of calculating sensitivity is to calculate the vector correlation between the Fisher z-transformed cue validities and cue utilization correlations across cues. This method is, however, biased and should, in our view, not be used. First, it does not control for the intercorrelation of cues. Second, it is affected by the arbitrary labeling of cues because variances of cue validities and cue utilizations depend on the sign of the single correlations, which themselves depend on how the cue is labeled. Relabeling the cue *nervous facial expression* as *calm facial expression*, for example, would reverse the sign of the respective cue validity and cue utilization correlation, thereby changing the variability of the correlation vectors and, thus, the sensitivity vector correlation.

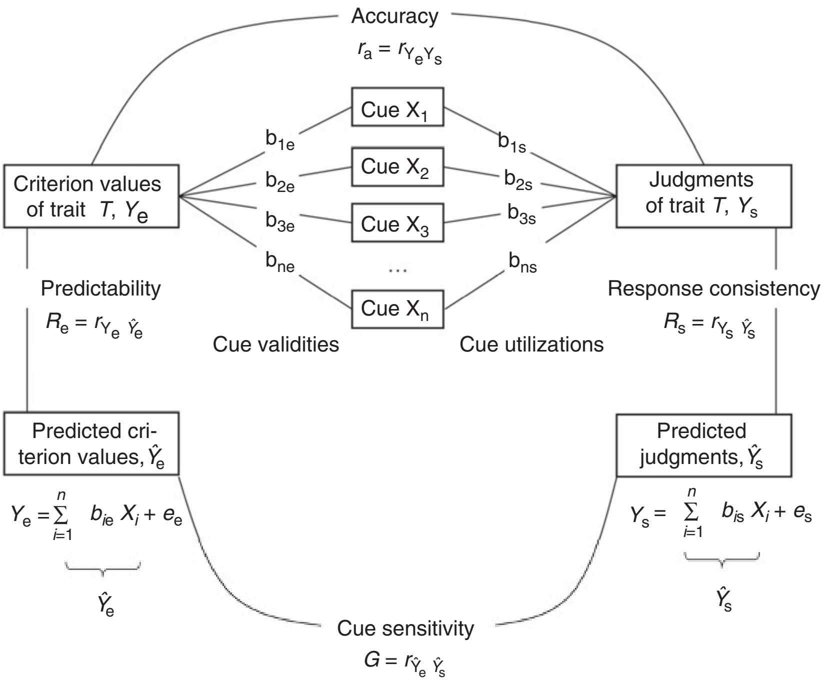

Figure 5.3 The lens model with superimposed statistical parameters.

These lens model indices can be directly applied to explain the amount of accuracy for a given perceiver, trait, and context. A perceiver can be inaccurate for example because s/he does not apply his/her judgment model consistently across targets. This would be mirrored in relatively small cue utilizations and, thus, a low response consistency. Alternatively (or in addition), s/he might lack the sensitivity for what relevant and irrelevant cues are. In this case, there would be a lack of match between the strength with which s/he uses a cue and its validity and, thus, a low-cue sensitivity for this perceiver. Similarly, lens model results can be applied to explain accuracies across perceivers in a specific context or concerning a specific trait. In these cases, the number of valid cues available (in a context/ for a trait) and, thus, the predictability (in a context or of a trait) is the most important lens model indicator. To apply lens model analyses, data are needed that aim at a full coverage of observable cues in the given judgment context. Importantly, this necessarily needs to include invalid cues – only if one covers the full range of valid and invalid cues one is able to meaningfully calculate a perceiver's sensitivity to the validity of available cues (i.e., to distinguish between valid and invalid cues).[4] Complementary to these classic lens model analyses, one can try to fit mediational models by (a) identifying the most important (valid and utilized) cues (e.g., expressive gestures, loud voice, smiling in the case of extraversion), and (b) using them as process variables mediating the relation between the personality criterion (e.g., self-reported extraversion) and each judgment (e.g., extraversion judgments; see Hirschmüller, Egloff, Schmukle, Nestler, & Back, 2015; Küfner et al., 2010; Stopfer, Egloff, Nestler, & Back, 2014).

To date, research in the context of personality judgments has only scratched the surface of what can be revealed by lens model analyses. Most prior empirical work has looked at cue validities and cue utilizations on a more descriptive correlational level. This work already resulted in a number of interesting findings, showing that individual differences in personality are related to specific behavioral expressions and behavioral residues that are picked up by lay perceivers, thereby explaining a certain amount of judgment accuracy. In terms of the Big Five, for example, some relevant specific cues are expressive gestures, smiling, a lot of talking, a loud voice, flashy and fashionable appearances, and personal environments (extraversion); nervous gestures and facial expressions, verbal nervousness, and negative self-references (neuroticism); compliant behaviors, formal and orderly appearances, and personal environments (conscientiousness); social orientation in verbal content (agreeableness); and sophisticated verbal content, and diverse and creative personal environments (openness to experience) (e.g., Back, Schmukle, & Egloff, 2008, 2011; Borkenau & Liebler, 1992; Gosling et al., 2002; Küfner et al., 2010; Naumann et al., 2009). Individual differences in intelligence have been found to be related to fluent speaking (Borkenau & Liebler, 1995) as well as responsiveness to conversation partner and looking while speaking (Murphy, 2007; Murphy et al., 2003). Of course, in addition to using these valid cues, perceivers also make use of invalid cues regarding all of these traits and across contexts. Many of these insensitive cue usages can be explained by halo effects caused by easily and positively perceived cues (e.g., physical attractiveness, nonverbal expressiveness) or by invalid stereotypes associated with surface characteristics of targets (such as gender).

[4] Lens model analyses aim at a comprehensive coverage of the observable target cues, including the invalid ones. This, however, makes it difficult to interpret specific cue-validity and cue-utilization correlations due to the large number of cues included, and, therefore, statistical tests applied. Correction for alpha inflation is not an option, because it would necessarily also mask valid cue associations that are real. The only option is, thus, to replicate cue associations across samples and studies (see Küfner et al. 2010, for an example).

Importantly, the lens model in its present form has been predominantly used to explain trait accuracy and has not been sufficiently elaborated to the understanding of profile accuracy. Logically, to render accurate profile judgments, some kind of cues associated with the relative ordering of traits within targets need to be utilized by perceivers (i.e., profiles of cues). Human, Biesanz, Finseth, Pierce, and Le (2014) took a first step in this direction and explained the accuracy in judging others' personality profiles via personality-congruent behavior profiles: Targets' personality profiles were associated with specific profiles of behavioral cues which were associated with corresponding profiles of perceivers' judgments about the targets.

Moderators: when are personality judgments accurate?

The outlined cue expression and cue utilization processes can be influenced by a number of personal, situational, and trait-specific factors, which thereby moderate the strength of the association between criterion measures of personality and personality judgments, i.e., the accuracy of personality judgments. Funder (1999, 2012) distinguished four classes of moderator variables that have been found to influence the degree to which personality judgments are accurate: differences between traits ("good trait"), differences regarding the quantity and quality of available information ("good information"), differences between perceivers ("good judge"), and differences between targets ("good target").

What traits can be judged most accurately ("good trait")?

Across contexts, targets, and judges, some traits tend to be easier to judge accurately than others. Research has identified two important characteristics of traits that affect their judgeability: observability and evaluativeness (Funder & Dobroth, 1987; John & Robins, 1993; Vazire, 2010). How much a trait expresses itself in observable appearances, behaviors, and behavioral residues (*observability*) enhances its accuracy, due to the higher number of available valid cues perceivers can base their judgment on. In addition, the degree to which a trait tends to be seen as very desirable or very undesirable (*evaluativeness*) has been found to weaken its accuracy – neutral traits are judged more accurately (but see Kenny & West, 2010). This might be due to response factors that affect both targets' self-descriptions (e.g., individual differences in overly positive self-descriptions) as well as perceivers' judgments (e.g., individual differences in overly positive other descriptions; Leising, Erbs, & Fritz, 2010), thereby affecting both the cue validity (self-cue relations) and cue utilization (cue-judgment relations) stages of the judgment process. Consequently, traits that are revealed in many well-observable behaviors across contexts and are not particularly evaluative like extraversion can be judged with highest accuracies, whereas traits that are less observable like neuroticism or highly evaluative traits like agreeableness tend to be difficult to judge (Borkenau & Liebler, 1993; Connelly & Ones, 2010; Connolly et al., 2007; Funder & Dobroth, 1987; Hall et al., 2008; Kenny, Albright, Malloy, & Kashy, 1994; Kenny et al., 1992). Interestingly, although only poorly judged, agreeableness is the most commonly inferred trait when collecting open-ended impressions of lay perceivers based on photographs, videos, and face-to-face encounters (Ames & Bianchi, 2008).

What information leads to accurate judgments ("good information")?

The quantity and quality of information on which judgments are based also moderates their accuracy. The amount of information available to perceivers typically enhances the accuracy of their judgments. This is illustrated by the fact that well-acquainted others make more accurate judgments than strangers (Connelly & Ones, 2010; Connolly et al., 2007; Funder & Colvin, 1988) but has also been shown in cross-sectional (Biesanz, West, & Millevoi, 2007) and longitudinal (Kurtz & Sherker, 2003; Paulhus & Bruce, 1992) designs, in which increasing acquaintance is associated with increasing accuracy, as well as in experimental designs where an increasing length (Blackman & Funder, 1998; Carney, Colvin, & Hall, 2007; Letzring, Wells, & Funder, 2006) or number (Borkenau et al., 2004) of behavioral slices (i.e., short excerpts of social behavior; Ambady et al., 2000; Murphy et al., 2015) was associated with more accurate judgments of strangers. Importantly, however, it is not the amount of time targets are observed or acquaintance per se that enhances accuracy, but the increasing amount of different (i.e., incremental) valid information that is available. The increase in accuracy, for example, levels off relatively early when the same information about targets is presented (Carney et al., 2007; Hall et al., 2008). It can even temporarily drop, when the increase in information is accompanied by a decreasing use of valid stereotypes (Berry, 1990; Biesanz et al., 2007; Jussim, Cain, Crawford, Harber, & Cohen, 2009; Kenny, 1991).

Beyond the quantity of information, the kind of information or its quality also plays an important role. Two qualitative features of the judgment context are particularly important: its "strength" and its "richness." First, situations that allow for more variation in targets' behaviors (that are less "strong") provide perceivers with more valid cues, thereby triggering more accurate judgments. Targets behaving in a free conversation were, for instance, easier to judge than those behaving in a highly structured interaction (Letzring et al., 2006). Second, contexts that incorporate qualitatively different kinds of information allow for the observation of more different valid cues, resulting in more accurate judgments: Judgments based on audible information alone are less accurate than judgments that additionally include visible nonverbal information (Blackman, 2002; Borkenau & Liebler, 1992), and the accuracy of judgments further increase when personal information about thoughts, feelings, and behaviors is added (Beer & Brooks, 2011; Beer & Watson, 2010; Letzring & Human, 2013).

Who can judge others accurately ("good judge")?

The question of who is a good judge of personality is one of the oldest (e.g., Adams, 1927; Allport, 1937; Vernon, 1933) but also one of the most difficult to answer. Most basically, it is relatively hard to find reliable individual differences in accuracy between judges and as a result, there are only few replicated effects of perceiver's personality on accuracy in personality judgments. Three meta-analyses on a very large and broad range of judgment domains (mostly affect but also trait judgments) and perceiver characteristics found some evidence for effects of cognitive abilities and indicators of good psychological adjustment on accuracy (Davis & Kraus, 1997; Hall, Andrzejewski, & Yopchick, 2009; Taft, 1955). To date, there is less systematic overview in the field of personality judgments in particular. Regarding trait-based accuracy, individual accuracy scores tend to correlate only weakly across judged trait dimensions (e.g., Lippa & Dietz, 2000; Nestler et al., 2012). Regarding the effects of the perceiver's personality on accuracy, there is some evidence for intelligence but no consistent evidence for other traits (e.g., Ambady et al., 1995; Bernieri et al., 1994; Lippa & Dietz, 2000).

In the context of profile accuracy, Biesanz and Human have recently shown that the variability in accuracy across perceivers is rather small and substantially smaller than the variability across targets (Biesanz, 2010; Human & Biesanz, 2011b, 2013). This might also explain why prior studies have only identified a small percentage of the investigated characteristics to be correlated with accuracy scores of perceivers, which were, moreover, inconsistent across studies (e.g., Bernieri et al., 1994; Christiansen, Wolcott-Burnam, Janovics, Burns, & Quirk, 2005; Letzring, 2008; Vogt & Colvin, 2003). When distinguishing between components of profile accuracy, accuracy in judging the personality profile of the average other (normative profile accuracy) seems to be higher for female, agreeable, and emotionally stable individuals (Chan, Rogers, Parisotto, & Biesanz, 2011; Human & Biesanz, 2011b; Vogt & Colvin, 2003), a finding, that does, however, not generalize to the accuracy in judging others' unique personality profiles (distinctive profile accuracy).

More effects of perceiver characteristics on the amount of judgmental accuracy might evolve in contexts in which the perceiver has the chance to more directly interact with the target individuals s/he judges. Here, individual differences in evoking the behaviors needed to judge targets (i.e., evoking valid cues) might play an important role, in addition to mere differences in judging the same set of available behaviors (i.e., consistency and sensitivity in using cues). Letzring (2008), for example, showed that targets could be more easily judged by outside perceivers who watched them interacting in groups of three, when targets interacted with other good perceivers, indirectly providing evidence to the contention that good perceivers are able to elicit more valid cues.

Who can be judged accurately ("good target")?

Research on the good target has almost exclusively focused on profile-based accuracy (see Ambady et al., 1995, for an exception). Interestingly, individual differences in being judged accurately in one's personality profile might be more pronounced than individual differences in judging others' profiles accurately (Biesanz, 2010; Human & Biesanz, 2013; Kenny, 1994). Following the lens model, traits that affect the number and consistency of valid behavioral cues the targets express should play an important role. Consistent with this idea, traits related to expressiveness such as extraversion and a lack of emotional suppression, as well as traits related to psychological adjustment such as emotional stability, self-esteem, and well-being, have been identified as features of the good target (Colvin, 1993; Human & Biesanz, 2011a). Targets who simply act more, who express their felt emotions, and whose current acts are more reliable indicators of their typical behavior show more of their "true" personality, and are therefore easier to judge accurately (Human et al., 2014).

Also consistent with the idea that the number of valid behavioral cues expressed is key to being a good target, the instruction to self-present (i.e., to make a "good impression") led to more accurate Big Five profile judgments (Human, Biesanz, Parisotto, & Dunn, 2012). Similarly, individual differences in intelligence were more accurately judged when targets engaged in impression management (trying to appear intelligent) than without (Murphy, 2007). Obviously, when self-presenting, targets engage in more observable behaviors, making it easier for judges to detect individual differences.

Complex moderators

Interactions between the aforementioned domains of moderators can also be relevant. Certain traits can, for example, be better judged in specific kinds of contexts but less so in others. Openness to experience, for instance, can be judged accurately when personal environments and leisure preferences are observed (Back et al., 2010; Gosling et al., 2002; Rentfrow & Gosling, 2006). Here, openness-related cues such as the diversity and creativity in how people think about and create their surroundings are more easily expressed. Neuroticism is judgeable in socially stressful situations, in which individual differences in this trait are expressed in observable cues of visible and audible nervousness (Hirschmüller et al., 2015). Agreeableness has been shown to be judgeable based on written materials (e.g., Küfner et al., 2010; Qiu et al., 2012). All three kinds of traits are, however, difficult to judge based on behavioral expressions and interactions in other contexts (e.g., Borkenau & Liebler, 1992; Wall, Taylor, Dixon, Conchie, & Ellis, 2013).

Moreover, particular judges can be good in judging some but not other traits. In a detailed lens model analysis of personality ratings after dyadic interaction, Hartung and Renner (2011) showed that individual differences in social curiosity of perceivers led to more accurate judgments of extraversion and openness due to a higher sensitivity to valid cues for these traits. Finally, underlining the idea that human observers are particularly good at judging other humans in pragmatically meaningful contexts (Swann, 1984), perceivers were found to be more accurate in judging their partners for traits relevant to romantic relationships (Gill & Swann, 2004).

Future prospects

In our view, future research will profit from a stronger process understanding of accurate personality judgments. In a way, this seems ironic, since accuracy research has once been abandoned by a more "process-oriented" research on biases (see Funder, 1999, and Jussim, 2012, for historical analyses). Of course, however, bias processes and the explanation of the accuracy phenomenon have always been two sides of the same coin. It seems like a rather strange by-product of different methodological traditions that research on the processes underlying social perception biases had so little to say about the factors influencing the amount of accuracy in personality judgments. In trying to unravel accuracy/bias processes further, we will now briefly highlight three lines of investigations, we regard as particularly fruitful, following a realistic process approach to accuracy (Brunswik, 1956, Funder, 1999; Nestler & Back, 2013): (1) the establishment of systematic overviews of relevant cue processes, (2) the integration of different approaches to accuracy beyond personality judgments alone, and (3) longitudinal investigations of accuracy antecedents and consequences.

Toward a systematic overview of cue processes

More original as well as meta-analytical work on the specific processes underlying the accuracy of personality judgments in realistic social settings is needed. What are the most prevalent and important behavioral expression processes by which certain traits of certain targets in certain contexts become observable? What are the key impression formation processes by which certain perceivers succeed or fail to make use of valid information regarding certain traits of certain targets in certain contexts? Following the lens model, future research might, for example, start to provide overviews of replicable valid and utilized cues to personality as well as systematic comparisons of predictability, consistency, and sensitivity results for specific personality traits in specific social contexts. Related to this, future research might apply the lens model as a framework to unravel the cue processes underlying moderators of accuracy (e.g., Why are well-adjusted individuals easier to read?).

Toward an integrative approach to the accuracy of interpersonal judgments

More integration across conceptual approaches to accuracy is needed within and beyond the field of personality judgments. This might include more firm comparisons of variable- and person-centered approaches, both conceptually and methodologically. We regard it as unfortunate that most prior research has applied the one or the other approach making it difficult to directly compare and integrate results across approaches. Both approaches provide valuable insights into the accuracy of personality judgments and they do so in complementary ways. Personality judgments also do not occur in mental isolation but are closely tied to other interpersonal perceptions and states of mind (e.g., Kenny, 1994), including self-perceptions (Back & Vazire, 2012), meta-perceptions (see Chapter 8), and affect (see Chapter 2). Lens model approaches (Funder, 1999; Nestler & Back, 2013) can be expanded in straightforward ways to investigate personality judgments in conjunction with other judgment and perception domains, thereby allowing for more integrated approaches to social judgment processes (e.g., Letzring & Hall, 2012; Stopfer et al., 2014). Breaking several theoretical accounts to judgmental accuracy down to a common set of core cue-expression and cue-utilization processes might also serve as a framework against which more complex and diverse theoretical approaches can be described and compared.

Toward an understanding of the development and consequences of accuracy

We would embrace more research on the antecedents and consequences of accurate personality judgments in longitudinal real-life contexts. How and based on which interaction processes does accuracy develop across the acquaintance process? Are people who are more accurate in judging others or those who can be more accurately judged happier, more popular, or more successful than others? And if so, why? To date, there exist only very few empirical studies on such topics (e.g., Human, Sandstrom, Biesanz, & Dunn, 2013). Integrating research on the accuracy of personality judgments with research on the joint development of social relationships and personality (Back, Baumert et al., 2011; Back & Vazire, 2015) seems to be a fruitful avenue for future research.

References

Adams, H. F. (1927). The good judge of personality. *Journal of Abnormal and Social Psychology*, 22, 172–181.

Albright, L., Kenny, D. A., & Malloy, T. E. (1988). Consensus in personality judgments at zero acquaintance. *Journal of Personality and Social Psychology*, 55, 387–395.

Allport, G. W. (1937). *Personality: A psychological interpretation*. New York, NY: Holt, Reinhart & Winston.

Ambady, N., Bernieri, F. J., & Richeson, J. A. (2000). Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. *Advances in Experimental Social Psychology*, 32, 201–271.

Ambady, N., Hallahan, M., & Rosenthal, R. (1995). On judging and being judged accurately in zero acquaintance situations. *Journal of Personality and Social Psychology*, 69, 518–529.

Ames, D. R., & Bianchi, E. C. (2008). The agreeableness asymmetry in first impressions: Perceivers' impulse to (mis)judge agreeableness and how it is moderated by power. *Personality and Social Psychology Bulletin*, 34, 1719–1736.

Back, M. D., Baumert, A., Denissen, J. J. A., Hartung, F.-M., Penke, L., Schmukle, S. C., Schönbrodt, F. D., Schröder-Abé, M., Vollmann, M., Wagner, J., & Wrzus, C. (2011). PERSOC: A unified framework for understanding the dynamic interplay of personality and social relationships. *European Journal of Personality*, 25, 90–107.

Back, M. D., Schmukle, S. C., & Egloff, B. (2008). How extraverted is inferring personality traits from email addresses. *Journal of Research in Personality*, 42, 1116–1122.

Back, M. D., Schmukle, S. C., & Egloff, B. (2011). A closer look at first sight: Social relations lens model analyses of personality and interpersonal attraction at zero acquaintance. *European Journal of Personality*, 25, 225–238.

Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B., & Gosling, S. D. (2010). Facebook profiles reflect actual personality, not self-idealization. *Psychological Science*, 21, 372–374.

Back, M. D., & Vazire, S. (2012). Knowing our personality. In S. Vazire & T. D. Wilson (Eds.) *Handbook of self knowledge* (pp. 131–156). New York: Guilford

Back, M. D., & Vazire, S. (2015). The social consequences of personality: Six suggestions for future research. *European Journal of Personality*, 29, 296–307.

Beer, A., & Brooks, C. (2011). Information quality in personality judgment: The value of personal disclosure. *Journal of Research in Personality*, 45, 175–185.

Beer, A., & Watson, D. (2008). Personality judgment at zero acquaintance: Agreement, assumed similarity, and implicit simplicity. *Journal of Personality Assessment*, 90, 250–260.

Beer, A., & Watson, D. (2010). The effects of information and exposure on self-other agreement. *Journal of Research in Personality*, 44, 38–45.

Bernieri, F. J., Zuckerman, M., Koestner, R., & Rosenthal, R. (1994). Measuring person perception accuracy: Another look at self-other agreement. *Personality and Social Psychology Bulletin*, 20, 367–378.

Berry, D. S. (1990). Taking people at face value: Evidence for a kernel of truth hypothesis. *Social Cognition*, 8, 343–361.

Biel, J., & Gatica-Perez, D. (2013). The YouTube lens: Crowdsourced personality impressions and audiovisual analysis of vlogs. *Multimedia, IEEE Transactions on*, 15, 41–55.

Biesanz, J. C. (2010). The social accuracy model of interpersonal perception: Assessing individual differences in perceptive and expressive accuracy. *Multivariate Behavioral Research*, 45, 853–885.

Biesanz, J. C., West, S. G., & Millevoi, A. (2007). What do you learn about someone over time? The relationship between length of acquaintance and consensus and self-other agreement in judgments of personality. *Journal of Personality and Social Psychology*, 92, 119–135.

Blackman, M. C. (2002). The employment interview via the telephone: Are we sacrificing accurate personality judgments for cost efficiency? *Journal of Research in Personality*, 36, 208–223.

Blackman, M. C., & Funder, D. C. (1998). The effect of information on consensus and accuracy in personality judgment. *Journal of Experimental Social Psychology*, 34, 164–181.

Bodenhausen, G. V., & Macrae, C. N. (1998). Stereotype activation and inhibition. In R. S. Wyer, Jr. (Ed.), *Advances in social cognition* (Vol. 11, pp. 1–52). Mahwah, NJ: Erlbaum.

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

Borkenau, P., & Liebler, A. (1992). Trait inferences: Sources of validity at zero acquaintance. *Journal of Personality and Social Psychology*, 62, 645–657.

Borkenau, P., & Liebler, A. (1993). Consensus and self-other agreement for trait inferences from minimal information. *Journal of Personality*, 61, 477–496.

Borkenau, P., & Liebler, A. (1995). Observable attributes as cues and manifestations of personality and intelligence. *Journal of Personality*, 63, 1–25.

Borkenau, P., Mauer, N., Riemann, R., Spinath, F. M., & Angleitner, A. (2004). Thin slices of behavior as cues of personality and intelligence. *Journal of Personality and Social Psychology*, 86, 599–614.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments* (2nd ed.). Berkeley: University of California Press.

Buffardi, L. E., & Campbell, W. K. (2008). Narcissism and social networking web sites. *Personality and Social Psychology Bulletin*, 34, 1303–1314.

Carney, D. R., Colvin, C. R., & Hall, J. A. (2007). A thin slice perspective on the accuracy of first impressions. *Journal of Research in Personality*, 41, 1054–1072.

Cattell, R. B. (1946). *The description and measurement of personality*. New York, NY: World Book.

Chan, M., Rogers, K. H., Parisotto, K. L., & Biesanz, J. C. (2011). Forming first impressions: The role of gender and normative accuracy in personality perception. *Journal of Research in Personality*, 45, 117–120.

Christiansen, N. D., Wolcott-Burnam, S., Janovics, J. E., Burns, G. N., & Quirk, S. W. (2005). The good judge revisited: Individual differences in the accuracy of personality judgments. *Human Performance*, 18, 123–149.

Colvin, C. R. (1993). “Judgable” people: Personality, behavior, and competing explanations. *Journal of Personality and Social Psychology*, 64, 861–873.

Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers’ accuracy and predictive validity. *Psychological Bulletin*, 136, 1092–1122.

Connolly, J. J., Kavanagh, E. J., & Viswesvaran, C. (2007). The convergent validity between self and observer ratings of personality: A metaanalytic review. *International Journal of Selection and Assessment*, 15, 110–117.

Cronbach, L. J. (1955). Processes affecting scores on understanding of others and assumed similarity. *Psychological Bulletin*, 52, 177–193.

Davis, M. H., & Kraus, L. A. (1997). Personality and empathic accuracy. In W. Ickes (Ed.), *Empathic accuracy* (pp. 144–168). New York: Guilford.

Estes, S. G. (1938). Judging personality from expressive behavior. *Journal of Abnormal and Social Psychology*, 33, 217–236.

Fiske, S. T., & Taylor, S. E., (1991). *Social cognition*. (2nd ed.). New York: McGraw-Hill.

Funder, D. C. (1999). *Personality judgment: A realistic approach to person perception*. San Diego, CA: Academic Press.

Funder, D. C. (2012). Accurate personality judgment. *Current Directions in Psychological Science*, 21, 177–182.

Funder, D. C., & Colvin, C. R. (1988). Friends and strangers: Acquaintanceship, agreement, and the accuracy of personality judgment. *Journal of Personality and Social Psychology*, 55, 149–158.

Funder, D. C., & Dobroth, K. M. (1987). Differences between traits: Properties associated with interjudge agreement. *Journal of Personality and Social Psychology*, 52, 409–418.

Funder, D. C., & Sneed, C. D. (1993). Behavioral manifestations of personality: An ecological approach to judgmental accuracy. *Journal of Personality and Social Psychology*, 64, 479–490.

Furr, M. R. (2008). A framework for profile similarity: Integrating similarity, normativeness, and distinctiveness. *Journal of Personality*, 76, 1267–1316.

Gangestad, S. W., Simpson, J. A., DiGeronimo, K., & Biek, M. (1992). Differential accuracy in person perception across traits: Examination of a functional hypothesis. *Journal of Personality and Social Psychology*, 62, 688–698.

Gifford, R. (1994). A lens-mapping framework for understanding the encoding and decoding of interpersonal dispositions in nonverbal behaviors. *Journal of Personality and Social Psychology*, 66, 398–412.

Gill, M. J., & Swann, W.B., Jr. (2004) On what it means to know someone: A matter of pragmatics. *Journal of Personality and Social Psychology*, 86, 405–418.

Gilovich, T. (1991). *How we know what isn't so: The fallibility of human reason in everyday life*. New York: The Free Press.

Gosling, S. D., Ko, S. J., Mannarelli, T., & Morris, M. E. (2002). A room with a cue: Personality judgments based on offices and bedrooms. *Journal of Personality and Social Psychology*, 82, 379–398.

Hall, J. A., Andrzejewski, S. A., Murphy, N. A., Schmid Mast, M., & Feinstein, B. (2008). Accuracy of judging others' traits and states: Comparing mean levels across tests. *Journal of Research in Personality*, 42, 1476–1489.

Hall, J. A., Andrzejewski, S. A., & Yopchick, J. E. (2009). Psychosocial correlates of interpersonal sensitivity: A meta-analysis. *Journal of Nonverbal Behavior*, 33, 149–180.

Hall, J. A., & Bernieri, F. (Eds.). (2001). *Interpersonal sensitivity: Theory and measurement*. Mahwah, NJ: Erlbaum.

Hall, J. A., Bernieri, F. J., & Carney, D. R. (2005). Nonverbal behavior and interpersonal sensitivity. In J. A. Harrigan, R. Rosenthal, & K. R. Scherer (Eds.), *The new handbook of methods in nonverbal behavior research* (pp. 237–281). Oxford: Oxford University Press.

Hammond, K. R. (1996). *Human judgment and social policy: Irreducible uncertainty, inevitable error, unavoidable injustice*. Oxford, England: Oxford University Press.

Harris, M. J., & Garris, C. P. (2008). You never get a second chance to make a first impression: Behavioral consequences of first impressions. In N. Ambady & J. J. Skowronski (Eds.), *First impressions* (pp. 147–168). New York: Guilford Press.

Hartung, F.-M., & Renner, B. (2011). Social curiosity and interpersonal perception: A judge x trait interaction. *Personality and Social Psychology Bulletin*, 37, 796–814.

Haselton, M. G., & Funder, D. C. (2006). The evolution of accuracy and bias in social judgment. In M. Schaller, D. T. Kenrick, & J. A. Simpson (Eds.), *Evolution and social psychology* (pp. 15–37). New York: Psychology Press.

Hirschmüller, S., Egloff, B., Nestler, S., Back, M. D. (2013). The dual lens model: A comprehensive framework for understanding self-other agreement of personalty judgments at zero acquaintance. *Journal of Personality and Social Psychology*, 104, 335–353.

Hirschmüller, S., Egloff, B., Schmukle, S. C., Nestler, S., & Back, M. D. (2015). Accurate judgments of neuroticism at zero acquaintance: A question of relevance. *Journal of Personality*, 83, 221–228.

Holleran, S. E., & Mehl, M. R. (2008). Let me read your mind: Personality judgments based on a person's natural stream of thought. *Journal of Research in Personality*, 42, 747–754.

Holleran, S. E., Mehl, M. R., & Levitt, S. (2009). Eavesdropping on social life: The accuracy of stranger ratings of daily behavior from thin slices of natural conversations. *Journal of Research in Personality*, 43, 660–672.

Human L. J., & Biesanz, J. C. (2011a). Target adjustment and self-other agreement: Utilizing trait observability to disentangle judgeability and self-knowledge. *Journal of Personality and Social Psychology*, 101, 202–216.

Human L. J., & Biesanz, J. C. (2011b). Through the looking glass clearly: Accuracy and assumed similarity in well-adjusted individuals' first impressions. *Journal of Personality and Social Psychology*, 100, 349–364.

Human, L. J., & Biesanz, J. C. (2013). Targeting the good target: An integrative review of the characteristics and consequences of being accurately perceived. *Personality and Social Psychology Review*, 17, 248–272.

Human, L. J., Biesanz, J. C., Finseth, S., Pierce, B, & Le, M. (2014). To thine own self be true: Psychological adjustment promotes judgeability via personality-behavior congruence. *Journal of Personality and Social Psychology*, 106, 286–303.

Human, L. J., Biesanz, J. C., Parisotto, K. L., & Dunn, E. W. (2012). Your best self helps reveal your true self: Positive self-presentation results in more accurate personality impressions. *Social Psychology and Personality Science*, 3, 23–30.

Human, L. J., Sandstrom, G. M., Biesanz, J. C., & Dunn, E. W. (2013). Accurate first impressions leave a lasting impression: The long-term effects of accuracy on relationship development. *Social Psychological and Personality Science*, 4, 395–402.

Hursch, C. J., Hammond, K. R., & Hursch, J. L. (1964). Some methodological considerations in multiple-probability studies. *Psychological Review*, 71, 42–60.

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues. In

O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 114–158). New York: Guilford Press.

John, O. P., & Robins, R. W. (1993). Determinants of interjudge agreement on personality traits: The Big Five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality*, 61, 521–551.

Jones, E. E. (1990). *Interpersonal perception*. New York: W. H. Freeman and Company.

Judd, C. M, & Park, B. (1993). Definition and assessment of accuracy in social stereotypes. *Psychological Review*, 100, 109–128.

Jussim, L. (2012). *Social perception and social reality: Why accuracy dominates bias and self-fulfilling prophecy*. New York: Oxford University Press.

Jussim, L., Cain, T., Crawford, J., Harber, K., & Cohen, F. (2009). The unbearable accuracy of stereotypes. In T. Nelson (Ed.), *Handbook of prejudice, stereotyping, and discrimination* (pp. 199–227). Hillsdale, NJ: Erlbaum.

Karelaia, N., & Hogarth, R. M. (2008). Determinants of linear judgment: A meta-analysis of lens model studies. *Psychological Bulletin*, 134, 404–426.

Kenny, D.A. (1991). A general model of consensus and accuracy in interpersonal perception. *Psychological Review*, 98, 155–163.

Kenny, D. A. (1994). *Interpersonal perception*. New York: Guilford.

Kenny, D. A., Albright, L., Malloy, T. E., & Kashy, D. A. (1994). Consensus in interpersonal perception: Acquaintance and the Big Five. *Psychological Bulletin*, 116, 245–258.

Kenny, D. A., Horner, C., Kashy, D. A., & Chu, L. (1992). Consensus at zero acquaintance: Replication, behavioral cues, and stability. *Journal of Personality and Social Psychology*, 62, 88–97.

Kenny, D. A., & West, T. V. (2008). Zero acquaintance: Definitions, statistical model, findings, and process. In N. Ambady & J. J. Skowronski (Eds.), *First impressions* (pp. 129–146). New York: Guilford Press.

Kenny, D. A., & West, T. V. (2010). Similarity and agreement in self- and other perception: A meta-analysis. *Personality and Social Psychology Review*, 14, 196–213.

Kenny, D. A., West, T. V., Cillessen, A. H. N., Coie, J. D., Dodge, K. A., Hubbard, J. A., & Schwartz, D. (2007). Accuracy in judgments of aggressiveness. *Personality and Social Psychology Bulletin*, 33, 1225–1236.

Kilianski, S. E. (2008). Who do you think I think I am? Accuracy in perceptions of others' self-esteem. *Journal of Research in Personality*, 42, 386–398.

Küfner, A. C. P., Back, M. D., Nestler, S., & Egloff, B. (2010). Tell me a story and I will tell you who you are! Lens model analyses of personality and creative writing. *Journal of Research in Personality*, 44, 427–435.

Kurtz, J. E., & Sherker, J. L. (2003). Relationship quality, trait similarity, and self-other agreement on personality ratings in college roommates. *Journal of Personality*, 71, 21–48.

Leising, D., Erbs, J., & Fritz, U. (2010). The letter of recommendation effect in informant ratings of personality. *Journal of Personality and Social Psychology*, 98, 668–682.

Letzring, T. D. (2008). The good judge of personality: Characteristics, behaviors, and observer accuracy. *Journal of Research in Personality*, 42, 914–932.

Letzring, T. D., & Hall, J. A. (2012). *Accuracy of judging personality traits and affective states: A combined judgment model*. Unpublished manuscript.

Letzring, T. D., & Human, L. J. (2013). An examination of information quality as a moderator of accurate personality judgment: Information about thoughts and feelings and behaviors increases distinctive accuracy. *Journal of Personality*, 82, 440–451.

Letzring, T. D., Wells, S. M., & Funder, D. C. (2006). Quantity and quality of available information affect the realistic accuracy of personality judgment. *Journal of Personality and Social Psychology*, 91, 111–123.

Levesque, M. J., & Kenny, D. A. (1993). Accuracy of behavioral predictions at zero acquaintance: A social relations analysis. *Journal of Personality and Social Psychology*, 65, 1178–1187.

Lippa, R. A., & Dietz, J. K. (2000). The relation of gender, personality, and intelligence to judges' accuracy in judging strangers' personality from brief video segments. *Journal of Nonverbal Behavior*, 24, 25–43.

Murphy, N. A. (2007). Appearing smart: The impression management of intelligence, person perception accuracy, and behavior in social interaction. *Personality and Social Psychology Bulletin*, 33, 325–339.

Murphy, N. A., Hall, J. A., & Colvin, C. R. (2003). Accurate intelligence assessments in social interaction: Mediators and gender effects. *Journal of Personality*, 71, 465–493.

Murphy, N. A., Hall, J. A., Schmid Mast, M., Ruben, M. A., Frauendorfer, D., Blanch-Hartigan, D., Roter, D. L., & Nguyen, L. (2015). Reliability and validity of nonverbal thin slices in social interactions. *Personality and Social Psychology Bulletin*, 41, 199–213.

Naumann, L. P., Vazire, S., Rentfrow, P. J., & Gosling, S. D. (2009). Personality judgments based on physical appearance. *Personality and Social Psychology Bulletin*, 35, 1661–1671.

Nestler, S., & Back, M.D. (2013). Applications and extensions of the lens model to understand interpersonal judgments at zero acquaintance. *Current Directions in Psychological Science*, 22, 374–379.

Nestler, S., & Back, M.D. (in press). Using cross-classified structural equation models to examine the accuracy of personality judgments. *Psychometrika*.

Nestler, S., Egloff, B., Küfner, A. C. P., & Back, M. D. (2012). An integrative lens model approach to bias and accuracy in human inferences: Hindsight effects and knowledge updating in personality judgments. *Journal of Personality and Social Psychology*, 103, 698–717.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.

Norman, W. T., & Goldberg, L. R. (1966). Raters, ratees, and randomness in personality structure. *Journal of Personality and Social Psychology*, 4, 681–691.

Paulhus, D. L., & Bruce, M. N. (1992). The effect of acquaintanceship on the validity of personality impressions: a longitudinal study. *Journal of Personality and Social Psychology*, 63, 816–824.

Paunonen, S. V., & Kam, C. (2014). The accuracy of roommate ratings of behaviors versus beliefs. *Journal of Research in Personality*, 52, 55–67.

Qiu, L., Lin, H., Ramsay, J., & Yang, F. (2012). You are what you tweet: Personality expression and perception on Twitter. *Journal of Research in Personality*, 46, 710–718.

Rentfrow, P. J., & Gosling, S. D. (2006). Message in a ballad: The role of music preferences in interpersonal perception. *Psychological Science*, 17, 236–242.

Reynolds, D. J. Jr., & Gifford, R. (2001). The sounds and sights of intelligence: A lens model channel analysis. *Personality and Social Psychology Bulletin*, 27, 187–200.

Rosenthal, R., & Rubin, D. B. (1989). Effect size estimation for one-sample multiple-choice-type data: Design, analysis, and meta-analysis. *Psychological Bulletin*, 106, 332–337.

Stern, W. (1911). *Die differentielle Psychologie in ihren methodischen Grundlagen*. Leipzig: Barth (Reprint 1994, Bern: Huber).

Stopfer, J. M., Egloff, B., Nestler, S., & Back, M. D. (2014). Personality expression and impression formation in online social networks: An integrative approach to understanding the processes of accuracy, impression management, and meta-accuracy. *European Journal of Personality*, 28, 73–94.

Swann, W. B., Jr. (1984). Quest for accuracy in person perception: A matter of pragmatics. *Psychological Review*, 91, 457–477.

Taft, R. (1955). The ability to judge people. *Psychological Bulletin*, 52, 1–23.

Tskhay, K. O., & Rule, N. O. (2014). Perceptions of personality in text-based media and OSN: A meta-analysis. *Journal of Research in Personality*, 49, 25–30.

Tucker, L. R. (1964). A suggested alternative formulation in the developments by Hursch, Hammond, and Hursch and by Hammond, Hursch, and Todd. *Psychological Review*, 71, 528–530.

Vazire, S. (2010). Who knows what about a person? The self–other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology*, 98, 281–300.

Vazire, S., & Gosling, S. D. (2004). E-perceptions: Personality impressions based on personal websites. *Journal of Personality and Social Psychology*, 87, 123–132.

Vernon, P. E. (1933). Some characteristics of the good judge of personality. *Journal of Social Psychology*, 4, 42–57.

Vogt, D. S., & Colvin, R. C. (2003). Interpersonal orientation and the accuracy of personality judgments. *Journal of Personality*, 71, 267–295.

Wall, H. J., Taylor, P. J., Dixon, J. A., Conchie, S. M., & Ellis, D. A. (2013). Rich contexts do not always enrich the accuracy of personality judgements. *Journal of Experimental Social Psychology*, 49, 1190–1195.

Watson, D. (1989). Strangers' ratings of the five robust personality factors: Evidence of a surprising convergence with self-report. *Journal of Personality and Social Psychology*, 57, 120–128.

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

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, 238–249.

Zebrowitz, L. A., Voinescu, L., & Collins, M. A. (1996). "Wide eyed" and "crooked-faced": Determinants of perceived and real honesty across the life span. *Personality and Social Psychology Bulletin*, 22, 1258–1269.

6 Accuracy of perceiving social attributes

Ravin Alaei and Nicholas O. Rule

Abstract

A wealth of research shows that people can achieve accurate interpersonal judgments of others based on brief observations of their nonverbal cues. Here, we review evidence demonstrating that people can accurately judge others' kinship, sexual orientation, religious identity, political ideology, and professional success from subtle cues in their physical appearance and expressive behavior. Following this discussion, we detail some of the major factors that can influence the accuracy of these judgments. Finally, we end by reflecting on what this research has elucidated about basic processes in person perception and nonverbal behavior more generally.

From exchanging glances on the street to meeting for the first time at a party, people consistently infer others' social attributes. Such "snap judgments" are principally achieved by categorizing individuals into social groups, from which perceivers extrapolate additional evaluations using group-based stereotypes (Macrae & Bodenhausen, 2000). Although a rich literature documents the biases rife in social judgments (e.g., Merton, 1948), these impressions can also be accurate, as evidenced by individuals' near-perfect categorizations of others' age, race, and sex (e.g., Macrae & Martin, 2007).

Most social attributes are not demarcated as clearly as these "Big 3" dimensions, however. Here, we review the literature investigating judgments of social attributes that are perceptually ambiguous. We begin by discussing kin recognition, showing that the ability to accurately judge others' social attributes occurs across species. We then describe how the cognitive and perceptual machinery underlying such judgments is adaptive and flexible – evidenced by work demonstrating an average of approximately 64.5% accuracy in judging sexual orientation, religious identity, and political ideology from brief observations of nonverbal cues (Tskhay & Rule, 2013). We then extend this to research investigating the predictive validity of inferences based on subtle cues. Next, we argue for a nuanced understanding of accuracy by detailing some factors that affect one's ability to make accurate judgments. Finally, we conclude by discussing what this work has revealed about basic processes in person perception and nonverbal behavior more generally.

Kinship

Like many other animals, humans preferentially invest resources into their close relatives (Smith, Kish, & Crawford, 1987). Indeed, nepotism is evolutionarily favorable, as any gene that leads an individual to promote the welfare of his or her relatives will also promote its own survival (Hamilton, 1964). Alongside this benefit, accurate kin recognition can also help to prevent inbreeding, a costly mistake in terms of evolutionary fitness (Keller & Waller, 2002). Finally, accurate kin recognition can also advantage individuals to identify other people's kin for ascertaining alliances (Cheney & Seyfarth, 2004). Given these benefits, one would expect kin recognition to be accurate and pervasive; indeed, this is so.

In one early study, researchers found that people could accurately judge family relationships from short (two minutes or less) naturalistic videos of one to four people based on verbal and nonverbal cues – such as correctly judging that a woman conversing on the telephone was speaking with her mother (Costanzo & Archer, 1989). Brédart and French (1999) showed that kinship judgments could be made with even less information, reporting that people could accurately match children and parents from photos of their faces. Indeed, static facial cues can communicate kinship between grandparents and grandchildren (Kaminski, Dridi, Graff, & Gentaz, 2009), and between siblings (DeBruine et al., 2009; Maloney & Dal Martello, 2006). More intriguing, humans can also reliably judge the kinship of other (nonhuman) primates from photos of the offspring and parent faces (Alvergne, Huchard et al., 2009).

People can detect kinship from olfactory cues as well. For example, Porter, Cernoch, and Balogh (1985) found that strangers could accurately match mothers and children from shirts worn while sleeping (controlling for personal hygiene products), but could not match spouses, suggesting that olfactory kinship cues arise from genetic similarity, rather than environmental similarity alone. Indeed, further studies showed that people regard non-cohabiting identical twins (who are genetically indistinguishable) as smelling more alike than non-cohabiting dizygotic twins (who are genetically distinguishable; Roberts et al., 2005), and that mothers cannot recognize their cohabiting stepchildren (who share no genes with them) from their odor (Weisfeld, Czilli, Phillips, Gall, & Lichtman, 2003).

Olfactory cues also predict kin recognition within families. Mothers, for instance, can correctly recognize their neonates from their odors even only 20 hours after delivery (Porter, Cernoch, & McLaughlin, 1983). Reciprocally, neonates prefer their own mothers' breast pad odors to those of other women (MacFarlane, 1975). Moreover, odors allow parents to distinguish between their individual children, and allow children and adults to distinguish their parents and siblings (Porter & Moore, 1981; Weisfeld et al., 2003). Extended family members (e.g., grandmothers and aunts) also accurately judge kinship from odors (Porter, Balogh, Cernoch, & Franchi, 1986).

Research has therefore pervasively demonstrated that people can judge their own and strangers' kin through minimal information, reinforcing previous findings that this ability is shared across species (Lieberman, Tooby, & Cosmides, 2007). Moreover, such attunement of people's visual and olfactory perceptions to specific individuals suggests that the social perceptual system can flexibly discern subtle cues relevant to the current social environment. Below, we review research investigating accurate judgments of sexual orientation, religious identity, and political ideology to illustrate this further.

Sexual orientation

Consistent with gay men's and lesbians' reports, a growing literature shows evidence that sexual orientation can be accurately perceived from subtle cues (colloquially referred to as "gaydar;" Nicholas, 2004). Berger, Hank, Rauzi, and Simkins (1987) first tested this by presenting judges with 2–3-minute videotaped interviews of gay, lesbian, and straight individuals. Although they found no evidence of accuracy, a more sensitive reanalysis of the same data by Hallahan (1998) did. Ambady, Hallahan, and Conner (1999) then provided further evidence that people could accurately judge sexual orientation from dynamic nonverbal cues. They presented judges with 1-s or 10-s-silent video clips, or photographs of gay, lesbian, and straight individuals speaking, and found that judges could accurately categorize the speakers' sexual orientation across all conditions. Thus, both dynamic and static cues accurately communicated sexual orientation.

The robustness of these effects was extended by Rule, Ambady, Adams, and Macrae (2008), who showed that sexual orientation could be reliably judged from static cues in photos of gay and straight men's faces collected from online personal advertisements.[1] In this study, they demonstrated that a static face suffices to communicate sexual orientation, as do its individual features (with rates of approximately 65.7% accuracy for the entire face, 56% accuracy for the eyes alone, 57.5% accuracy for the mouth alone, and 62% accuracy for the hair alone). The human social perceptual system therefore seems calibrated to accurately perceive sexual orientation, even when only one facial cue is available (see Tskhay, Feriozzo, & Rule, 2013, for similar results with women's faces). Further investigation into the features underlying such judgments revealed that gay men tend to have wider and shorter faces, smaller and shorter noses, and bigger and more rounded jaws than straight men do (Valentova, Kleisner, Havlíček, & Neustupa, 2014; see also Skorska, Geniole, Vrysen, McCormick, & Bogaert, 2015). Thus, facial morphology alone can provide valid cues to sexual orientation, complementing other work using samples of expressive behavior (Ambady et al., 1999; Tskhay & Rule, 2015).[2]

Vocal cues also allow for accurate judgments of sexual orientation (Munson & Babel, 2007). Indeed, people are about 4% more accurate in judging sexual orientation from speech samples than from visual cues, on average (Tskhay & Rule, 2013). Despite the folk belief that gay men speak like straight women (with relatively high, variable pitch) and that lesbians speak like straight men (with relatively low, monotonous pitch; Levon, 2007), several studies have failed to detect such differences (e.g., Rendall, Vasey, & McKenzie, 2008). Rather, Linville (1998) found that judges were accurate when they used the duration and frequency of speakers' "s" sound (i.e., the voiceless alveolar fricative) to judge men's sexual orientation, and other researchers have detected differences in particular vowel sounds (Rendall et al., 2008).

As suggested by work on vocal cues, the extent to which gender inversion (i.e., the possession of characteristics typical of the opposite sex) accurately communicates sexual orientation may be exaggerated. Still, gendered cues, such as from facial appearance (Freeman, Johnson, Ambady, & Rule, 2010) and body movement (Johnson, Gill,

[1] One might expect that people will be especially motivated to accurately communicate their sexual orientation in online personal advertisements. However, people usually communicate traits that are counter-stereotypical in such ads (Bailey, Kim, Hills, & Linsenmeier, 1997). Indeed, judges' accuracy appears to be generally worse when based on self-selected photos taken from the Internet versus other sources (Tskhay & Rule, 2013) and sexual orientation, in particular, appears to be legible regardless of whether photos are self-posted by online daters (Rule & Ambady, 2008a), posted by friends (Rule et al., 2008), or photographed under standardized conditions in the lab (Stern, West, Jost, & Rule, 2013).

[2] Notably, expression can also influence facial morphology such that the two may be somewhat inextricable (see Malatesta, Fiore, & Messina, 1987).

Reichman, & Tassinary, 2007) can allow for accurate inferences of sexual orientation (e.g., explaining roughly 37% of the variance between gay and straight faces in Freeman et al., 2010); indeed, even home videos of gender-nonconforming children can be used to predict their sexual orientation in adulthood (Rieger, Linsenmeier, Gygax, & Bailey, 2008). Thus, there seem to be at least some cases in which gendered cues are valid indicators of sexual orientation.

Religious identity and political ideology

The Holocaust stimulated research on the accurate judgment of Jewish identity. During this time, people commonly believed that Jewish people could be identified through observation. Substantiating this, Allport and Kramer (1946) found that judges could categorize Jewish and non-Jewish individuals better than chance from yearbook photos, and Lund and Berg (1946) found that even preschoolers could discern Jewish identity from live observations providing both appearance and speech information, thus demonstrating that expressive cues communicate religious identity (though speech cues generally decreased accuracy). Further studies that statistically accounted for response biases also revealed above-chance accuracy somewhat consistently (e.g., Dorfman, Keeve, & Saslow, 1971), yet other studies did not (e.g., Elliott & Wittenberg, 1955). Two meta-analyses subsequently clarified this discord by demonstrating an overall significant, albeit small, level of accuracy for identifying Jewish individuals from nonverbal cues (approximately 55% accuracy for judgments made from the static face; Andrzejewski, Hall, & Salib, 2009; Rice & Mullen, 2003; but see Lund & Berg, 1946, for much higher accuracy from live presentations). In addition, people can differentiate Mormons and non-Mormons from photos of their faces with approximately 58% accuracy (Rule, Garrett, & Ambady, 2010a).

Aside from group-based differences in religious identity, individual variation in political ideology is also legible from facial cues. Samochowiec, Wänke, and Fiedler (2010) found that Swiss and German politicians' party memberships and political attitudes (i.e., right-wing versus left-wing) could be reliably judged from 15-s videos and photographs of their faces. This suggests that people are sensitive to facial cues communicating not only party membership, but also the extent to which one supports liberal or conservative views. These effects also extend to American politicians (e.g., Olivola, Sussman, Tsetsos, Kang, & Todorov, 2012): conservatives tend to be perceived as powerful, whereas liberals tend to be perceived as warm, facilitating accurate judgments of political ideology through facial morphology alone (Rule & Ambady, 2010).

Thus, the social perceptual system can detect subtle cues communicating perceptually ambiguous information, such as one's kinship, sexual orientation, and beliefs (e.g., political ideology, and even attitudes—see Chapter 7). We now review research indicating that people are sensitive to subtle cues that predict later success. It is worth pointing out that, although most of these studies demonstrate predictive validity rather than accuracy per se, we include them to show that perceivers can attune to how individuals' appearances correlate with real-world outcomes.

Professional success

On September 26, 1960, John Kennedy and Richard Nixon participated in the first televised US presidential debate. Undoubtedly an important event in American political history, this momentous day is also thought to have been an inadvertent demonstration of the striking influence that appearances can hold for real-world outcomes: those who had watched the debate on television believed that Kennedy had won, whereas those who had listened to the debate on the radio hailed Nixon the winner (Krauss, 1996). Today, a large body of research reflects what was suggested by reactions to the Kennedy–Nixon debate nearly 60 years ago: appearances can predict people's achievements.

Indeed, appearance seems to be an important factor in political outcomes. Todorov, Mandisodza, Goren, and Hall (2005) found that naïve ratings of American political candidates' competence from photos of their faces predicted the winning candidate, even when viewed for only 100 milliseconds (e.g., predicting the outcomes of 68.5% of gubernatorial races in Ballew & Todorov, 2007). Although this relationship between inferences of candidates' competence and their electoral success is meaningful, it is indirect and therefore does not measure accuracy. Moreover, the subjective nature of both the predictor variable (laboratory participants' opinions) and the outcome variable (voters' opinions) may simply suggest that candidates' faces are useful polls, rather than measures of political leaders' actual traits or effectiveness in office.

Other studies have demonstrated that direct inferences of success can predict individuals' actual performance, however. Rule and Ambady (2008b), for instance, found that first impressions of chief executive officers' (CEOs') leadership ability from their faces correlated with their companies' profitability – the standard for success in business. Although this association could arise because more profitable companies hire people who look like better leaders, Wong, Ormiston, and Haselhuhn (2011) found that CEOs' facial morphology predicted their companies' profits when controlling for the companies' financial performance prior to their tenure as CEO. This suggests that CEOs' appearances may validly indicate their leadership ability. Moreover, such inferences may be stable: Rule and Ambady (2011) found that evaluations of business leaders' power from their faces predicted their companies' profits across different photos taken decades apart – even before the individuals became business leaders.

In addition to static faces, other studies have found that dynamic, expressive nonverbal behavior also allows for accurate judgments of success. Benjamin and Shapiro (2009) showed that perceivers could predict election winners from 10-s silent videos of debates, and Tsay (2013) found that people could accurately judge the winners of music competitions from silent videos of their performances. Similarly, Tskhay, Xu, and Rule (2014) observed that naïve perceivers judged conductors' relative fame from brief silent videos of their live performances with approximately 62% accuracy. Given that conductors' success requires eliciting specific behaviors from their followers in a very intimate setting, these data show that the nonverbal behaviors of leaders of small groups relate to measures of their success like they do for the leaders of large groups who are very distant from their followers (e.g., CEOs and politicians), as described above.

Correlates and moderators

Alongside research that seeks to identify the cues leading to accurate judgments of social attributes, researchers have investigated some of the variables that moderate the relationships between individuals' perceptions and outcome criteria. Importantly, this work joins the efforts of other research to ascertain various correlates of nonverbal judgment accuracy (Hall, Andrzejewski, & Yopchick, 2009). Here, we review how research into the accurate judgments of kinship, sexual orientation, religious identity, and political ideology reveals several consistently influential correlates and moderators, thus offering a nuanced view of accuracy.

Context

Accuracy can vary in different contexts. For instance, people judge the sexual orientation of gay men, straight men, and straight women more accurately from speech samples in which they are conversing with a gay individual than they do from speech samples in which they are conversing with a straight individual (Carahaly, 2000). Thus, social context can affect the accurate perception of sexual orientation and the stereotypes present in a given social context may affect accuracy as well. For example, a meta-analysis showed that the years in which studies were published moderated the effect of prejudice on the accuracy of judging Jewish identity: prejudice related to greater accuracy in the past but diminished accuracy today (Andrzejewski et al., 2009). The authors speculated that higher prejudice against Jewish individuals previously resulted in greater accuracy because such prejudicial views were once normative (and thus an indicator of better social adjustment, which is associated with greater interpersonal accuracy), whereas now the reverse is true. Similarly, people who are more familiar with sexual minorities (including gay and lesbian perceivers) tend to be more accurate judges of sexual orientation from nonverbal cues (e.g., Brambilla, Riva, & Rule, 2013), and people with higher self-reported levels of anti-gay prejudice tend to perform worse (Rule, Tskhay, Brambilla, Riva, Andrzejewski, & Krendl, 2015).

Culture and race

Researchers have documented accuracy across numerous ethnic, racial, and cultural lines for a variety of judgments in the nonverbal communication literature (e.g., Zebrowitz et al., 1993; see also Chapter 16). Recent research suggests that this consistency also applies to the kinship judgments described above: both Senegalese and French judges displayed similar levels of accuracy when judging strangers' kin from both countries (Alvergne, Oda et al., 2009). Accuracy also extends across group boundaries for the other social attributes we have discussed, with some also showing an in-group advantage. For example, gay men judge male sexual orientation more accurately from faces than straight men do (Rule, Ambady, Adams, & Macrae, 2007), and Mormons can distinguish between Mormons and non-Mormons better than non-Mormons can (Rule, Garrett, & Ambady, 2010b).

Rule, Ishii, Ambady, Rosen, and Hallet (2011) asked perceivers from cultures with low (Japan), medium (the US), and high (Spain) acceptance of homosexuality to judge the sexual orientation of targets from all three nations, finding that natives of each country were able to accurately judge targets' sexual orientation regardless of their culture of origin, with Americans being the most accurate, possibly because of their greater propensity for intuitive judgments (see the Perceptual and Cognitive Mechanisms Underlying Accuracy section). Consistent with the cultures' level of acceptance, however, American and Japanese participants were less likely to categorize targets as gay compared to Spanish participants, suggesting that culture can affect one's openness to consider another individual as gay. Similarly, Valentova, Rieger, Havlíček, Linsenmeier, and Bailey (2011) demonstrated accuracy in judging the sexual orientation of Czech targets, and also identified an in-group advantage: US judges were more accurate for US targets, and Czech judges for Czech targets. Moreover, although target and participant race do not generally affect the accuracy of judging sexual orientation (Rule, 2011), the combination of racial and gender inversion stereotypes can facilitate accurate judgments. For instance, because Asian individuals are perceived as being feminine (relative to Caucasian individuals), the sexual orientation of Asian women is relatively easier to judge because any gender-atypical features, which are valid cues to homosexuality (Freeman et al., 2010), will be more salient and thereby facilitate judgments of homosexuality (Johnson & Ghavami, 2011).

Sex

Women are often better judges of nonverbal cues than men (Hall, 1984; see Chapter 15). This may apply to sexual orientation judgments based on dynamic cues but not the static face (e.g., Ambady et al., 1999; Rule, 2011). Moreover, some research indicates that women's sexual orientation is judged more accurately than men's sexual orientation from static cues (Tabak & Zayas, 2012), whereas judgments of men's sexual orientation may be more legible than women's when inferred from dynamic cues (Ambady et al., 1999; Johnson et al., 2007).

Motivation

Motivation may also influence accuracy (see Chapter 19). In the case of recognizing kin, Kaminski, Ravary, Graff, and Gentaz (2010) found that individuals with older siblings performed better than first-born individuals when judging kinship among strangers. They speculated that this arose because first-born individuals could rely on the fact that their siblings were born to their parents after them to judge kinship (e.g., they were present for their siblings' births). Later-born individuals, however, presumably had greater implicit motivation to develop alternative ways of detecting kinship, such as by facial cues, because they would not have been present to associate their siblings with their parents when their siblings were first born.

The ecological theory of social perception predicts that social perception functions to facilitate social action: when observing another person's nonverbal cues, people glean information that can guide the realization of their social goals (Zebrowitz & Montepare, 2006). Thus, perceivers should be able to quickly and accurately judge characteristics that are relevant to adaptive action (Gibson, 1979). Indeed, it would be adaptive for individuals to accurately judge sexual orientation in order to identify potential mates, especially when such motivations are heightened. Along these lines, Rule, Rosen, Slepian, and Ambady (2011) found that heterosexual women were significantly more accurate judges of men's (but not women's) sexual orientation the closer they were to peak ovulation, when they are most capable of conception. This accords with previous studies showing that women are more attentive to mating-related facial cues when ovulating (e.g., Penton-Voak et al., 1999). Moreover, when Rule, Rosen et al. (2011) experimentally manipulated women's motivation by priming them to think about romance, heterosexual women again showed better accuracy at judging men's but not women's sexual orientation, supporting the conclusion that women's motivational state encouraged the increase in accuracy.

Political ideology

Perceivers' political ideology also affects their accuracy in judging sexual orientation. Buttressed by the finding that conservatives tend to show a greater desire to reach certainty and typically rely more heavily on stereotypes in making judgments (e.g., Jost, Glaser, Kruglanski, & Sulloway, 2003), Stern, West, Jost, and Rule (2013) found that conservatives more accurately judged sexual orientation as the validity of the gender inversion stereotype increased. Moreover, forcing liberals to rely on their initial snap judgments increased the influence of gender inversion stereotypes on their categorizations as well, rendering their judgments similar to conservatives' because they were also then more likely to rely on stereotypes. Thus, to the extent that stereotypes about gender inversion accurately distinguish gay and straight individuals, conservatives are more effective than liberals in judging others' sexual orientation.

Perceptual and cognitive mechanisms underlying accuracy

Research on the accurate judgment of perceptually ambiguous social attributes has allowed the field to develop a more complete account of how people judge the majority of social dimensions, a question not fully addressed by studies that only examine obvious characteristics (e.g., sex). These efforts have revealed that the perceptual and cognitive processes underlying judgments of perceptually ambiguous social attributes parallel those supporting the accurate judgment of perceptually obvious social attributes. For example, just as people process perceptually obvious social attributes categorically (e.g., age), judgments of sexual orientation rely on a straight–non-straight dichotomy, such that bisexual individuals are judged as being different from heterosexual individuals, but are judged as belonging to the same category as gay and lesbian individuals (Ding & Rule, 2012). In other words, people judge others' sexual orientation in terms of discrete categories, rather than conceptualizing sexual orientation along a continuum (cf. Kinsey, Pomeroy, Martin, & Gebhard, 1953/1998).

Second, just as members of perceptually obvious social groups are categorized automatically (Macrae & Bodenhausen, 2000), this seems to extend to perceptually ambiguous groups as well. For instance, research generally shows that nonverbal kinship communication is primarily implicit: strangers posing as couples are accurately distinguished from actual couples simply told to pose together for a photo, suggesting that part of kinship is communicated unintentionally (Barnes & Sternberg, 1989; Sternberg & Smith, 1985). Kinship recognition, whether based on face matching or olfaction, may also be implicit, as participants still perform above chance when they are unaware of their performance level and feel that they are merely guessing (e.g., Arantes & Berg, 2012; Lundström, Boyle, Zatorre, & Jones-Gotman, 2009).

Similarly, in addition to the controllable and intentional cues that communicate sexual orientation (e.g., eye gaze, clothing; Nicholas, 2004; Rudd, 1996), sexual orientation can be automatically perceived from subtle cues as well. Cosmetics notwithstanding, people typically do not deliberately style the appearance of their eyes and mouths (Ekman & Friesen, 1969), yet both features independently permit accurate judgments of sexual orientation when perceived in isolation, suggesting that individuals may unintentionally provide cues to their sexual orientation through these features (e.g., Rule et al., 2008). Moreover, people can accurately judge sexual orientation with as little as a 40-ms glimpse of a person's face (Rule, Ambady, & Hallett, 2009). Thus, only very small amounts of time are needed to distinguish sexual orientation. This efficiency suggests that sexual orientation may be processed automatically, and direct tests of automaticity have supported this: people identified words relating to gay and straight stereotypes faster when preceded by photographs of gay and straight men's faces, respectively (Rule, Macrae, & Ambady, 2009), and deliberated judgments – which disrupt intuitive processes – impaired perceivers' judgments of sexual orientation (Rule, Ambady, & Hallett, 2009). Finally, sexual orientation can still be perceived from nonverbal cues when gay and lesbian targets deliberately attempt to conceal it (Sylva, Rieger, Linsenmeier, & Bailey, 2010).

These findings therefore suggest that kinship and sexual orientation are either part of the "master status" categories like age, race, and sex, or that the cognitive and perceptual machinery involved in these accurate judgments may be adaptive and flexible to perceiving a number of group distinctions. If the latter, one would expect that a variety of social attributes could be perceived accurately (i.e., better than chance). Considering the findings for religious identity, political ideology, and professional success reviewed above, this appears to be the case. What is more, some research also suggests that success and religious ideology are processed nonconsciously (e.g., Ballew & Todorov, 2007; Rule et al., 2010b). Thus, the cognitive machinery responsible for identifying perceptually ambiguous social attributes seems to be (i) fundamental to how social groups are perceived, (ii) flexible in its processing of group distinctions, and (iii) applicable beyond perceptually obvious categories.

Conclusion

In a meta-analysis of 47 articles investigating the accurate perception of ambiguous social groups, Tskhay and Rule (2013) found the aggregate effect size to be $r = .29$, indicating that people are correct in 64.5% of their judgments, on average (see Rosenthal & Rubin, 1982). Although 64.5% is much lower than the near-perfect accuracy attained when judging perceptually obvious groups (e.g., 99.2% accuracy for race; Remedios, Chasteen, Rule, & Plaks, 2011), it still demonstrates the mind's remarkable ability to glean important social information, whether for judging kinship, sexual orientation, religious identity, political ideology, or professional success. Of course, given such imperfect accuracy, a range of factors can affect people's ability to correctly judge others' social attributes, such as an individual's culture or sex. Nevertheless, such judgments appear to occur beneath conscious awareness, delicately making sense of the infinitely complicated social world in which we live.

References

Allport, G. W., & Kramer, B. M. (1946). Some roots of prejudice. *Journal of Psychology*, 22, 9–39.

Alvergne, A., Huchard, E., Caillaud, D., Charpentier, M. J. E., Setchell, J. M., Ruppli, C., Féjan, D., Martinez, L., Cowlishaw, G., & Raymond, M. (2009). Human ability to visually recognize kin within primates. *International Journal of Primatology*, 30, 199–210.

Alvergne, A., Oda, R., Faurie, C., Matsumoto-Oda, A., Durand, V., & Raymond, M. (2009). Cross-cultural perceptions of facial resemblance between kin. *Journal of Vision*, 9, 1–10.

Ambady, N., Hallahan, M., & Conner, B. (1999). Accuracy of judgments of sexual orientation from thin slices of behavior. *Journal of Personality and Social Psychology*, 77, 538–547.

Andrzejewski, S. A., Hall, J. A., & Salib, E. R. (2009). Anti-Semitism and identification of Jewish group membership from photographs. *Journal of Nonverbal Behavior*, 33, 47–58.

Arantes, J., & Berg, M. E. (2012). Kinship recognition by unrelated observers depends on implicit and explicit cognition. *Evolutionary Psychology*, 10, 210–224.

Bailey, J. M., Kim, P. Y., Hills, A., & Linsenmeier, J. A. (1997). Butch, femme, or straight acting? Partner preferences of gay men and lesbians. *Journal of Personality and Social Psychology*, 73, 960–973.

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.

Barnes, M. L., & Sternberg, R. J. (1989). Social intelligence and decoding of nonverbal cues. *Intelligence*, 13, 263–287.

Benjamin, D. J., & Shapiro, J. M. (2009). Thin-slice forecasts of gubernatorial elections. *The Review of Economics and Statistics*, 91, 523–536.

Berger, G., Hank, L., Rauzi, T., & Simkins, L. (1987). Detection of sexual orientation by heterosexuals and homosexuals. *Journal of Homosexuality*, 13, 83–100.

Brambilla, M., Riva, P., & Rule, N. O. (2013). Familiarity increases the accuracy of categorizing male sexual orientation. *Personality and Individual Differences*, 55, 193–195.

Brédart, S., & French, R. M. (1999). Do babies resemble their fathers more than their mothers? A failure to replicate Christenfeld and Hill (1995). *Evolution and Human Behavior*, 20, 129–135.

Carahaly, L. (2000). *Listener accuracy in identifying the sexual orientation of male and female speakers*. Unpublished manuscript, The Ohio State University.

Cheney, D. L., & Seyfarth, R. M. (2004). The recognition of other individuals' kinship relationships. In B. Chapais & C. M. Berman (Eds.), *Kinship and behavior in primates* (pp. 347–365). Oxford: Oxford University Press.

Costanzo, M., & Archer, D. (1989). Interpreting the expressive behavior of others: The Interpersonal Perception Task. *Journal of Nonverbal Behavior*, 13, 225–245.

DeBruine, L. M., Smith, F. G., Jones, B. C., Craig Roberts, S., Petrie, M., & Spector, T. D. (2009). Kin recognition signals in adult faces. *Vision Research*, 49, 38–43.

Ding, J. Y. C., & Rule, N. O. (2012). Gay, straight, or somewhere in between: Accuracy and bias in the perception of bisexual faces. *Journal of Nonverbal Behavior*, 36, 165–176.

Dorfman, D. D., Keeve, S., & Saslow, C. (1971). Ethnic identification: A signal detection analysis. *Journal of Personality and Social Psychology*, 18, 373–379.

Ekman, P., & Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry*, 32, 88–106.

Elliott, D. N., & Wittenberg, B. H. (1955). Accuracy of identification of Jewish and non-Jewish photographs. *Journal of Abnormal and Social Psychology*, 57, 339–341.

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.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton-Mifflin.

Hall, J. A. (1984). *Nonverbal sex differences: Communication accuracy and expressive style*. Baltimore, MD: Johns Hopkins University Press.

Hall, J. A., Andrzejewski, S. A., & Yopchick, J. E. (2009). Psychosocial correlates of interpersonal sensitivity: A meta-analysis. *Journal of Nonverbal Behavior*, 33, 149–180.

Hallahan, M. (1998). *Reanalysis of Berger, Hank, Rauzi, and Simkins, 1987*. Unpublished manuscript, Clemson University.

Hamilton, W. D. (1964). The genetical evolution of social behaviour I. *Journal of Theoretical Biology*, 7, 1–16.

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

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

Jost, J. T., Glaser, J., Kruglanski, A. W., & Sulloway, F. (2003). Political conservatism as motivated social cognition. *Psychological Bulletin*, 129, 339–375.

Kaminski, G., Dridi, S., Graff, C., & Gentaz, E. (2009). Human ability to detect kinship in strangers' faces: Effects of the degree of relatedness. *Proceedings of the Royal Society B: Biological Sciences*, 276, 3193–3200.

Kaminski, G., Ravary, F., Graff, C., & Gentaz, E. (2010). Firstborns' disadvantage in kinship detection. *Psychological Science*, 21, 1746–1750.

Keller, L. F., & Waller, D. M. (2002). Inbreeding effects in wild populations. *Trends in Ecology and Evolution*, 17, 230–241.

Kinsey, A. C., Pomeroy, W. B., Martin, C. E., & Gebhard, P. H. (1998). *Sexual behavior in the human female*. Bloomington, IN: Indiana University Press. (Original work published 1953).

Krauss, S. (1996). Winners of the first 1960 televised presidential debate between Kennedy and Nixon. *Journal of Communication*, 46, 78–96.

Levon, E. (2007). Sexuality in context: Variation and the sociolinguistic perception of identity. *Language in Society*, 36, 533–554.

Lieberman, D., Tooby, J., & Cosmides, L. (2007). The architecture of human kin detection. *Nature*, 445, 727–731.

Linville, S. E. (1998). Acoustic correlates of perceived versus actual sexual orientation in men's speech. *Folia Phoniatrica et Logopaedica*, 50, 35–48.

Lund, F. H., & Berg, W. C. (1946). Identifiability of nationality characteristics. *Journal of Social Psychology*, 24, 77–83.

Lundström, J. N., Boyle, J. A., Zatorre, R. J., & Jones-Gotman, M. (2009). The neuronal substrates of human olfactory based kin recognition. *Human Brain Mapping*, 30, 2571–2580.

Macfarlane, A. (1975). Olfaction in the development of social preferences in the human neonate. In R. Porter & M. O'Connor (Eds.), *Parent-infant interaction* (pp. 103–113). Amsterdam: Elsevier.

Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology*, 51, 93–120.

Macrae, C. N., & Martin, D. (2007). A boy primed Sue: Feature-based processing and person construal. *European Journal of Social Psychology*, 37, 793–805.

Malatesta, C. Z., Fiore, M. J., & Messina, J. J. (1987). Affect, personality, and facial expressive characteristics of older people. *Psychology and Aging*, 2, 64–69.

Maloney, L. T., & Dal Martello, M. F. 2006. Kin recognition and the perceived facial similarity of children. *Journal of Vision*, 6, 1047–1056.

Merton, R. K. (1948). The self-fulfilling prophecy. *The Antioch Review*, 8, 193–210.

Munson, B., & Babel, M. (2007). Loose lips and silver tongues, or, projecting sexual orientation through speech. *Language and Linguistics Compass*, 1, 416–449.

Nicholas, C. L. (2004). Gaydar: Eye-gaze as identity recognition among gay men and lesbians. *Sexuality and Culture*, 8, 60–86.

Olivola, C. Y., Sussman, A. B., Tsetsos, K., Kang, O. E., & Todorov, A. (2012). Republicans prefer republican-looking leaders: Political facial stereotypes predict candidate electoral success among right-leaning voters. *Social Psychological and Personality Science*, 3, 605–613.

Penton-Voak, I. S., Perrett, D. I., Castles, D. L., Kobayashi, T., Burt, D. M., Murray, L. K., & Minamisawa, R. (1999). Menstrual cycle alters face perception. *Nature*, 399, 741–742.

Porter, R. H., & Moore, J. D. (1981). Human kin recognition by olfactory cues. *Physiology and Behavior*, 27, 493–495.

Porter, R. H., Balogh, R. D., Cernoch, J. M., & Franchi, C. (1986). Recognition of kin through characteristic body odors. *Chemical Senses*, 11, 389–395.

Porter, R. H., Cernoch, J. M., & Balogh, R. D. (1985). Odor signatures and kin recognition. *Physiology and Behavior*, 34, 445–448.

Porter, R. H., Cernoch, J. M., & McLaughlin, F. J. (1983). Maternal recognition of neonates through olfactory cues. *Physiology and Behavior*, 30, 151–154.

Remedios, J. D., Chasteen, A. L., Rule, N. O., & Plaks, J. E. (2011). Impressions at the intersection of ambiguous and obvious social categories: Does gay + Black = likable? *Journal of Experimental Social Psychology*, 47, 1312–1315.

Rendall, D., Vasey, P. L., & McKenzie, J. (2008). The Queen's English: An alternative, biosocial hypothesis for the distinctive features of "gay speech." *Archives of Sexual Behavior*, 37, 188–204.

Rice, D. R., & Mullen, B. (2003). Isaac, Ishmael, and Janus: Past and future lessons regarding the ethnic categorization of faces. *Applied Cognitive Psychology*, 17, 1129–1147.

Rieger, G., Linsenmeier, J. A., Gygax, L., & Bailey, J. M. (2008). Sexual orientation and childhood gender nonconformity: Evidence from home videos. *Developmental Psychology*, 44, 46–58.

Roberts, S. C., Gosling, L. M., Spector, T. D., Miller, P., Penn, D. J., & Petrie, M. (2005). Body odor similarity in noncohabiting twins. *Chemical Senses*, 30, 651–656.

Rosenthal, R., & Rubin, D. B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology*, 74, 166–169.

Rudd, N. A. (1996). Appearance and self-presentation research in gay consumer cultures: Issues and impact. *Journal of Homosexuality*, 31, 109–134.

Rule, N. O. (2011). The influence of target and perceiver race in the categorization of male sexual orientation. *Perception*, 40, 830–839.

Rule, N. O., & Ambady, N. (2008a). Brief exposures: Male sexual orientation is accurately perceived at 50-ms. *Journal of Experimental Social Psychology*, 44, 1100–1105.

Rule, N. O., & Ambady, N. (2008b). The face of success: Inferences from Chief Executive Officers' appearance predict company profits. *Psychological Science*, 19, 109–111.

Rule, N. O., & Ambady, N. (2010). Democrats and Republicans can be differentiated from their faces. *PLoS ONE*, 5, e8733.

Rule, N. O., & Ambady, N. (2011). Judgments of power from college yearbook photos and later career success. *Social Psychological and Personality Science*, 2, 154–158.

Rule, N. O., Ambady, N., Adams, R. B., Jr., & Macrae, C. N. (2007). Us and them: Memory advantages in perceptually ambiguous groups. *Psychonomic Bulletin & Review*, 14, 687–692.

Rule, N. O., Ambady, N., Adams, R. B., Jr., & Macrae, C. N. (2008). Accuracy and awareness in the perception and categorization of male sexual orientation. *Journal of Personality and Social Psychology*, 95, 1019–1028.

Rule, N. O., Ambady, N., & Hallett, K. C. (2009). Female sexual orientation is perceived accurately, rapidly, and automatically from the face and its features. *Journal of Experimental Social Psychology*, 45, 1245–1251.

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

Rule, N. O., Garrett, J. V., & Ambady, N. (2010b). Faces and places: Geographic environment influences the ingroup memory advantage. *Journal of Personality and Social Psychology*, 98, 343–355.

Rule, N. O., Ishii, K., Ambady, N., Rosen, K. S., & Hallett, K. C. (2011). Found in translation: Cross-cultural consensus in the accurate categorization of male sexual orientation. *Personality and Social Psychology Bulletin*, 37, 1449–1507.

Rule, N. O., Macrae, C. N., & Ambady, N. (2009). Ambiguous group membership is extracted automatically from faces, *Psychological Science*, 20, 441–443.

Rule, N. O., Rosen, K. S., Slepian, M. L., & Ambady, N. (2011). Mating interest improves women's accuracy in judging male sexual orientation. *Psychological Science*, 22, 881–886.

Rule, N. O., Tskhay, K. O., Brambilla, M., Riva, P., Andrzejewski, S. A., & Krendl, A. C. (2015). The relationship between anti-gay prejudice and the categorization of sexual orientation. *Personality and Individual Differences*, 77, 74–80.

Samochowiec, J., Wänke, M., & Fiedler, K. (2010). Political ideology at face value. *Social Psychological and Personality Science*, 1, 206–213.

Skorska, M. N., Geniole, S. N., Vrysen, B. M., McCormick, C. M., & Bogaert, A. F. (2015). Facial structure predicts sexual orientation in both men and women. *Archives of Sexual Behavior*, 44, 1377–1394.

Smith, M. S., Kish, B. J., & Crawford, C. B. (1987). Inheritance of wealth as human kin investment. *Ethology and Sociobiology*, 8, 171–182.

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.

Sternberg, R. J., & Smith, C. (1985). Social intelligence and decoding skills in nonverbal communication. *Social Cognition*, 3, 168–192.

Sylva, D., Rieger, G., Linsenmeier, J. A., & Bailey, J. M. (2010). Concealment of sexual orientation. *Archives of Sexual Behavior*, 39, 141–152.

Tabak, J. A., & Zayas, V. (2012). The roles of featural and configural face processing in snap judgments of sexual orientation. *PloS ONE*, 7: e36671.

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

Tsay, C. J. (2013). Sight over sound in the judgment of music performance. *Proceedings of the National Academy of Sciences*, 110, 14580–14585.

Tskhay, K. O., & Rule, N. O. (2013). Accuracy in categorizing perceptually ambiguous groups: A review and meta-analysis. *Personality and Social Psychology Review*, 17, 72–86.

Tskhay, K. O., & Rule, N. O. (2015). Emotions facilitate the communication of ambiguous group memberships. Emotion, 15, 812–826.

Tskhay, K. O., Feriozzo, M. M., & Rule, N. O. (2013). Facial features influence the categorization of female sexual orientation. *Perception*, 42, 1090–1094.

Tskhay, K. O., Xu, H., & Rule, N. O. (2014). Perceptions of leadership success from nonverbal cues communicated by orchestra conductors. *The Leadership Quarterly*, 25, 901–911.

Valentova, J. V., Kleisner, K., Havlíček, J., & Neustupa, J. (2014). Shape differences between the faces of homosexual and heterosexual men. *Archives of Sexual Behavior*, 43, 353–361.

Valentova, J., Rieger, G., Havlíček, J., Linsenmeier, J. A., & Bailey, J. M. (2011). Judgments of sexual orientation and masculinity–femininity based on thin slices of behavior: A cross-cultural comparison. *Archives of Sexual Behavior*, 40, 1145–1152.

Weisfeld, G. E., Czilli, T., Phillips, K. A., Gall, J. A., & Lichtman, C. M. (2003). Possible olfaction-based mechanisms in human kin recognition and inbreeding avoidance. *Journal of Experimental Child Psychology*, 85, 279–295.

Wong, E. M., Ormiston, M. E., & Haselhuhn, M. P. (2011). A face only an investor could love: CEOs' facial structure predicts their firms' financial performance. *Psychological Science*, 22, 1478–1483.

Zebrowitz, L. A., Montepare, J. M., & Lee, H. K. (1993). They don't all look alike: Individuated impressions of other racial groups. *Journal of Personality and Social Psychology*, 65, 85–101.

Zebrowitz, L. A., & Montepare, J. M. (2006). The ecological approach to person perception: Evolutionary roots and contemporary offshoots. In M. Schaller, J. A. Simpson, & D. T. Kenrick (Eds.), *Evolution and social psychology* (pp. 81–113). New York: Psychology Press.

7 Accuracy of judging group attitudes

Tessa V. West

Abstract
How accurate are people's stereotypes about groups? And how accurate are people in knowing what others think of the groups they belong to? The goal of this chapter is to provide an overview of conceptual and methodological approaches to studying accuracy in people's attitudes about in-groups and out-groups, and to provide a brief review of empirical findings that address such accuracy. I focus on two central questions that scholars have addressed: One, are people accurate in their judgments about groups? And two, are people accurate in reading what others think about groups to which they belong (i.e., meta-perceptions of attitudes about groups)? I first discuss methodological and conceptual approaches to studying group-based attitudes, including a discussion of the process through which a valid truth criterion is selected, the different ways in which the relationship between the truth and the judgment can be operationalized, and the level at which accuracy is measured. I then review findings from research on accuracy of group-based attitudes using a motivation-based framework to understand why perceivers might be accurate or inaccurate in their judgments. Finally, I propose several avenues for future research, with an emphasis on research designed to provide evidence of the process through which perceivers become accurate and biased in their own attitudes about groups and their perceptions of others' attitudes about groups.

Imagine that Jen has started a new job at an engineering firm, and she has been assigned to work with her new colleague Darnell on a project. In this interaction context, Jen's and Darnell's abilities to accurately detect each other's thoughts, feelings, and intentions can directly affect their ability to communicate effectively with each other. If Jen is one of a handful of women in her firm, gender will likely be a salient social category that serves as a lens through which Jen and Darnell evaluate each other. Does Darnell believe that women are not as competent as men at engineering tasks, in general?

If so, are Darnell's attitudes grounded in reality? Do Darnell's attitudes about women in general bias his perceptions of Jen's competence? Is Jen accurate in detecting Darnell's attitudes about women's competence, and importantly, about whether he thinks she is particularly competent?

In this opening example, I have touched upon a few questions that scholars can address in the study of how accurate people are in their attitudes about groups (e.g., their attitudes about female engineers in general), and about individuals as representative members of those groups (e.g., particular female engineers like Jen). The goal of this chapter is to provide an overview of conceptual and methodological approaches to studying people's accuracy of attitudes about groups. In my review of the literature, I focus on two central questions in the study of accuracy throughout: One, are people's attitudes about groups accurate? Here, I will draw from prior research on *stereotype accuracy*, using the working definition of a stereotype as "a set of beliefs about the personal attributes of a social group" (Ashmore & Del Boca, 1981, p. 21; see also Jussim, Cain, Crawford, Harber, & Cohen, 2009). For example, are people accurate in their judgments of how talkative women are? Two, are people accurate in their judgments of other people's attitudes about groups? For example, are women accurate in their judgments of how talkative men *think* women are? And are they accurate in their judgments of how talkative women *think* women are? Here, I will focus on the accuracy of metaperceptions, or our beliefs about other people's group-based attitudes. I provide a brief review of the research to date that has examined these questions.

For both types of questions, I will review different conceptual and methodological approaches to studying accuracy of beliefs that people hold about groups, including the process through which a valid truth criterion is selected, the different ways in which the relationship between the truth and the judgment is operationalized, and the level at which accuracy is measured (i.e., is accuracy measured for individual perceivers who judge individual targets such as in dyadic interactions, for multiple perceivers who judge multiple targets as in group interactions, or for individual perceivers who rate groups as a whole). I will then discuss the role of motivation in achieving accuracy of attitudes about groups. Finally, I propose several avenues for future research, with an emphasis on new lines of research designed to provide evidence of the process through which perceivers become accurate and biased in their own group-based attitudes and their perceptions of other's group-based attitudes. Generally speaking, an attitude can be any psychological tendency that one can have toward an attitude object (Eagly & Chaiken, 1998), and many chapters in this book focus on accuracy for other types of attitudes that individuals hold that are not about groups in particular. As such, I review research that exclusively addresses accuracy about attitudes that are about groups in particular.

Conceptual and methodological approaches to studying accuracy of group-based attitudes

In studying the accuracy of attitudes about groups, there are important methodological decisions that one must make when designing a study. In this section, I briefly discuss three issues: the selection of a valid truth criterion, the way in which the relationship between the judgment and the truth is conceptualized and measured, and the level at which accuracy is measured.

Selecting a truth criterion

The selection of the truth criterion is a critical methodological step for any researcher who plans to study accuracy (see Jussim, 2012; Stern, West, & Schoenthaler, 2013; West & Kenny, 2011). When researchers examine whether people's attitudes about groups are accurate, they need to select a valid truth criterion that reflects where groups actually stand on the trait(s) they are examining, and one important step is selecting criteria groups that are *representative* of the groups that perceivers provide attitudes about (Jussim et al., 2009). Failing to do so can lead to a disconnect between who is being judged and who is providing the truth for that group.

For example, imagine a study that examines the question: Are people's attitudes about the relative athletic abilities of Blacks compared to Whites accurate? People provide ratings of the extent to which they think Black athletes, in general, are more skilled than White athletes. Because perceivers are judging Black and White athletes *in general*, the truth criteria data need to be drawn from a representative sample of Black and White athletes that is sufficiently large and includes athletes who play many different sports (e.g., a large sample of football players, tennis players, basketball players, and baseball players). If the criteria groups of Blacks and Whites are not representative samples (e.g., data from only pro-basketball players are used), or the groups from whom criteria data are drawn are limited (data are from one team only), then the researchers would be limited in the conclusions they can draw about accuracy of beliefs about Whites' and Blacks' athletic abilities. It would also be problematic if participants are told to make judgments about one group (e.g., athletes in general), but their judgments are compared to a truth criterion drawn from another group (e.g., the Lakers). In this case the researchers might be "stacking the deck" against accuracy by not making clear to perceivers whom they are supposed to be responding about (for a more developed argument on giving accuracy a "fighting chance," see Funder, 1995). These issues could be easily resolved by making clear to perceivers who is providing the data for the criterion groups.

How might researchers go about selecting a representative sample for criterion groups? One option is to capture actual group differences using meta-analytic data. For example, Swim (1994) and Hall and Carter (1999) examined accuracy of perceivers' beliefs about differences between men and women. Are men and women's verbal abilities, leadership abilities, and happiness (among other traits) as different as people think they are? Swim (1994) compared participants' estimates of mean differences and variability between men and women's attitudes with results of actual differences in means and variability on these dimensions, using Hyde and Linn's (1986) meta-analysis of actual gender differences, where the truth criteria included objective measures, such as verbal tests and nonverbal behaviors. Swim (1994) also assessed accuracy using items for which she had behavioral truth criteria data (e.g., SAT scores) and that perceivers (male and female college students) had familiarity with. As such, Swim (1994) was able to avoid asking participants to make judgments about groups on dimensions that they had no prior knowledge about. Her results indicated some evidence of accuracy in that perceptions of mean and variability differences mapped onto actual differences (Study 1), and some evidence of inaccuracy, in that perceivers underestimated actual gender differences (Study 2).

In some cases, meta-analytic data are not available, but researchers can utilize prior research to obtain truth criteria. As an example of this approach, Chan et al. (2012) had participants from 26 countries rate the personality of typical adolescents, adults, and older people. Participants made ratings of each group using the National Character Survey (Terracciano et al., 2005), which consisted of 30 bipolar items. The truth criterion for each of the groups was drawn from published research that provided self- or observer reports for each of the age groups from the countries in which the judgment ratings were collected. Raters across nations tended to share similar beliefs about different age groups (e.g., adolescents were seen as impulsive, rebellious, undisciplined). These consensual age group stereotypes correlated strongly with published age differences on the same dimensions. One potential issue with this study, however, is that some of the truth criteria data were assessed using self-report, which might be biased by stereotypes as well. People may see themselves as consistent with stereotypes about their group (i.e., self-stereotype). For this reason, a behavioral accuracy criterion might be ideal for examining accuracy here.

In addition to using published empirical findings, another option for obtaining a truth criterion is to use data from large-scale survey data. For example, Saguy, Tausch, Dovidio, and Pratto (2009) examined accuracy in Israelis' meta-perceptions of Palestinians' attitudes about them using poll data (studies described in more detail below), and McCauley and Stitt (1978) used census data to examine the accuracy of attitudes about Whites and African Americans. Ashton and Esses (1999) examined accuracy of attitudes about the relative achievements of nine Canadian ethnic groups; their truth criterion was grades of the nine ethnic groups published by the Toronto Board of Education (more details described later).

In summary, there are many important considerations for selecting a proper truth criterion for studies that assess accuracy in people's attitudes about groups. Potential criterion data include data from meta-analyses, prior research findings, and survey data.

Measuring the relationship between the truth and the judgment

A second important consideration when studying accuracy of attitudes people hold about groups is how to best conceptualize the relationship between the truth criterion and the judgment – a decision that should be guided by the theoretical question of interest. There are many statistical approaches to date, which I will briefly review.

Two of the most common methods of operationalizing accuracy, especially in the study of accuracy of attitudes about groups, are (a) to compute discrepancy scores between the judgment and the truth, and (b) to compute correlations between the judgment and the truth for each kind of judgment, or across judgments within the perceiver. The first method allows one to examine whether perceivers over- or underestimate where the target group stands on a trait, over- or underestimate what others think their attitudes are, or whether they over- or underestimate the differences between groups. In some cases, researchers are not interested in the direction of inaccuracy, and they compute absolute difference scores (e.g., Saguy et al., 2009). The second method captures the strength of the relationship between the truth and the judgment, and allows one to assess whether some perceivers show a stronger overlap between the truth and the judgment (when estimated across judgments within perceivers) than others, and for what types of judgments there is a stronger overlap between the truth and judgment (when estimated across perceivers separately for each judgment). Some conceptual models estimate both forms of accuracy simultaneously. For example, in the Truth and Bias model (West & Kenny, 2011), *directional bias* captures the mean discrepancy between the truth and the judgment – stronger positive values indicate greater overestimation of the truth (e.g., people see others as more extraverted than they actually are), and stronger negative values indicate greater underestimation of the truth (e.g., people see others as less extraverted than they actually are). The *truth force* represents the strength of the effect of the truth on the judgment. The model can be estimated by treating the judgment as the outcome variable, and the truth criterion as the predictor variable, using a regression-based approach. By subtracting the mean of the truth off the judgment (assuming both are measured using the same scale), the intercept in this model represents directional bias and the main effect of the truth represents the truth force (estimated as a B weight).

Another method of assessing accuracy using the correlational approach is to compare a perceiver's rank ordering of attitudes to how these items actually rank among a group (akin to a profile correlation in personality research). The rank-ordering approach allows researchers to assess accuracy in attitudes about the relative frequency of certain traits – are perceivers accurate in knowing how common certain traits are relative to others? (Jussim, 2012). For example, are women more communal than they are agentic? There are two potential challenges with the rank-ordering approach: one, determining the actual frequency of traits for a group (i.e., establishing a valid truth criterion) and two, choosing traits for which it makes conceptual sense to compare their relative frequency. For example, it might be difficult to establish what it means for communality to occur more frequently than agency. Do perceivers have an accurate working model of what a lot and little agency look like, and what a lot and a little communality look like? Are the thresholds for a lot and a little of these two traits the same? Although challenging, addressing these issues is likely worth the effort, as this approach could be used to answer previously unexplored questions such as whether perceivers understand the intricate complexities of groups, and also, whether they understand how traits uniquely fit together to inform a "big picture" understanding of groups.

Accuracy of attitudes about the level of heterogeneity of groups

In addition to examining whether attitudes about particular traits are accurate, researchers can also ask the question: Do people accurately perceive how much heterogeneity there is within groups and between groups, and do they perceive differential amounts of heterogeneity depending on if they are judging their own group, or an out-group? For example, do people know the extent to which students at inner city schools in New York score similarly on standardized tests? Do they think students' test scores are more similar to each other than they actually are, or more different than they actually are? According to Park and Judd (1990) (see also Judd, Ryan, & Park, 1991), two forms of accuracy of heterogeneity can be empirically assessed that capture accuracy of knowing how diverse groups are. One, *perceived group dispersion* tests whether the perceived dispersion of individuals around their central tendency matches how these individuals actually disperse around their central tendency. A group that is perceived to be less variable is one that is perceived to be tightly bunched around the central tendency. A group that is perceived to be more variable is more dispersed around the central tendency. For example, imagine that students at the inner city school have an actual range of test scores that vary from 0 to 100, with a mean of 50 but a standard deviation of 10. If the group is perceived to be less variable, people might estimate that students on average have a score of 50 but that the standard deviation is quite small, 2. In other words, they assume that students' scores are "tightly bunched" around the mean. However, if they believe that the mean is 50 but the standard deviation is 20, then they believe that students' scores are loosely dispersed around the mean. The other, *perceived group stereotypicality*, tests whether people's beliefs about the prevalence of individuals who are stereotypical matches the prevalence of individuals who are actually stereotypical. A group that is perceived as less variable by this definition is one in which a relatively large percentage of the group is perceived as possessing the stereotype and a small percentage is counterstereotypical. For example, imagine that in the school study, people reported on the percentage of students who failed the high-school entrance exam. They may underestimate the percentage of students who do so, or overestimate it. These two forms of variability are conceptually related, but they are often empirically distinct and only moderately correlated.

How are accuracy of perceived group dispersion and perceived group assessed? Judd et al. (1991) describe three tasks that can be used to assess them. One, the group distribution dot task (Park & Judd, 1990), in which perceivers are asked to think about the group as a whole (e.g., Asians) and indicate the relative number of group members who would fall at each point along a dimension; two, the percentage estimation task, in which perceivers are asked to provide the percentage of group members who would have a trait (e.g., are good at math) or who would endorse an attitude; and three, the mean and range estimation task, in which perceivers rate where on average the group falls on a scale, and where on the scale the most extreme members fall (e.g., the Asian who is the worst at math, and the Asian who is best at math). The scores from these three tasks are then compared to the actual mean and standard deviation of a truth criterion that includes that actual mean and standard deviation for the group, and the percentage of group members who fall into each level of the trait (e.g., the percentage of Asians with math scores of 70, and with math scores of 90). To measure accuracy, correlations are computed between the truth and the judgment. For example, the perceived standard deviation from the dot task is correlated with the actual standard deviations from the truth criteria data (for more details see Judd et al., 1991). Judd et al. (1991) used this strategy to demonstrate that people overestimated stereotypicality for out-group members – that is, they show that stereotypes are overgeneralizations and that people are actually less stereotypical than people think they are.

Comparing accuracy of attitudes about in-group to out-group members

Finally, scholars may be interested in making comparisons between in- and out-group accuracy: Are individuals more accurate when judging the in-group versus out-group? Does accuracy of in-group versus out-group judgments differ as a function of the valence of the attitude, and of the stereotypicality of the attitude? Judd and Park's (1993) approach is designed to answer these questions. Targets and perceivers are separated into two groups (e.g., in- vs. out-groups, such as men and women), and judgments are made by and of in-group and out-group members on dimensions that are and not stereotypical for that group. For example, in a study looking at accuracy of attitudes about men and women's competence and warmth (men being more stereotypically competent and women being stereotypically warm), one could have men rate men and women, and women rate men and women, on competence and warmth. Actual levels of competence and warmth would also be needed.

Judd and Park's (1993) model is a three-way analysis of variance with the following factors: perceiver (male vs. female), target (male vs. female), and attribute (stereotypical for the in- or out-group, such as competence for men and warmth for women). The model yields the following parameters of interest: *Elevation accuracy*, which is the extent to which perceivers over- or underestimate attributes, averaged over all perceivers, targets, and attributes; the *perceiver group effect*, which is an overall tendency of one group to over- or underestimate all attributes of other groups (above and beyond elevation accuracy); *the target group effect*, which is an overall tendency for one group of targets to have all their attributes (added together) over- or underestimated (beyond the elevation effect). For example, are women seen as more of (all traits) than men? The *attribute effect* is the tendency to over- or underestimate a type of attribute (e.g., those about physical appearance). For example, if men are judging how agentic women are and how much leadership they show in the workplace (using a set of attributes to tap into each), then this effect refers to the tendency to see the group as more or less "stereotypical" than they actually are (assuming a shared understanding of the stereotype).

The effect that is the most of interest for studies of accuracy of in-group versus out-group attitudes is the three-way interaction between the perceiver group, target group, and attribute effects. This interaction tests whether under- or overestimation of stereotype is most likely to occur when people are judging in-group versus out-group members, and if so, whether this is the case for both groups (i.e., men judging women, women judging men). For example, are women seen as more stereotypical than men, but only when they are judged by other men? I provide a further example of this model in the next section on the motivational determinants of accuracy.

In sum, there are many ways of conceptualizing accuracy for attitudes about groups, all of which can provide unique insight into when people are accurate (e.g., in judging in-groups and out-groups), and for what types of accuracy (e.g., in gauging mean levels of a target group, in comparing groups to each other, or in estimating the level of heterogeneity within a group). Next, I briefly review the research that examines the accuracy of group-based attitudes, cutting across a diverse set of attitudes, types of groups, and methodologies. Throughout this review, I will highlight one mechanism that plays a central role in how scholars have theorized about the process through which accuracy of group-based attitudes is achieved: motivation.

Motivational determinants of accuracy of group-based attitudes

How might perceivers become accurate in their attitudes about groups? Many studies of accuracy of attitudes about groups have emphasized the role of motivational factors in determining how accurate people are. Motivation has been theorized to affect accuracy directly – that is, people should be motivated to be accurate and this motivation might influence how accurate they are – and indirectly by affecting another psychological process that affects accuracy, such as a drive to perceive similarity. I review research that illustrates the indirect and direct effects of motivation on accuracy of people's attitudes about groups.

Motivation indirectly affects accuracy

In this section, I will discuss the research that tests the idea that motivation, broadly construed, will affect one or more psychological processes (e.g., assuming similarity, stereotype utilization; what West & Kenny, 2011, refer to as bias), which in turn affect accuracy. Thus, motivation *indirectly* affects accuracy via its effects on another psychological process.

Ashton and Esses (1999) theorized that the motivation to not use stereotypes when evaluating groups might lead perceivers to be less accurate, to the extent that these stereotypes are grounded in reality. In other words, motivation might indirectly affect accuracy via its effect on stereotype usage. As detailed in the section on selecting a truth criterion, the authors examined accuracy of rank ordering of the achievement of nine Canadian ethnic groups, and used data from the Toronto Board of Education to obtain actual achievement data for each group. They found that people were quite accurate in their rank ordering of group means, and they also had accurate notions of between-group variability in academic achievement.

Moreover, they found that people who underestimated between-group variability (i.e., assumed the ethnic groups were more similar than they were in terms of achievement) were lower in Right Wing Authoritarianism (RWA) than were accurate estimators and overestimators. The authors propose that people who are low on RWA are committed to equality, and this commitment can serve as a motivation to reject the notion that ethnic groups truly differ on a socially important variable such as academic performance. In his review of their work, Jussim (2012) couches this finding in terms of liberal versus conservative politics (suggesting that RWA is really a proxy for ideology in this study). He proposes that liberals might be motivated to deny the existence of stereotype accuracy, and insofar as stereotypes are accurate, the failure to utilize stereotypes when comparing groups will lead to inaccuracy. Thus, stereotype utilization might be one indirect path through which perceivers become accurate, and the motivation to be egalitarian affects stereotype utilization and therefore affects accuracy indirectly.

In a similar vein, Stern, West, Jost, and Rule (2013) tested the hypothesis that liberals are motivated to not utilize stereotypes when categorizing perceptual ambiguous group members – that is, group members for whom appearance alone is not a valid indicator of group membership – into distinct categories. The authors examined whether liberals and conservatives categorize men into the gay and straight based only on their facial appearance. They also tested whether people utilize stereotypes about the association between facial appearance and sexual orientation (specifically, that gay men have more feminine features than straight men, and straight men have more masculine features than gay men), when doing so. The authors found that liberals, but not conservatives, corrected away from stereotype utilization when categorizing men as gay or straight; their categorization judgments did not correlate with the actual femininity and masculinity of the targets. To the extent that gay men are actually more feminine than straight men (and straight men are more masculine than gay men), the failure to utilize stereotypes about appearance would lead liberals to be less accurate because perceivers are failing to take stereotype accuracy into account when categorizing targets. In further support of a motivated correction process, the authors showed that when liberals were under cognitive load, which inhibits the correction process, they utilized stereotypes to the same degree as conservatives. These findings suggest that liberals are motivated to not utilize group-based stereotypes when making judgments, and to the extent to which these stereotypes are grounded in reality (which is an open question) they would be less accurate.

Ryan and Bogart (2001) examined a different motivational factor that might indirectly affect accuracy. The authors examined how accuracy of in-group and out-group members changes over time when people join a new group, utilizing Judd et al.'s (1991) approach. Accuracy of *perceived dispersion* was measured (i.e., variation of group members around the mean of the group on stereotypic attributions). The authors hypothesized that when people first join a new group, the motivation to reduce uncertainty and anxiety that characterize the socialization phase (Ryan & Bogart, 1997) leads new members to focus on the similarities of group members, resulting in less accuracy in dispersion. As the socialization process proceeds, new members shift their focus from how everyone is similar to the ways in which they differ. This process is more likely to occur for in-group than out-group judgments for functional reasons – individuals need to accurately read how different in-group members are from each other in order to function more effectively within day-to-day social interactions (see also Swann, 1984, for a similar argument regarding "pragmatic accuracy," or accuracy needed to navigate one's social world). Participants were sorority members who reported their attitudes about their own sorority (in-group) and other sororities (out-groups) during the first year of membership. They made evaluations that were stereotypic of that particular sorority, or counterstereotypic, and also positively (e.g., competitive, sophisticated) or negatively valenced (conceited, loud). Self-ratings were used as the truth criterion. They found that participants initially underestimated in-group more than out-group dispersion (i.e., they assumed in-group members were more similar to each other than they were, and more so than they assumed that out-group members were more similar to each other than they were). But over time, in-group dispersion judgments became more accurate whereas out-group dispersion judgments became less accurate. This study is an example of how the motivation to want to see similarity can indirectly lead perceivers to be less accurate, to the extent that group members are more different from each other than they are perceived to be.

Swim (1994) also examined accuracy of in-group as compared to out-group attitudes in the context of gender to test the following questions: One, do perceivers accurately know men and women's attitudes, and more specifically, how similar and different they are, and two, does in-group favoritism moderate accuracy in estimating differences between men and women's attitudes? To the extent that perceivers demonstrate in-group favoritism, they might be more motivated to see their own group more positively than they actually are, which could decrease accuracy. In this study, accuracy was operationalized in two ways. One, as the correspondence between individual's perceptions of the size of gender differences and meta-analytic findings about gender differences in social behaviors (i.e., the truth criterion); two, as sensitivity correlations between judgment and the truth (Judd & Park, 1991) which allow one to assess whether people are sensitive to relative differences among attributes. She found that for most traits, perceivers were accurate or underestimated differences; they only overestimated men's tendency to be aggressive and women's verbal abilities. Women perceived greater gender differences in the ability to decode nonverbal behaviors than did men, and men's failure to recognize these differences – that is, thinking that men were just as good as women at reading nonverbal behaviors when they were actually worse – led them to be less accurate. Women engaged in greater in-group favoritism for ratings of helping when alone and leadership; that is, they perceived women to be more positive on these traits. This bias to see women more positively explains in part why women were less accurate than men for these traits.

Hall and Carter (1999) made important headway in understanding why some people are more accurate in their knowledge of sex differences than others. The authors calculated stereotype accuracy judgment scores for each participant and then provided them with a standardized test to assess accuracy at decoding nonverbal behaviors (Profile of Nonverbal Sensitivity (PONS) test), as well as self-reported measures that assessed the extent to which they were likely to use stereotypes. They found that people who were more accurate at reading nonverbal behaviors had greater accuracy in knowing actual gender differences, whereas those who were more likely to endorse stereotypes were less accurate. These findings suggest that accuracy in knowing actual differences between men and women was obtained through actual observation, rather than simply endorsing stereotypes. In other words, stereotype usage biased perceivers' judgments, leading to less accuracy.

As a final example, Li and Hong (2001) examined accuracy in the ability of Mainland Chinese and Hong Kong students to read in-group and out-group members' values (i.e., items related to collectivism, such as striving for common good, altruism, cooperation, and individualism). The study was conducted in Hong Kong, which provides a unique context to study accuracy of intergroup attitudes between majority and minority group members. Interactions between local Hong Kongers (the majority group) and people from Mainland China (the minority group) have become more frequent following the return of sovereignty in 1997, but historically the groups have been separated. The authors proposed that because of relatively few interactions with the minority group, members of the majority group would assume that minority groups would differ from themselves, and to the extent that majority groups perceived themselves to be higher status, they would be motivated to achieve distinctiveness and differentiation from the out-group (see also Mummendey, Otten, Berger, & Kessler, 2000; Schwartz, Struch, & Bilsky, 1990; Wilder, 1986). To the extent that groups underestimate similarity, they will be less accurate when they use these similarity judgments to guide their estimates (i.e., an indirect effect of bias on accuracy in West & Kenny, 2011). Both the mainland Chinese and Hong Kongers were more accurate at estimating in-group than out-group values (consistent with Judd et al., 1991). The mainland group was also more accurate than the Hong Kong (minority) group, and they projected more onto the out-group than did the Hong Kong group – that is, they assumed more similarity. Both groups engaged in greater in-group than out-group projection (consistent with Krueger & Zeiger, 1993). These findings support the notion that the higher status majority group (Hong Kongers) perceived greater out-group distinctiveness – i.e., assumed less similarity – than the minority group. To the extent that these groups were not actually distinct, then this bias would lead to less accuracy.

Motivation directly affects accuracy

In addition to the indirect effect that motivation can have on accuracy, scholars have also theorized that certain types of perceivers should be more motivated to be accurate than others, because it helps them navigate their social environments. That is, there should be a direct effect of motivation on accuracy. For example, Hehman, Leitner, Deegan, and Gaertner (2013) examined how the differential motivation of Whites and Blacks to accurately read Whites targets' levels of prejudice leads minorities to achieve greater accuracy in these judgments because in order to navigate their social worlds, they need to have an accurate understanding of who might be prejudiced against them. The authors examined the ability of people to read prejudiced attitudes using the facial width-to-height ratio (fWHR) of male targets. They found that White men's fWHR correlated with explicit racial attitudes (the fWHR ratio is a visible manifestation of testosterone exposure, and testosterone is associated with social dominance motives). The authors theorized that men with higher fWHR are more likely to report prejudicial beliefs because they are less inhibited than men with lower fWHR. In support of this argument, they found that fWHR correlated with explicit prejudice (measured using the Attitudes toward Blacks scale, and the Internal Motivation to Respond Without Prejudice scale; Brigham, 1993, and Plant & Devine, 1998), suggesting that fWHR is a valid cue through which prejudice can be perceived. Perceivers then rated how racist they thought each participant was, on a 1–6 scale.

The authors found that perceivers were able to accurately detect the target's self-reported explicit prejudice, via the utilization of fWHR as a cue. They further showed that for Blacks, the motivation to be accurate (which was higher than it was for Whites), contributed in part to them making accurate judgments.

Similarly, Richeson and Shelton (2005) found that Black perceivers, relative to White perceivers, were able to detect racism at above-chance levels in White people engaging in interracial interactions by observing nonverbal interactions (20-s clips). The authors argued that Blacks should be particularly motivated to accurately detect Whites' racial attitudes, given the pragmatic utility of accuracy for navigating social interactions, even when given very little information about Whites, and none in which racial attitudes are directly expressed.

As another example, Saguy and Kteily (2011) examined in- and out-group members' accuracy in knowing each other's attitudes about conflict in the context of Israeli–Palestinian relations. They argued that the ability to accurately know the out-group during times of war has implications for the strategies groups' use, such as facilitating constructive initiatives or antagonistic ones, and this should especially be the case for low-power groups. In two studies, they used a mean difference approach to examine the extent to which people under- or overestimated in- and out-group members' goals pertaining to a conflict (i.e., they rated the extent to which each of eight goals guided an Israeli operation called the flotilla incident, such as "undermine Hamas," and "strengthen the image of the Israeli defense forces"). The authors used absolute difference scores because they were not interested in the direction of inaccuracy. In Study 1, they found that Israelis, who perceived political loss to their group (following the flotilla incident), were more accurate in predicting out-group (Palestinian) views, than were Palestinians, who perceived out-group political gains.

The authors also tested the idea that the more people perceive in-group losses within a conflict, the more motivated they would be to seek relevant information about the conflict from the out-group and to understand the out-group's perspective on the incident. In Study 2, they examined how accurate Israelis were in their perceptions of Palestinian views about them, specifically testing the hypothesis that for Israelis, accuracy in reading Palestinian views was predicted by perceived political losses. In other words, perceived in-group losses should motivate accuracy in knowing the out-group's thoughts. They examined accuracy in knowing whether Palestinians supported or opposed hurting Israeli civilians inside the green line (Israelis' borders prior to the 1967 War), measured using poll data. The authors found that the more Israelis perceived in-group losses, the more accurate they were in knowing Palestinians' attitudes about hurting civilians inside the 1967 borders. Taken together, these studies show that perceived losses to one's in-group predict accuracy in reading the out-group member's beliefs about the in-group. Across three situations in which Palestinians were viewed as gaining politically from the incident, Israelis were more accurate in reading Palestinians' views than vice versa. For Israelis who were strongly identified with their group, those who felt that their group was losing politically were even more accurate. These findings provide strong support for a motivational explanation for accuracy.

In summary, I have provided a brief review of studies that have addressed how perceivers become accurate in perceptions of attitudes about groups, with an emphasis on the role of motivation, and the different ways in which motivation has been conceptualized. In the next section, I outline several different avenues of future research that aim to extend and complement preexisting research on this question.

Future research

One question that is relatively underexplored in the study of accuracy of attitudes about groups is: How does accuracy for one operationalization of attitude – such as one's explicit attitude – relate to their accuracy for another operationalization of their attitude – such as their implicit attitude? For example, are people equally accurate at detecting implicit forms of racial attitudes – such as those that tap into nonconsciously held beliefs that often predict more subtle, nonverbal behaviors (Dovidio, Kawakami, & Gaertner, 2002) – as they are at detecting explicit forms of racial attitudes that predict more deliberative behaviors? Moreover, under what circumstances might they be especially likely to accurately perceive an out-group member's racial attitudes at the implicit and explicit levels?

As an example, Heyman et al. (2013) theorized that explicit but not implicit attitudes would be readable via the fWHR because this ratio is associated with psychological constructs related to dominance, which is related to explicit but not implicit prejudice. These findings suggest that utilizing fWHR to perceive implicit bias would not lead to accurate judgment because it is not a valid cue of targets' actual implicit prejudice levels. If this is the case, we can ask the question, how do perceivers know what cues are valid indicators of implicit prejudice, and what cues are valid indicators of explicit prejudice? In many social interaction contexts, individuals do not provide clear, valid indicators of their attitudes (e.g., telling people they are prejudiced); hence it would be interesting to examine whether some perceivers have developed the ability to simultaneously perceive an interaction partner's implicit and explicit prejudice, and if they are able to do so because they have an accurate working model of how these cues fit together to form a "prejudice profile" for the target.

In one of the first studies to examine how accurate people are in detecting different forms of sexism, Rudman and Fetterolf (2014) examined men and women's accuracy in detecting hostile and benevolent sexism – two forms of sexism that although fall under the umbrella of sexism are conceptually and empirically distinct. Male and female participants completed the Ambivalent Sexism Inventory (ASI; Glick & Fiske, 1996) to measure their own endorsement and perceptions of a typical out-group member's endorsement of hostile sexism and benevolent sexism (i.e., men rated women and women rated men). Here, the authors operationalized accuracy as the mean level difference between perceived and actual levels of sexism. They found that women overestimated men's hostile sexism, but they underestimated men's benevolent sexism. Men, in contrast, overestimated women's benevolent sexism but underestimated women's hostile sexism. The authors discuss how their findings provide support for the idea of an "illusion of antagonism" between hostile and benevolent sexism; individuals falsely assume that these two constructs are negatively related, but they are actually positively related and represent two underlying constructs of sexism that together reinforce the gender hierarchy (Rudman & Glick, 2008). Interestingly, benevolent sexists are often not labeled as sexist, and women, who often score just as high as men on the benevolent sexism scale, are unaware of its negative consequences. Unlike hostile sexism, which may be easier to identify, benevolent sexism can be more difficult to identify in others, leading to a mismatch in the levels of accuracy that perceivers achieve for these two forms of sexism. Rudman and Fetterolf's (2014) work is a nice example of how certain types of attitudes – such as those that masquerade as positive attitudes about certain groups – can be difficult to read because perceivers do not have an accurate working model of the attitude construct in question. As a consequence, perceivers do not know what cues they should be attending to in order to detect these attitudes in others.

In an interesting extension of this work, Goh, Rad, and Hall (2015) tested whether Rudman and Fetterolf's (2014) findings hold in a dyadic context, in which men and women judged each other's hostile and benevolent sexism during an interpersonal interaction. In Studies 1 and 2, participants provided ratings of their own and their perceived partner's hostile and benevolent sexism after a brief social interaction. The authors largely replicated Rudman and Fetterolf (2014). On the one hand, women overestimated men's hostile sexism, but estimation of men's benevolent sexism was not significant. Men, on the other hand, underestimated their female partners' hostile sexism and overestimated benevolent sexism. In their Study 3, they found that looking at 30-second silent clips produced above-chance levels for both kinds of sexism.

The work of Rudman and Fetterolf (2014) and Goh et al. (2015) are the first to test how accurate perceivers are for two forms of sexism – hostile and benevolent – at the generalized and dyadic levels. Future work could further explore the mechanisms through which perceivers become accurate in reading these two forms of sexism, and how the lack of understanding that benevolent sexism is an attitude construct might directly hinder their ability to perceive it in others.

Another area for future research is to further explore how accurate people's meta-perceptions are of how out-groups see their in-groups. To date, much research has explored the accuracy of attitudes about out-groups, but far less research has explored how accurate people are in knowing how others see their groups. For example, one could extend Ashton and Esses (1999) to test the question: Do Blacks know how Whites rank their levels of achievement relative to other groups? That is, do they know where, on the totem pole of achievement, Whites place Blacks? Further, is it better to be accurate or positively biased in knowing how an out-group sees an in-group (e.g., to assume that out-group members place your group higher on the totem pole than they actually do)? On the one hand, a positivity bias might help Blacks perform well in achievement settings that are threatening and anxiety provoking, such as being in a predominantly White school. To the extent that Blacks believe that out-group members have a more positive attitude about them than they actually do, they might experience a number of benefits, including greater self-efficacy, less anxiety and threat, and greater comfort working directly with Whites. However, there might be a dark side to positivity biases in perceptions of outgroups' attitudes as well, such as creating false expectations that one will be treated as an equal when in reality they are discriminated against (Saguy et al., 2009). Future research could test these competing hypotheses within numerous "high-stakes" contexts in which accuracy and positivity bias could be both beneficial and costly for perceivers.

As an extension of this idea, one could also examine the process through which people become accurate in their meta-perceptions of how the out-group perceives the in-group. One possibility is to test how having a strong identification with the in-group indirectly leads to greater accuracy. For example, it is possible that people who strongly identify with the in-group are more likely to see themselves in stereotypical terms (i.e., they self-stereotype), and upon making judgments of how out-group members see them, they project their stereotype-based self-perceptions onto those judgments. That is, they are accurate because they see themselves as stereotypical, and so they assume that others do too. To the extent that others utilize the same stereotypes when evaluating them, they will be accurate.

Indeed, there is a strong history of research demonstrating that meta-perceptions are strongly tied to perceivers' self-perceptions (see Kenny & DePaulo, 1993, and also Chapter 8), and this research would be extending this work to the domain of attitudes about one's group. There are some groups for whom this might be likely to occur. For example, Stern, West, and Schmitt (2014) found that conservatives are more likely to assume that they are similar to other conservatives, whereas liberals see themselves as distinct from other liberals. Conservatives therefore might be more likely to self-stereotype – readily identify with stereotypes about their group – and see themselves in a stereotype-consistent manner. In evaluating how the out-groups (liberals) see them, they might project their self-perceptions – not as individual but as a member of the group "conservatives" – onto how they think liberals see them. Insofar as liberals see conservatives in a stereotype-consistent manner, conservatives will achieve meta-accuracy. These findings would suggest that conservatives' accurate knowledge of how liberals see them would be achieved through a bias of self-stereotyping. The different paths through which people achieve meta-accuracy in how groups view them likely have important implications for intergroup relations, such as how conservatives and liberals work together.

In this section, I have highlighted a few avenues of future research that scholars might take in studying accuracy of group-based attitudes. These are certainly not exclusive, and there are a number of other potential directions that I have not touched upon, and contexts in which accuracy should be studied. Scholars should continue to dive further into the process through which accuracy of group-based attitudes is achieved, including utilizing new methods to examine these processes beyond the individual, dyad, and small group. For example, recent theoretical models of how person perception processes operate within broad social networks (e.g., Smith & Collins, 2009; see also Denrell, 2005) could be utilized to study accuracy of group-based attitudes in new and exciting ways.

Conclusion

The goal of this chapter is to provide an overview of research that addresses how accurate people are in their attitudes about groups, and how accurate they are in reading others' attitudes about the groups they belong to. I have touched upon a number of conceptual and methodological considerations when studying these questions, with the goal to not only highlight the complexity involved in studying accuracy of attitudes about groups, but also provide readers with an overview of the many possible ways they can conceptualize accuracy to go about asking new and important questions. Although a full review of all empirical findings is beyond the scope of this chapter, I have outlined the role of motivation in shaping how people might become more or less accurate, and how motivation might be theorized to directly influence accuracy, or indirectly, but affecting a psychological process that influences accuracy. I also touched upon new research that taps into different components of attitudes that might underlie a common attitude construct, and offered some suggestions for future research. In so doing, I emphasized the importance of studying the "when" and "how" of accuracy rather than the "either/or," as doing so will provide much needed insight into the mechanisms of accuracy of group-based attitudes.

References

Ashmore, R. D., & Del Boca, F. K. (1981). Conceptual approaches to stereotypes and stereotyping. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 1–35). Hillsdale, NJ: Lawrence Erlbaum Associates.

Ashton, M. C., & Esses, V. M. (1999). Stereotype accuracy: Estimating the academic performance of ethnic groups. *Personality and Social Psychology Bulletin*, 25, 225–236.

Brigham, J. (1993). College students' racial attitudes. *Journal of Applied Social Psychology*, 23, 1933–1967.

Chan, W., Mccrae, R. R., Fruyt, F. D., Jussim, L., Löckenhoff, C. E., Bolle, M. D., & Terracciano, A. (2012). Stereotypes of age differences in personality traits: Universal and accurate? *Journal of Personality and Social Psychology*, 103, 1050–1066.

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.

Eagly, A. H., & Chaiken, S. (1998). Attitude structure and function. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., pp. 269–322). New York: McGraw-Hill.

Funder, D. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, 102, 652–670.

Glick, P., & Fiske, S. T. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology*, 70, 491–512.

Goh, J. X., Rad, R. A., & Hall, J. A. (2015). *Bias and accuracy in detecting sexism in mixed gender interactions*. Unpublished manuscript.

Hall, J. A., & Carter, J. D. (1999). Gender-stereotype accuracy as an individual difference. *Journal of Personality and Social Psychology*, 77, 350–359.

Hehman, E., Leitner, J., Deegan, M., & Gaertner, S. (2013). Facial structure is indicative of explicit support for prejudicial beliefs. *Psychological Science*, 24, 289–296.

Hyde, J. S., & Linn, M. C. (Eds.). (1986). *The psychology of gender: Advances through meta-analysis*. Baltimore, MD: Johns Hopkins University Press.

Judd, C. M., Ryan, C. S., & Park, B. (1991). Accuracy in the judgment of in-group and out-group variability. *Journal of Personality and Social Psychology*, 61, 366–379.

Judd, C. M., & Park, B. (1993). Definition and assessment of accuracy in social stereotypes. *Psychological Review*, 100, 109–128.

Jussim, L., Cain, T., Crawford, J., Harber, K., & Cohen, F. (2009). The unbearable accuracy of stereotypes. In T. D. Nelson (Ed.), *Handbook of prejudice, stereotyping, and discrimination* (pp. 199–227). New York: Psychology Press Taylor & Francis Group.

Jussim, L. (2012). *Social perception and social reality: Why accuracy dominates bias and self-fulfilling prophecy*. New York: Oxford University Press.

Kenny, D. A., & DePaulo, B. M. (1993). Do people know how others view them? An empirical and theoretical account. *Psychological Bulletin*, 114, 145–161.

Krueger, J., & Zeiger, J. S. (1993). Social categorization and the truly false consensus effect. *Journal of Personality and Social Psychology*, 65, 670–680.

Li, Q., & Hong, Y. (2001). Intergroup perceptual accuracy predicts real-life intergroup interactions. *Group Processes & Intergroup Relations*, 4, 341–354.

McCauley, C., & Stitt, C. (1978). An individual and quantitative measure of stereotypes. *Journal of Personality and Social Psychology*, 36, 929–940.

Mummendey, A., Otten, S., Berger, U., & Kessler, T. (2000). Positive–negative asymmetry in social discrimination: Valence of evaluation and salience of categorization. *Personality and Social Psychology Bulletin*, 26, 1258–1270.

Park, B., & Judd, C. M. (1990). Measures and models of perceived group variability. *Journal of Personality and Social Psychology*, 59, 173–191.

Plant, E., & Devine, P. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, 75, 811–832.

Richeson, J. A., & Shelton, J. N. (2005). Thin slices of racial bias. *Journal of Nonverbal Behavior*, 29, 75–86.

Rudman, L. A., & Fetterolf, J. C. (2014). How accurate are metaperceptions of sexism? Evidence for the illusion of antagonism between hostile and benevolent sexism. *Group Processes and Intergroup Relations*, 17, 275–285.

Rudman, L. A., & Glick, P. (2008). *The social psychology of gender: How power and intimacy shape gender relations*. New York: Guilford.

Ryan, C. S., & Bogart, L. M. (1997). Development of new group members' in-group and out-group stereotypes: Changes in perceived variability and ethnocentrism. *Journal of Personality and Social Psychology*, 73, 719–732.

Ryan, C. S., & Bogart, L. M. (2001). Longitudinal changes in the accuracy of new group members' in-group and out-group stereotypes. *Journal of Experimental Social Psychology*, 37, 118–133.

Saguy, T., Tausch, N., Dovidio, J. F., & Pratto, F. (2009). The irony of harmony: Intergroup contact can produce false expectations for equality. *Psychological Science*, 20, 114–121.

Saguy, T., & Kteily, N. (2011). Inside the opponent's head: Perceived losses in group position predict accuracy in metaperceptions between groups. *Psychological Science*, 22, 951–958.

Schwartz, S. H., Struch, N., & Bilsky, W. (1990). Values and intergroup social motives: A study of Israeli and German students. *Social Psychology Quarterly*, 53, 185–198.

Smith, E. R., & Collins, E. C. (2009). Contextualizing person perception: Distributed social cognition. *Psychological Review*, 116, 343–364.

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.

Stern, C. V., West, T. V., & Schmitt, P. G. (2014). The liberal illusion of uniqueness. *Psychological Science*, 25, 137–144.

Stern, C., West, T. V., & Schoenthaler, A. (2013). The dynamic relationship between accuracy and bias in social perception research. *Social and Personality Psychology Compass*, 7, 303–314.

Swann, W. B. (1984). Quest for accuracy in person perception: A matter of pragmatics. *Psychological Review*, 91, 457–477.

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

Terracciano, A., Abdel-Khalek, A. M., Adam, N., Adamovova, L., Ahn, C. K., Ahn, H. N., . . . & Meshcheriakov, B. (2005). National character does not reflect mean personality trait levels in 49 cultures. *Psychological Science*, 310, 96–100.

West, T. V., & Kenny, D. A. (2011). The truth and bias model of judgment. *Psychological Review*, 118, 357–378.

Wilder, D. A. (1986). Cognitive factors affecting the success of intergroup contact. In S. Worchel & W. G. Austin (Eds.), *Psychology of intergroup relations* (2nd ed., pp. 49–66). Chicago, IL: Nelson-Hall.

8 Metaperceptions

Do people know how others perceive them?

Erika N. Carlson and Maxwell Barranti

Abstract

Metaperceptions, or beliefs about how other people perceive the self, are the implicit maps people use to navigate complex social environments. Are metaperceptions accurate? The answer to this question is complex and depends on several factors, such as how insight is measured, the attribute in question, and the social context. We first review several ways in which the accuracy of metaperceptions is typically conceptualized and measured. We then summarize for which attributes (e.g., intelligent, likeable) and in which contexts (e.g., among friends or coworkers) metaperceptions are accurate as well as for whom (e.g., personality traits, status) and in which situations. Next, we consider the process of metaperception and which sources of information lead people to form accurate beliefs about how others perceive them. Finally, we discuss future directions that may shed more light on when people know how others experience them and how to potentially improve this type of insight.

Did I make a good first impression? Does my boss think I am competent? Implicitly or explicitly, people often think about the impressions they make on others, and these beliefs are called *metaperceptions*. Ultimately, metaperceptions are the implicit maps people use to navigate their social worlds, and metaperceptions powerfully shape people's behavior, the quality of their relationships, and their identity (Felson & Reed, 1986; Lemay & Dudley, 2009). Given the pivotal role metaperceptions play in everyday life, the natural question is are they accurate? The main goal of this chapter is to summarize for which types of attributes, in which social contexts, and for whom metaperceptions are accurate. Our discussion also outlines mechanisms that might facilitate or hinder people's ability to infer how others see them and provides several future directions about how people can improve the accuracy of their metaperceptions.

The accuracy of metaperceptions

When are metaperceptions accurate? The answer to this question is complex, because there are many ways to conceptualize and measure the degree to which people know how others perceive them. For the sake of brevity, we use the term *insight* to refer to people's ability to infer how others perceive them. Before outlining if and when people have insight, we first explain how insight is typically measured, using the example of Pete and Meg. Conceptually, Pete, the perceiver, forms an impression of Meg, the metaperceiver, and then Meg forms a belief about how Pete views her (i.e., a metaperception). Insight is the relationship between Pete's impression of Meg and Meg's metaperception. As with most social judgments, there are at least two ways to think about this relationship: (a) accuracy, or the correlation between impressions and (b) bias, or the mean-level difference between impressions (West & Kenny, 2011). In the context of metaperception, *meta-accuracy* is the correlation between metaperceptions and impressions and represents, for example, whether the people who think they are seen as funnier are actually seen as being funnier. *Metaperception bias* is the mean-level difference between metaperceptions and impressions and represents, for example, whether people overestimate or underestimate how funny others think they are. Meta-accuracy and metaperception bias are conceptually and statistically independent indices of insight. For example, relative to the people who think they are seen as less humorous, the people who think they are seen as more humorous may actually be seen as more humorous (i.e., meta-accuracy), but at the same time, metaperceivers might systematically overestimate how humorous they are seen by other people (i.e., metaperception bias). Indeed, a meta-analysis revealed that people's metaperceptions are generally accurate but also slightly, positively biased (Carlson & Kenny, 2012).

Insight can also be conceptualized and measured as awareness of how a specific person perceives the self (e.g., how Meg's romantic partner, Jon, sees her) or as awareness of one's reputation across several perceivers (e.g., Meg's reputation among her friends). The former index of insight can be measured between two people, but the latter form of insight requires multiple perceivers, such as in a round robin, block, or one-with-many design. In a round robin design, each person in a group (e.g., Meg, Pete, Jill, James, and Joan) provides impressions of each group member and estimates how every other person in their group perceived them (i.e., provides metaperceptions). For example, individuals in a work group might rate every other group member's leadership abilities and guess how each member rated their leadership abilities (Malloy & Janowski, 1992). A block design is similar, but a subset of members (e.g., women) rate and are rated by another subset (e.g., men). In a one-with-many design, a single metaperceiver (e.g., Meg) estimates how multiple perceivers (e.g., Pete, Jill, James, and Joan) perceive him or her, and perceivers provide impressions of the metaperceiver. The one-with-many design is especially useful when the metaperceiver and perceivers have unique roles (e.g., supervisor and subordinates) or when perceivers cannot rate one another because they are not acquainted. The main benefit of these multiple-perceiver designs is that they can reveal whether people know their reputation, which is called *generalized meta-accuracy* (GMA), as well as whether people know who in a group sees them as higher or lower on a trait, which is called *dyadic meta-accuracy* (DMA) (Kenny, 1994). With these distinctions in mind, the following sections summarize for which traits and in which social contexts people have more or less insight.

Generalized meta-accuracy: when do people know their reputation?

Overall, people have GMA, or insight into their reputation, for a wide range of attributes, such as the Big Five personality traits, leadership abilities, attractiveness, and well-being attributes (e.g., happiness, depression; Carlson & Kenny, 2012). In fact, the strength of GMA is about the same as that observed for self–other agreement (Vazire & Carlson, 2010), suggesting people have a fair amount of insight into their reputation. A meta-analysis revealed that GMA tends to be strongest for observable attributes, such as extraversion, and weakest for positively and negatively evaluative attributes, such as intelligence or arrogance (Carlson & Kenny, 2012). People are also aware of their reputation for having personality problems (Carlson, Vazire, & Oltmanns, 2011; Oltmanns, Gleason, Klonsky, & Turkheimer, 2005). For example, Air Force recruits were aware of the degree to which their training group described them as having personality disorder symptoms outlined by the DSM-5 (Oltmanns et al., 2005). In sum, there are differences in the strength of GMA across attributes, but overall, people seem to have some insight about their reputation.

People are generally aware of their reputation in a variety of social contexts as well, ranging from zero acquaintance contexts, where the metaperceiver never meets the individual(s) who rate him or her, to close relationships, such as family members or friends. For example, people can accurately infer how individuals who view their social media profiles will perceive their Big Five traits, with the exception of impressions of neuroticism (Stopfer, Egloff, Nestler, & Back, 2014). People are also aware of how their friends, coworkers, and family members view them on all of the Big Five traits (Malloy, Albright, Kenny, Agatstein, & Winquist, 1997). While not empirically tested, Malloy and colleagues (1997) observed descriptive differences in the strength of GMA, such that the relationship between metaperceptions and impressions was weaker among coworkers and stronger for family members than for friends.

Interestingly, in contrast to personality attributes people often have poor insight about their reputation for affective attributes, such as liking or popularity (Kenny & DePaulo, 1993). For example, one study found that roommates were aware of their reputation for all of the Big Five traits except for openness, but they did not have insight into their reputation for being liked (Levesque, 1997). Poor GMA for affective attributes may be due to the fact that these impressions tend to be idiosyncratic. Put another way, people typically agree about others' personality attributes (e.g., who is more or less outgoing), but people agree less about who is especially liked, which makes it difficult for a metaperceiver to guess his or her reputation.

With respect to metaperception bias, or the tendency to assume others see the self in more positive or negative ways than they really do, a meta-analysis based on findings from face-to-face relationships found that people are slightly, positively biased about personality attributes but slightly, negatively biased about relational attributes (e.g., liking; Carlson & Kenny, 2012). This pattern suggests that people assume the best about how others experience their personalities but are too humble with respect to how much others like them. However, at least two zero acquaintance studies suggest people can be positively biased about their popularity. One study showed that people who guessed how they would be seen based on their online profile overestimated how likeable others would rate them (Sherman et al., 2001). The same positivity bias was observed when people were asked to guess how a potential romantic partner might perceive them based on their online romantic profile (Preuss & Alicke, 2009). Thus, the degree to which people are positively or negatively biased about how people perceive them might depend on the attribute and social context.

Dyadic meta-accuracy: when do people know the unique impressions they make?

DMA indexes whether people know who sees them as higher or lower on a given attribute, or whether they know the unique impressions they make on specific people. In general, people's insight into their reputation (i.e., GMA) is stronger than DMA for personality attributes (Carlson & Kenny, 2012; Kenny & DePaulo, 1993). For example, college roommates know their reputation for most of the Big Five traits, but they are not necessarily aware of which roommate sees them as higher or lower on these attributes (Levesque, 1997). However, DMA tends to be quite strong for affective attributes (e.g., social value; Elfenbein, Eisenkraft, & Ding, 2009). For example, while college roommates cannot detect the unique impressions they make on the Big Five, they do know who sees them as more or less interesting (Levesque, 1997). This difference is likely driven by the fact that people agree substantially about personality attributes (e.g., Jan and Jon agree that Meg is outgoing), but affective perceptions are more idiosyncratic (i.e., Jan likes Meg but Jon does not).

People's ability to detect the unique impressions they make on specific people varies across social contexts in at least two ways. First, when people first meet, DMA seems to be stronger for agentic traits (e.g., extraversion, openness) than for communal traits (e.g., agreeableness), but this pattern is reversed among close acquaintances (Carlson & Kenny, 2012), suggesting DMA may change over time for some attributes. Second, DMA is substantially weaker when assessed in a single social context, such as among coworkers, than when it is assessed across social contexts, such as among hometown friends, college friends, and family members (Carlson & Furr, 2009). This pattern may be due to the fact that people make unique impressions on people they know from different social contexts (e.g., parents versus friends; Funder, Kolar, & Blackman, 1995). Put another way, people struggle to differentiate among the impressions they've made among people in the exact same context because there is little variability to detect, but in everyday life, where people interact with perceivers across contexts, people can detect the unique impressions they make on others.

How do people have insight into the impressions they make?

The process of forming metaperceptions is complex and involves integrating several sources of information. Intuitively, the main source of information should be verbal and nonverbal feedback, but several other sources of information seem to be important in metaperception, such as self-perceptions, heuristics, and stereotype information (Carlson & Kenny, 2012; Kenny, 1994; Kenny & DePaulo, 1993). The following sections review the degree to which these various sources of information facilitate or hinder insight.

Feedback

Presumably, people provide some form of feedback that metaperceivers detect and utilize. Yet, direct feedback is rarely provided, especially when it comes to evaluative attributes (Blumberg, 1972; Herbert & Vorauer, 2003; Shrauger & Schoeneman, 1979). Rather than being direct, people often hide their true feelings or tell white lies to avoid hurt feelings or conflict (DePaulo & Bell, 1996; Swann, Stein-Seroussi, & McNulty, 1992), which ultimately forces metaperceivers to rely on nonverbal feedback or other sources of information. When people do provide explicit feedback, metaperceivers often fail to detect or correctly utilize this information (Shechtman & Kenny, 1994). One explanation is that social interactions are cognitively taxing, which makes it difficult to detect cues (Lieberman & Rosenthal, 2001; Patterson, 1995). For instance, Meg might not notice Pete's positive responses toward her because she is thinking about the next thing to say or paying attention to something else. A great deal of interpersonal feedback, especially negative feedback, also takes the form of a non-occurrence, or a lack of a response, such as not receiving a compliment (Carter & Dunning, 2008). When metaperceivers do notice cues, they may fail to correctly utilize the information because they read too much into feedback (Kaplan, Santuzzi, & Ruscher, 2009). Likewise, motivational biases, such as self-enhancement motives (Sedikides, 2007) or self-verification motives (i.e., the motive to believe others share one's self-view; Kwang & Swann, 2010), may skew how people utilize feedback. Indeed, at least one study suggests that people fail to correctly utilize direct, verbal feedback, largely because they assume others see them as they see themselves (Shechtman & Kenny, 1994). In sum, perceivers do not always provide direct feedback, and when they do, metaperceivers either fail to detect the cues or somehow misinterpret them. Counter to intuition, research thus far suggests feedback is generally a poor source of information to use when inferring how others perceive the self.

Self-perceptions of behavior

Rather than using feedback to infer how others see us, people can observe their own behavior and form metaperceptions based on that information (Bem, 1967; Kenny & DePaulo, 1993). Perceivers generally base their impressions of an individual on observations of that person's behavior; thus, self-perception may be a valid source of information when the self and others observe the same information. Going further, people tend to agree on which behaviors are indicative of a given attribute (Funder & Sneed, 1993), suggesting that metaperceptions and impressions that are based on the same observations of behavior will converge.

Some evidence suggests that self-observation of behavior does facilitate insight. For example, people are more aware of how new acquaintances perceive their neurotic attributes when they are able to view themselves interacting with those individuals (i.e., by watching themselves in a videotape; Albright & Malloy, 1999). Self-perception of behavior may also explain why people know the unique impressions they make in different social contexts (Carlson & Furr, 2009). People behave differently across social contexts (Furr & Funder, 2004) and consequently make unique impressions on people who know them in different contexts (Funder et al., 1995). Thus, people may think about how they behave when in a specific context, which allows them to accurately gauge the impressions they make across social contexts.

Yet, there are a few reasons to predict that self-observation hinders insight. People are not always aware of how they behave, especially on evaluative behaviors (e.g., rudeness) or behaviors that are difficult for the self to observe (e.g., facial expressions; Gosling, Pete, Craik, & Robins, 1998; Hall, Murphy, & Schmid Mast, 2007; Vazire & Mehl, 2008). People also tend to weigh internal factors, such as intentions and motivations, more than their actual behavior when forming self-perceptions, whereas the opposite is true when people form judgments of others (Malle & Pearce, 2001; Petees & Nisbett, 1972). Thus, people may not weigh their own behavior enough when forming metaperceptions. Indeed, several lines of work show that people engage in the *illusion of transparency* (Gilovich, Savitsky, & Medvec, 1998), whereby they assume others can detect their internal experiences (e.g., motivation, intention) more than is the case. Likewise, people can become so accustomed to their own behavior that they fail to realize how distinctive it is to other people (Leising, Rehbein, & Sporberg, 2006).

Self-perceptions of personality

Rather than observing their own behavior, people might simply assume others see their personality as they do. The relationship between self- and metaperceptions is quite strong suggesting that self-perceptions are the main source of information people use to infer how others see them (Kenny, 1994; Kenny & DePaulo, 1993). Going further, people are generally motivated to assume others see them as they see themselves in order to verify their own identity and to make the world a more predictable place (Swann, 1990). Indeed, people with positive self-views assume others see them positively, whereas people with negative self-views assume others see them negatively (Campbell & Fehr, 1990; Murray, Holmes, & Griffin, 2000; Murray, Holmes, MacDonald, & Ellsworth, 1998).

Self-perceptions of personality will facilitate insight when others actually agree with one's self-views, or when self–other agreement is high. In support of this hypothesis, insight is strongest for traits that tend to show strong self–other agreement, such as observable traits (e.g., extraversion; Carlson & Kenny, 2012). In contrast, self-perception of personality hinders insight when other people do not see us as we see ourselves. Thus, self-perception is a poor source of metaperception when perceivers' perceptions are more accurate than are self-perceptions, which tends to be true for evaluative attributes (e.g., intelligence, agreeableness; Carlson & Furr, 2013; Vazire, 2010). Likewise, self-perception is a poor source when perceivers are less accurate than is the self, which tends to be true for internal traits such as neuroticism (Vazire, 2010), or when perceivers base their impressions on personal biases rather than on reality (Srivastava, Guglielmo, & Beer, 2010; Wood, Harms, & Vazire, 2010). In sum, self-perception of personality is a major source of information that facilitates insight when others actually do agree with the self, but hinders insight when others do not share one's self-views.

Heuristics and stereotypes

There are likely several heuristics people use to guess how others perceive them. One type of heuristic is *assumed reciprocity* where a metaperceiver assumes that a perceiver will perceive him or her in the same way as the metaperceiver sees the perceiver. For example, Meg sees Pete as very intelligent and assumes that Pete sees her in the same way. There is some evidence that this heuristic facilitates insight for attributes that are reciprocal, such as liking. For example, coworkers tend to assume that the people they value also value them, an assumption that facilitates DMA (Elfenbein et al., 2009). However, assumed reciprocity is not especially effective for personality attributes (Kenny & DePaulo, 1993).

Another heuristic people use implicitly or explicitly is their beliefs about what the typical person is like, or *normativeness*. Normative information is a major component of initial personality judgments when people have little information to base their impressions on and can facilitate accuracy (Ames, 2004). For example, if Pete described Meg's personality before meeting her, his perceptions would be somewhat valid because he assumes she is similar to the typical person (e.g., more kind than cruel). The same normative information can be used when forming metaperceptions. For example, if Meg was asked to guess how Pete perceives her before ever meeting him, her guess about how Pete sees her would be somewhat accurate because Meg assumes Pete sees her in typical ways (e.g., more kind than cruel). Indeed, in first impression contexts, removing normative information substantially reduces meta-accuracy, suggesting that normative information is a valid source of information people use to infer how others see them (Carlson, Furr, & Vazire, 2010). The role of normative information in metaperception over the course of acquaintanceship is unclear, but similar to other perceptions, it might be used less over time (Biesanz, West, & Millevoi, 2007).

When thinking about how others perceive the self, people sometimes consider attributes that are associated with their group memberships as well as the group membership of the perceiver (e.g., ethnicity, social class; Frey & Tropp, 2006). For example, when meeting Pete for the first time, Meg might assume that Pete sees her in stereotypically feminine ways (e.g., kind, nurturing) even if she does not see herself in this way. Beliefs about how others view the self based on their social group membership are called *meta-stereotypes* (Vorauer, Hunter, Main, & Roy, 2000; Vorauer & Kumhyr, 2001; Vorauer, Main, & O'Connel, 1998). Most evidence suggests meta-stereotypes hurt rather than help insight. For example, research that explored meta-accuracy for personality attributes between heterosexual and homosexual men found that, while heterosexual men demonstrated normal levels of insight, homosexual men showed negative meta-accuracy, an effect that is rarely observed (Miller & Malloy, 2003). This effect was due to the fact that homosexual men erroneously assumed they would be seen in negative, stereotypical ways.

Summary

There are several available sources of information people can use to infer how others see them, but our review suggests that no one source seems to provide a clear path to insight. Instead, people likely use a combination of several sources of information. For example, recent work suggests that people are able to detect how others see them differently from how they see themselves, an ability called meta-insight (Carlson, Vazire, & Furr, 2011). Specifically, controlling for self-perceptions of personality (Carlson et al., 2011; Gallrein, Carlson, Holstein, & Leising, 2013; Oltmanns et al., 2005) or self-perceptions of behavior (Carlson et al., 2011), metaperceptions predict impressions, suggesting self-views are not the only source of information people use to achieve insight. The same findings have emerged for various heuristics, such as normativeness and assumed reciprocity; specifically, controlling for normative information (Carlson & Furr, 2013; Carlson et al., 2010) or assumed reciprocity (Elfenbein et al., 2009), people are aware of how others see them. In sum, people likely use several sources of information to accurately infer how they are seen.

Going one step further, research has identified individual differences and situational factors that affect insight, which suggests the sources of information that foster (or hinder) insight might vary across people and situations. For example, relative to people with fewer, people with more personality disorder symptoms listed in the DSM-5 tend to be less accurate about how a close acquaintance, known for approximately 30 years (e.g., romantic partner, family member), sees them on the Big Five traits, and they also tend to overestimate the negativity of the impressions they make on this person (Carlson & Oltmanns, 2015). Thus, unlike the typical person who has some accuracy and is somewhat positively biased about the Big Five traits (Carlson & Kenny, 2012), people with more personality problems are less accurate and more negatively biased than are people with fewer problems. (Chapter 9 discusses psychopathology in relation accurately judging others' own characteristics.) Supplementary analyses suggest these individuals are more likely to base their metaperceptions on information other than their self-perceptions, suggesting that their poor insight is driven by misinterpreting other cues (e.g., feedback).

For other forms of psychopathology, self-perception seems to play a key role in poor insight. For instance, people with low self-esteem and people higher in social anxiety tend to assume they are seen more negatively than they really are seen by new acquaintances and by a romantic partner, an effect largely driven by negative self-views (Christensen, Stein, & Means-Christensen, 2003; Kashdan & Savostyanova, 2011; Murray et al., 2000; Murray, Holmes, Griffin, Bellavia, & Rose, 2001; Murray et al., 1998). Likewise, people higher in subclinical narcissism, which is defined by exceptionally high self-esteem, underestimate how negatively new acquaintances see them (Lukowitsky & Pincus, 2013).

In addition to stable, individual differences, there are several situational factors that affect insight and the sources that may foster or hinder insight. One factor is the relative status of the metaperceiver. Low-status individuals may have more insight into the impressions they make because these individuals pay closer attention to higher status individuals, with the aim of better predicting outcomes that these individuals control (Fiske, 1993). Yet, high-status individuals tend to be more expressive (i.e., provide more feedback), which can make it easier for low-status individuals to detect interpersonal cues (Snodgrass, 1992; Snodgrass, Hecht, & Ploutz-Snyder, 1998). Taken together, there may be an asymmetry in insight between people of high and low status, such that people of lower status are more aware of the impressions they make than are higher status individuals due to differences in the attention to and availability of feedback.

Outcome-dependent situations, or situations where someone else's opinion of the metaperceiver has a direct impact on his or her life (e.g., job interviews) also have a major impact on insight by affecting which sources of information are more or less salient. In most situations, people tend to weigh their internal experiences more than their external behavior when describing themselves (Andersen, Glassman, & Gold, 1998), but in outcome-dependent situations, internal experiences (e.g., goals, motives, and self-views) become even more salient, which leads people to believe that their inner experiences are also salient to others (Garcia, 2002; Vorauer, 2006). For example, one study found that negotiators with less power in the negotiation assumed that their inner experiences were more transparent than did the higher-powered negotiator. Put another way, outcome-dependent situations can lead to people assuming that a perceiver knows what they are thinking and feeling and may lead them to erroneously conclude that others perceive them as they see themselves. In addition to overestimating one's transparency, outcome-dependent situations can also lead a metaperceiver to read too much into others' verbal and nonverbal feedback (Kaplan et al., 2009). In sum, outcome-dependent situations tend to hinder insight, because they lead to erroneously assuming others know more about the self than is really the case or to overthinking the meaning of social feedback.

Future directions

We know a great deal about if and when people have insight into the impressions they make on others, but there are at least five critical gaps in the literature that we hope future research will explore. First, most work has focused on the accuracy of metaperceptions rather than on mean-level differences. For example, in a meta-analysis of 26 studies, only eight studies included bias indices (Carlson & Kenny, 2012). The underlying processes that lead to accuracy and bias are distinct, which means that the factors that affect accuracy may not be the same factors that affect bias. Research from other domains of accuracy (e.g., self–other agreement) suggests that people are not only motivated to be *accurate* in their social judgments to make the world a predictable and navigable place, but also motivated to be *biased* to feel more secure in their relationships (Gagne & Lydon, 2004). In the context of metaperception, people might also be motivated to know who sees them as more or less intelligent, for example, to make better decisions (e.g., requesting professional references) while also being motivated to maintain self-esteem by assuming others generally see them as being more intelligent than they really do. Learning more about which traits, which individuals, and in which contexts metaperceptions are biased may shed more light on the process and consequences of metaperceptions. To accomplish this goal, future research might employ the Truth and Bias model (West & Kenny, 2011), which is essentially a centering model that can index accuracy and mean-level bias in a single model. This model can also identify factors that influence both accuracy and mean-level bias, such as an individual's self-perception (e.g., do self-perceptions facilitate or hinder insight?), perceptions of relationship quality (e.g., does a higher quality relationship affect insight?), or individual differences in personality (e.g., do some people have more insight?). For example, the Truth and Bias model was used to explore if people with personality disorder symptoms have insight into how a close acquaintance perceived them and whether self-perceptions foster or hinder their insight (Carlson & Oltmanns, 2015). Hopefully, future research will apply this model to many other attributes, social contexts, and individual differences.

Second, insight may vary in important ways across cultures, but few studies have directly compared the process of metaperception or the level of insight across cultures. Studies conducted in the United States and Mexico found that GMA is fairly strong and consistent across social contexts (e.g., family, friends, coworkers), but research in China suggests that meta-accuracy among Chinese participants may be weaker and more mixed across contexts (Malloy, Albright, Diaz-Loving, Dong, & Lee, 2004; Malloy et al., 1997). Future work is needed to better identify and explain cultural differences in insight. For example, differences might be driven by variability in the availability of feedback or the degree to which feedback is interpreted in self-enhancing ways. Theoretically, this line of work may also reveal which bright spots and blind spots in metaperception are more or less universal across cultures.

Third, future work is needed to better understand if and when people know how others experience them across modes of communication, such as writing or in social media. Technology is changing the way people communicate with one another in fundamental ways, and this shift might have a major effect on people's ability to understand how others perceive them. Indeed, communication in workplace settings tends to occur over e-mail rather than in face-to-face meetings, and romantic relationships are often initiated online rather than in person. A few studies have explored whether people know how others perceive their online personas

(Stopfer et al., 2014), but much more work is needed to better understand social perception among different types of social media (e.g., Twitter, Facebook, LinkedIn), given that the functions and the social norms for each platform may vary in ways that affect metaperception.

Fourth, people often want to learn how to improve the accuracy of their social judgments (see Chapter 12), but it is unclear if seeing social reality is a healthy goal. On one hand, metaperceptions play a pivotal role in people's ability to navigate complex social environments, suggesting that insight is best. Indeed, people use metaperceptions to gauge the success of their impression management goals (e.g., making good first impressions; Schlenker & Weigold, 1992) as well as their social value (Elfenbein et al., 2009; Leary, 2005). Going further, misunderstanding how others perceive us likely leads to a host of communication problems and hurt feelings (Cameron & Vorauer, 2008). On the other hand, there may be times where it is best to blindly assume others see us as we see ourselves or that others see the best in us. Interestingly, assuming a close other sees us more positively than we see ourselves or more positively than they really do seems to facilitate relationship quality (Boyes & Fletcher, 2007; Kenny & Acitelli, 2001; Lemay & Dudley, 2011). In sum, the evidence for the adaptiveness of insight is mixed, and future work is needed to disentangle if and when insight is adaptive.

Fifth, if insight is adaptive, future work might focus more on how to improve it. To date, the most valid source of information seems to be self-perceptions of one's behavior, suggesting that self-knowledge of behavior improves insight. However, there may be many important moderators that dictate which sources of information are best when the goal is to understand how others experience us. For example, focusing on one's behavior in outcome-dependent situations (e.g., job interviews) or situations where people are particularly concerned with being liked (e.g., romantic dates) may backfire, whereas this type of awareness in casual interactions may improve insight. Thus, future work is needed to better understand the contextual factors that affect the process of insight as well as which sources of information lead to accuracy.

Conclusion

Do people know how others see them? In a nutshell, people are aware of their reputation for most personality attributes across a host of social contexts but are less aware of the unique personality impressions they make on specific individuals. Yet, people are very attuned to who especially likes or values them. There are many sources of information that can lead to accurate metaperceptions, but thus far, research suggests that no one source provides a clear path to insight. Indeed, individual differences and situational factors seem to play a large role in which sources of information facilitate or hinder insight. Hopefully, future research will explore the fundamental process of metaperception to better understand how people come to understand their effect on other people.

References

Albright, L., & Malloy, T. E. (1999). Self-observation of social behavior in metaperception. *Journal of Personality and Social Psychology*, 77, 726–734.

Ames, D. R. (2004). Strategies for social inference: A similarity contingency model of projection and stereotyping in attribute prevalence estimates. *Journal of Personality and Social Psychology*, 87, 573–585.

Andersen, S. M. Glassman, N. S. & Gold, D. A. (1998). Mental representations of the self, significant others, and nonsignificant others: Structure and processing of private and public aspects. *Journal of Personality and Social Psychology*, 75, 845–861.

Bem, D. J. (1967). Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review*, 74, 183–200.

Biesanz, J. C., West, S. G., & Millevoi, A. (2007). What do you learn about someone over time? The relationship between length of acquaintance and consensus and self-other agreement in judgments of personality. *Journal of Personality and Social Psychology*, 92, 119–135.

Blumberg, H. H. (1972). Communication of interpersonal evaluations. *Journal of Personality and Social Psychology*, 23, 157–162.

Boyes, A. D., & Fletcher, G. J. O. (2007). Metaperceptions of bias in intimate relationships. *Journal of Personality and Social Psychology*, 92, 286–306.

Cameron, J. J., & Vorauer, J. D. (2008). Feeling transparent: On metaperceptions and miscommunications. *Social and Personality Psychology Compass*, 2, 1093–1108.

Campbell, J. D., & Fehr, B. (1990). Self-esteem and perceptions of conveyed impressions: Is negative affectivity associated with greater realism? *Journal of Personality and Social Psychology*, 58, 122–133.

Carlson, E. N., & Furr, R. M. (2009). Differential meta-accuracy: People understand the different impressions they make. *Psychological Science*, 20, 1033–1039.

Carlson, E. N., & Furr, R. M. (2013). Resolution for meta-accuracy: Should people trust their beliefs about how others see them? *Social Psychological and Personality Science*, 4, 419–426.

Carlson, E. N., Furr, R. M., & Vazire, S. (2010). Do we know the first impressions we make? Evidence for idiographic meta-accuracy and calibration of first impressions. *Social Psychological and Personality Science*, 1, 94–98.

Carlson, E. N., & Kenny, D. A. (2012). Meta-accuracy: Do we know how others see us? In S. Vazire & T. D. Wilson (Eds.), *Handbook of self-knowledge* (pp. 242–257). New York: Guilford Press.

Carlson, E. N., & Oltmanns, T. F. (2015). The role of metaperception in personality disorders: Do people with personality problems know how others experience their personality? *Journal of Personality Disorders*, 29, 449–467.

Carlson, E. N., Vazire, S., & Furr, R. M. (2011). Meta-insight: Do people really know how others see them? *Journal of Personality and Social Psychology*, 101, 831–846.

Carlson, E. N., Vazire, S., & Oltmanns, T. F. (2011). You probably think this paper's about you: Narcissists' perceptions of their personality and reputation. *Journal of Personality and Social Psychology*, 101, 185–201.

Carter, T. J., & Dunning, D. (2008). Faulty self-assessment: Why evaluating one's own competence is an intrinsically difficult task. *Social & Personality Psychology Compass*, 2, 346–360.

Christensen, P. N., Stein, M. B., & Means-Christensen, A. (2003). Social anxiety and interpersonal perception: A social relations model analysis. *Behaviour Research and Therapy*, 41, 1355–1371.

DePaulo, B. M., & Bell, K. L. (1996). Truth and investment: Lies are told to those who care. *Journal of Personality and Social Psychology*, 71, 703–716.

Elfenbein, H. A., Eisenkraft, N., & Ding, W. W. (2009). Do we know who values us? Dyadic meta-accuracy in the perception of professional relationships. *Psychological Science*, 20, 1081–1083.

Felson, R. B., & Reed, M. D. (1986). Reference groups and self-appraisals of academic ability and performance. *Social Psychology Quarterly*, 49, 103–109.

Fiske, S. T. (1993). Controlling other people: The impact of power on stereotyping. *American Psychologist*, 48, 621–628.

Frey, F. E., & Tropp, L. R. (2006). Being seen as individuals versus as group members: Extending research on metaperception to intergroup contexts. *Personality and Social Psychology Review*, 10, 265–280.

Funder, D. C., Kolar, D. C., & Blackman, M. C. (1995). Agreement among judges of personality: Interpersonal relations, similarity, and acquaintanceship. *Journal of Personality and Social Psychology*, 69, 656–672.

Funder, D. C., & Sneed, C. D. (1993). Behavioral manifestations of personality: An ecological approach to judgmental accuracy. *Journal of Personality and Social Psychology*, 64, 479–490.

Furr, R. M., & Funder, D. C. (2004). Situational similarity and behavioral consistency: Subjective, objective, variable-centered, and person-centered approaches. *Journal of Research in Personality*, 38, 421–447.

Gagne, F. M., & Lydon, J. E. (2004). Bias and accuracy in close relationships: An integrative review. *Personality and Social Psychology Review*, 8, 322–338.

Gallrein, M. B., Carlson, E. N., Holstein, M., & Leising, D. (2013). You spy with your little eye: People are "blind" to some of the ways in which they are consensually seen by others. *Journal of Research in Personality*, 47, 464–471.

Garcia, S. M. (2002). Power and the illusion of transparency in negotiations. *Journal of Business and Psychology*, 17, 133–144.

Gilovich, T., Savitsky, K., & Medvec, V. H. (1998). The illusion of transparency: Biased assessments of others' ability to read one's emotional states. *Journal of Personality and Social Psychology*, 75, 332–346.

Gosling, S. D., Pete, O. P., Craik, K. H., & Robins, R. W. (1998). Do people know how they behave? Self-reported act frequencies compared with on-line codings by observers. *Journal of Personality and Social Psychology*, 74, 1337–1349.

Hall, J. A., Murphy, N. A., & Schmid Mast, M. (2007). Nonverbal self-accuracy in interpersonal interaction. *Personality and Social Psychology Bulletin*, 33, 1675–1685.

Herbert, B. G., & Vorauer, J. D. (2003). Seeing through the screen: Is evaluative feedback communicated more effectively in face-to-face or computer mediated exchanges? *Computers in Human Behavior*, 19, 25–38.

Kaplan, S. A., Santuzzi, A. M., & Ruscher, J. B. (2009). Elaborative metaperceptions in outcome-dependent situations: The diluted relationship between default self-perceptions and metaperceptions. *Social Cognition*, 27, 601–614.

Kashdan, T. B., & Savostyanova, A. A. (2011). Capturing the biases of socially anxious people by addressing partner effects and situational parameters. *Behavior Therapy*, 42, 211–223.

Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis*. New York: Guilford Press.

Kenny, D. A., & Acitelli, L. K. (2001). Accuracy and bias in the perception of the partner in a close relationship. *Journal of Personality and Social Psychology*, 80, 439–448.

Kenny, D. A., & DePaulo, B. M. (1993). Do people know how others view them? An empirical and theoretical account. *Psychological Bulletin*, 114, 145–161.

Kwang, T., & Swann, W. B. (2010). Do people embrace praise even when they feel unworthy? A review of critical tests of self-enhancement versus self-verification. *Personality and Social Psychology Review*, 14, 263–280.

Leary, M. R. (2005). Sociometer theory and the pursuit of relational value: Getting to the root of self-esteem. *European Review of Social Psychology*, 16, 75–111.

Leising, D., Rehbein, D., & Sporberg, D. (2006). Does a fish see the water in which it swims? A study of the ability to correctly judge one's own interpersonal behavior. *Journal of Social and Clinical Psychology*, 25, 963–974.

Lemay, E. P., & Dudley, K. L. (2009). Implications of reflected appraisals of interpersonal insecurity for suspicion and power. *Personality and Social Psychology Bulletin*, 35, 1672–1686.

Lemay, E. P., & Dudley, K. L. (2011). Caution: fragile! Regulating the interpersonal security of chronically insecure partners. *Journal of Personality and Social Psychology*, 100, 681–702.

Levesque, M. J. (1997). Meta-accuracy among acquainted individuals: A social relations analysis of interpersonal perception and metaperception. *Journal of Personality and Social Psychology*, 72, 66–74.

Lieberman, M. D., & Rosenthal, R. (2001). Why introverts can't always tell who likes them: Multitasking and nonverbal decoding. *Journal of Personality and Social Psychology*, 80, 294–310.

Lukowitsky, M. R., & Pincus, A. L. (2013). Interpersonal perception of pathological narcissism: A social relations analysis. *Journal of Personality Assessment*, 95, 261–273.

Malle, B. F., & Pearce, G. E. (2001). Attention to behavioral events during interaction: Two actor–observer gaps and three attempts to close them. *Journal of Personality and Social Psychology*, 81, 278–294.

Malloy, T. E., Albright, L., Diaz-Loving, R., Dong, Q., & Lee, Y. T. (2004). Agreement in personality judgments within and between nonoverlapping social groups in collectivist cultures. *Personality and Social Psychology Bulletin*, 30, 106–117.

Malloy, T. E., Albright, L., Kenny, D. A, Agatstein, F., & Winquist, L. (1997). Interpersonal perception and metaperception in nonoverlapping social groups. *Journal of Personality and Social Psychology*, 72, 390–398.

Malloy, T. E., & Janowski, C. (1992). Perceptions and metaperceptions of leadership: Components, accuracy, and dispositional correlates. *Personality and Social Psychology Bulletin*, 18, 700–708.

Miller, S., & Malloy, T. E. (2003). Interpersonal behavior, perception, and affect in status-discrepant dyads: Social interaction of gay and heterosexual men. *Psychology of Men & Masculinity*, 4, 121–135.

Murray, S. L., Holmes, J. G., & Griffin, D. W. (2000). Self-esteem and the quest for felt security: How perceived regard regulates attachment processes. *Journal of Personality and Social Psychology*, 78, 478–498.

Murray, S. L., Holmes, J. G., Griffin, D. W., Bellavia, G., & Rose, P. (2001). The mismeasure of love: How self-doubt contaminates relationship beliefs. *Personality and Social Psychology Bulletin*, 27, 423–436.

Murray, S. L., Holmes, J. G., MacDonald, G., & Ellsworth, P. C. (1998). Through the looking glass darkly? When self-doubts turn into relationship insecurities. *Journal of Personality and Social Psychology*, 75, 1459–1480.

Oltmanns, T. F., Gleason, M. E. J., Klonsky, E. D., & Turkheimer, E. (2005). Metaperception for pathological personality traits: Do we know when others think that we are difficult? *Consciousness and Cognition: An International Journal*, 14, 739–751.

Patterson, M. L. (1995). Invited article: A parallel process model of nonverbal communication. *Journal of Nonverbal Behavior*, 19, 3–29.

Petees, E. E., & Nisbett, R. E. (1972). The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Petees, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 79–94). Hillsdale, NJ: Lawrence Erlbaum Associates.

Preuss, G. S., & Alicke, M. D. (2009). Everybody loves me: Self-evaluations and metaperceptions of dating popularity. *Personality and Social Psychology Bulletin*, 35, 937–950.

Schlenker, B. R., & Weigold, M. F. (1992). Interpersonal processes involving impression regulation and management. *Annual Review of Psychology*, 43, 133–168.

Sedikides, C. (2007). Self-enhancement and self-protection: Powerful, pancultural, and functional. *Hellenic Journal of Psychology*, 4, 1–13.

Shechtman, Z., & Kenny, D. A. (1994). Metaperception accuracy: An Israeli study. *Basic and Applied Social Psychology*, 15, 451–465.

Sherman, R. C., End, C., Kraan, E., Cole, A., Campbell, J., Klausner, J., & Birchmeier, Z. (2001). Metaperception in cyberspace. *CyberPsychology & Behavior*, 4, 123–129.

Shrauger, J. S., & Schoeneman, T. J. (1979). Symbolic interactionist view of self-concept: Through the looking glass darkly. *Psychological Bulletin*, 86, 549–573.
Snodgrass, S. E. (1992). Further effects of role versus gender on interpersonal sensitivity. *Journal of Personality and Social Psychology*, 62, 154–158.
Snodgrass, S. E., Hecht, M. A., & Ploutz-Snyder, R. (1998). Interpersonal sensitivity: Expressivity or perceptivity? *Journal of Personality and Social Psychology*, 74, 238–249.
Srivastava, S., Guglielmo, S., & Beer, J. S. (2010). Perceiving others' personalities: Examining the dimensionality, assumed similarity to the self, and stability of perceiver effects. *Journal of Personality and Social Psychology*, 98, 520–534.
Stopfer, J. M., Egloff, B., Nestler, S., & Back, M. D. (2014). Personality expression and impression formation in online social networks: An integrative approach to understanding the processes of accuracy, impression management and meta-accuracy. *European Journal of Personality*, 28, 73–94.
Swann, W. B. (1990). To be adored or to be known: The interplay of self-enhancement and self-verification. In R. M. Sorrentino & E. T. Higgins (Eds.), *Motivation and cognition* (Vol. 2, pp. 408–448). New York: Guilford Press.
Swann, W. B., Stein-Seroussi, A., & McNulty, S. E. (1992). Outcasts in a white-lie society: The enigmatic worlds of people with negative self-conceptions. *Journal of Personality and Social Psychology*, 62, 618–624.
Vazire, S. (2010). Who knows what about a person? The self–other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology*, 98, 281–300.
Vazire, S., & Carlson, E. N. (2010). Self-knowledge of personality: Do people know themselves? *Social and Personality Psychology Compass*, 4, 605–620.
Vazire, S., & Mehl, M. R. (2008). Knowing me, knowing you: The accuracy and unique predictive validity of self-ratings and other-ratings of daily behavior. *Journal of Personality and Social Psychology*, 95, 1202–1216.
Vorauer, J. D. (2006). An information search model of evaluative concerns in intergroup interaction. *Psychological Review*, 113, 862–886.
Vorauer, J. D., Hunter, A. J., Main, K. J., & Roy, S. A. (2000). Meta-stereotype activation: Evidence from indirect measures for specific evaluative concerns experienced by members of dominant groups in intergroup interaction. *Journal of Personality and Social Psychology*, 78, 690–707.
Vorauer, J. D., & Kumhyr, S. M. (2001). Is this about you or me? Self-versus other-directed judgments and feelings in response to intergroup interaction. *Personality and Social Psychology Bulletin*, 27, 706–719.
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.
West, T. V., & Kenny, D. A. (2011). The truth and bias model of judgment. *Psychological Review*, 118, 357–378.
Wood, D., Harms, P., & Vazire, S. (2010). Perceiver effects as projective tests: What your perceptions of others say about you. *Journal of Personality and Social Psychology*, 99, 174–190.

Part II

Correlates of interpersonal accuracy

9 Accuracy in perceiving facial expressions of emotion in psychopathology

Pip Griffiths and Chris Ashwin

Abstract

The accurate recognition of facial expressions is important for social interaction and interpersonal relationships. Many psychiatric disorders are characterized by social difficulties, including problems in recognizing the emotions of others. The disorders of autism, schizophrenia, depression, social anxiety, and borderline personality disorder all involve difficulties in social cognition, which include deficits in recognizing emotions. However, the nature of these difficulties within and between psychiatric disorders is not well understood. One problem is that differences in the methodologies and samples utilized within emotion research often make it difficult to make comparisons across studies and disorders. Despite these differences, some commonalities in deficits for perceiving emotional expressions have been noted within psychopathology, which include differences in processing of emotions involving negative valence. This may suggest dysfunction of a primary emotion system centered on the amygdala in the brain, although the exact nature of this dysfunction may vary between some disorders.

Humans are very social beings, so it is not surprising that we spend a lot of time looking at each other. Much of the visual information we gain about others comes from nonverbal communication, such as facial expressions of emotion (Argyle, 1972). Perceiving the facial expressions of others helps us to identify and understand their current emotional state, in order to respond to them in an appropriate way. Being able to accurately read emotional expressions is important for successful social functioning and interpersonal relationships. Reduced accuracy in reading the facial expressions of others would cause difficulties in the social world, and difficulties in this ability are associated with psychopathology. This chapter aims to review findings about the accuracy in recognizing facial displays of expression across various psychiatric disorders, in order to examine any commonalities between them that might be related to primary emotion systems.

Darwin proposed that human emotional expressions developed within more general emotion mechanisms that served to protect humans and prepare them for action (Darwin, 1872). For example, Darwin proposed that expressions of anger specifically involved a muscle pattern in the face involving a furrowed brow, flared nostrils, and a mouth showing the teeth. He believed this emotional expression evolved from the aggressive displays by animals showing their fangs, which serves to provide a signal of threat that encourages avoidance in others. Our human ancestors who could more accurately recognize emotional expressions would have benefitted from this in better understanding the behavior of others in terms of mental states and intentions, which likely facilitated the ability to pass their genes on to offspring, including a biological and brain basis for the mechanisms underlying this ability.

The universality of emotion expressions was shown through work by Paul Ekman and colleagues who studied the identification of facial expressions in a pre-literate society (Ekman & Friesen, 1971). The researchers showed participant photographs of American actors posing various emotional expressions and had them select the photograph of an emotional expression that best corresponded with it, and found high accuracy in matching the correct photograph with the corresponding emotion. From this work, the expressions of happiness, sadness, anger, disgust, fear, and surprise have become known as the "basic" emotions constant across the human race (Ekman & Friesen, 1975). Ekman and Friesen's landmark research stands as evidence for the innate ability to recognize emotional expressions, which has been used to support Darwin's notion that emotion expression recognition is an inborn and evolved ability within the human species. Failures in the accurate recognition of facial expressions of emotion would lead to difficulties in understanding the mental and emotional states of others and misattributions about the intentions underlying people's behavior.

There have been various ideas regarding the functions of emotional expressions as a communication tool and how they might be different in psychopathology (Flack & Laird, 1998; Kring & Bachorowski, 1999). (Chapter 2 discusses emotion recognition in typical populations.) Blair (2003) suggested that different patterns of social cognition ability seen in disorders such as autism, psychopathy, and sociopathy might relate to specific deficits in accurately recognizing different types of emotions, and the pattern of deficits might help to understand and differentiate between various disorders. He further proposed that the difficulties in accurately recognizing emotional expressions could be due to impairments in a primary emotion system in the brain.

The brain basis of emotion expression recognition in psychopathology

One such primary emotion system is centered on the amygdala region of the brain and is important for emotion processing (LeDoux, 1996), including the accurate identification of emotional expressions. The idea of a primary system for reading emotional expressions is supported by research showing that damage to the amygdala in humans produces deficits in the ability to accurately recognize the emotional expressions of others (Adolphs, Tranel, Damasio, & Damasio, 1994). Initial research suggested that the deficits were most evident for the basic emotion of fear, and anger to a lesser extent (Adolphs, Tranel, Damasio, & Damasio, 1995; Calder, 1996). However, further studies with amygdala patients revealed evidence for difficulties in perceiving other basic emotions expressing a negative valence, including deficits for disgust and sadness (Adolphs et al., 1999; Anderson & Phelps, 2000). To help clarify the role of the amygdala in the recognition of basic emotions from facial expressions, the data from different studies of amygdala-damaged patients were combined together in one analysis to provide a more powerful examination of the data (Adolphs et al., 1999). The findings with a larger group of nine patients with amygdala damage revealed that decreased accuracy was most evident for fear expressions, but that most patients also showed impairments for several negative emotions in addition to fear. Furthermore, no patient had impaired recognition for happy expressions, showing difficulties in accurately recognizing emotional expressions of negative valence after amygdala damage. These findings are thought to reflect the role of the amygdala in perceiving signals of threat and danger from facial expressions.

Methodology in emotion expression research

The explosion of emotion research in recent decades has spawned the development of various methods and stimuli for investigating the recognition of facial expressions. Static photographs of people displaying various different emotional expressions are often used, which are generally taken from standardized sets of emotional faces developed for research purposes. More recent research has utilized videos involving dynamic expressions of emotion, which often start with actors posing a neutral expression followed by the facial display of an emotional expression. Both the static and dynamic displays of emotion may also involve different intensities of the expressions, with lower intensities involving smaller muscle patterns in the face and higher intensities having more extreme movements of the relevant facial muscles. In addition to images of real people, sometimes schematic images or cartoon drawings are used to represent the expressions. Morphed images of emotions may be created by computers to mix emotional expressions together to varying degrees, which produce series of facial expression examples that blend from one emotional category into another.

Various paradigms have also been developed for assessing the accuracy of reading emotional expressions, with many categorized as being either labeling or match-to-sample tasks. Labeling tasks generally require participants to view a facial expression and choose the most appropriate emotional label out of a number of choices to best describe a displayed expression. In match-to-sample tasks, participants are required to observe a target image or video displaying an expression and then choose a matching image out of a set of choices. Threshold tasks represent a further type of paradigm, where participants view a dynamic facial expression or series of expressions and indicate the point in time a face starts to display an emotional valence. They often additionally label the expression after responding. Other aspects of the research methods can be varied, such as the time stimuli are presented to the participants and the amount of information about the emotion displayed.

Recognition of emotional expressions in autism

Autism spectrum conditions (ASCs) are characterized by difficulties in social reciprocity and communication alongside restricted interests and repetitive behavior (APA, 2013). The diagnostic criteria for ASC include reduced recognition of emotional expressions in others and impairments in understanding and using emotion. The emotion deficits have been prominent since the first reports about children with ASC noted about the difficulties in emotionally relating to other people (Kanner, 1943). The main difficulty was their inability to relate with other people from early in life, and instead show a preference to relate to inanimate and more predictable items. The empathizing–systemizing theory (Baron-Cohen, 2010) is a prominent psychological theory of ASC proposing that people with ASC have reduced interest in the social world, alongside an increased interest to understand how nonsocial systems work. The innate human drive toward the social world appears to be reduced in ASC and produces difficulties in social cognition (Chevallier, Kohls, Troiani, Brodkin, & Schultz, 2012). A number of reviews and meta-analyses have reported the presence of reduced accuracy in recognizing emotional expressions from the face in ASC, primarily for emotions with a negative valence (Gaigg, 2012; Harms, Martin, & Wallace, 2010; Uljarevic & Hamilton, 2013).

While some of the findings from early studies supported the idea of a deficit in accurately recognizing emotional expressions in ASC (Celani, Battacchui, & Arcidiacono, 1999; Hobson, 1986), other early studies failed to replicate results of reduced accuracy (Buitelaar, Van Der Wees, Swaab-Barneveld, & Van Der Gaag, 1999; Loveland et al., 1997). The mixed results in earlier research assessing emotion expression recognition in ASC are thought to be due to differences in methodologies, stimuli, and in the nature of the samples recruited (Gaigg, 2012). Hobson (1986) tested emotion recognition in children with and without ASC by testing their ability to match facial expressions of emotion with videotapes of a person expressing different emotional states. He found those with ASC had deficits compared to controls for accurately matching the facial expression with the correct emotion video. On the other hand, studies using match-to-sample paradigms often reported no group differences in emotion recognition (Buitelaar et al., 1999). However, the intact accuracy for recognizing emotional expressions by those with ASC in these tasks may arise from the use of low-level perceptual strategies, such as an individual feature or low-level visual characteristic of the images, rather than representing emotion recognition abilities that rely on more holistic processing of faces (Celani et al., 1999). There is much evidence for enhanced local processing of features in ASC, which involves perceiving the parts of an image at the expense of understanding the whole gestalt or meaning of the image (Shah & Frith, 1983). Therefore, those with ASC may be using their general visual–perceptual strengths in match-to-sample tasks in order to match local details or features within the images of facial expressions, rather than matching based on the emotional expressions and meaning in a holistic manner that healthy controls typically utilize.

When the ability to use lower-level perceptual strategies during emotion matching tasks is restricted in ASC, then deficits for accurately recognizing emotional expressions become more evident. For example, Celani et al. (1999) investigated emotion recognition accuracy in children using a match-to-sample procedure where the test image was presented for a limited amount of time (750 ms) before participants saw the response images. The children with ASC had reduced accuracy for identifying the correct emotional expression compared to control children, with the effect unlikely to be related to short-term memory deficits based on findings reported in other research. When emotion recognition tasks require more global processing, people with ASC generally show deficits compared to controls. It is thought they are using their strengths in nonsocial processing, such as enhanced local processing of features, at the expense of processing the entire emotional expression and its meaning.

A number of studies have also reported evidence for deficits in recognizing emotional expressions with a negative valence in ASC. For example, children with ASC are less able to accurately point out examples of specific negative emotional expressions than controls (Bormann-Kischkel, Vilsmeier, & Baude, 1995), and they are poorer at identifying the emotional expressions of sadness and anger within vignettes compared to the ones showing happiness (Feldman, McGee, Mann, & Strain, 1993). A study by Howard et al. (2000) found that adults with ASC had deficits accurately labeling fear expressions compared to matched controls, but no group differences were seen in the identification of the other five basic emotions. They further reported the deficits in recognizing fear were linked to abnormalities in volume of the amygdala in the ASC group, which is the primary emotion center in the brain. Neuroimaging studies have reported reduced amygdala activation in ASC compared to controls during the perception of facial expressions of emotion (Ashwin, Baron-Cohen, Wheelwright, O'Riordan, & Bullmore, 2007; Baron-Cohen et al., 2000).

While research has shown evidence for a deficit in accurately recognizing emotional expressions with a negative valence, many of the studies have included a small number of trials, stimuli, and participants, and the field lacks replication studies with consistent methodology and stimuli. Ashwin, Chapman, Colle, & Baron-Cohen (2006) addressed some of these issues across two studies of basic emotion recognition involving separate samples of ASC and control groups, but using the same general methodology and larger numbers of trials and stimuli than many previous studies. They tested adults with and without ASC in their ability to accurately label photographs of basic emotional expressions, and found that adults with ASC had reduced accuracy for the negative basic emotional expressions of sadness, fear, anger, and disgust. These results were found alongside no group difference for the expressions of happiness, surprise, and neutral, which were labeled as non-negative expressions. These general results were found across both studies which involved different participants. A similar pattern of deficits for negative emotional expressions in ASC has also been shown using subtle emotional expressions (Law Smith, Montagne, Perrett, Gill, & Gallagher, 2010), where emotional expressions are more difficult to recognize because the muscular pattern in the face expressing the emotion is less extreme.

In conclusion, research has shown a deficit for accurately perceiving the emotional expressions of others in ASC, which is particularly evident for emotions with a negative valence. These findings are consistent with theories about social cognition difficulties in autism and ideas linking the reduced accuracy in recognizing emotional expressions to hypo-activity in the primary emotion center in the brain, the amygdala.

Recognition of emotional expressions in schizophrenia

Schizophrenia is a disorder with severe abnormal behavior including hallucinations, delusions, and flattened affect (APA, 2013). The negative symptoms in schizophrenia represent a reduced or absent functioning in something that is typically present, such as a loss of motivation, social withdrawal, and blunted emotional responses. The positive symptoms are behaviors that were not previously present, such as hallucinations, delusions, and disorganized thought or behavior. Schizophrenia and autism share several common characteristics involving atypical social behavior, with the social cognition deficits in autism comparable to some of the negative symptoms in schizophrenia.

Many of the early studies of emotion recognition in schizophrenia were hampered by methodological limitations, including small and biased samples, nonrepresentative control groups, and non-standardized stimuli (Morrison, Bellack, & Mueser, 1988). This initially built an unclear picture of whether deficits in recognizing emotional expressions were truly evident in schizophrenia (Kring & Elis, 2013). Since then, a wealth of research has led to the conclusion that difficulties in accurately recognizing facial expressions of emotion are a core feature of schizophrenia (Kohler, Hoffman, Eastman, Healey, & Moberg, 2011; Tremeau, 2006). Patients with schizophrenia show deficits in identifying, discriminating, and judging the intensity of emotional expressions. Evidence for emotion recognition deficits has been reported across different types of tasks and stimuli. This has included faces that are familiar and unknown faces, and paradigms involving facial as well as vocal cues of emotion (Dondaine et al., 2014; Kohler, Walker, Martin, Healey, & Moberg, 2009).

There is now general agreement that deficits in accurately reading emotional expressions are a core feature of schizophrenia, although these effects can be moderated by other variables including antipsychotic medication, clinical characteristics, and current hospitalization (Kohler et al., 2011). A number of studies have revealed that emotion recognition difficulties may be specific to basic expressions with a negative valence (Morris, Weickert, & Loughland, 2009; van't Wout, Aleman, Bermond, & Kahn, 2007). For example, Kohler et al. (2003) tested individuals with schizophrenia and control participants by having them identify the emotion in photographs of people displaying different intensities of basic emotion expressions, as well as neutral faces. It was reported that schizophrenics had reduced accuracy compared to controls for fear, disgust, and neutral expressions, and that those with schizophrenia were more likely than controls to mislabel neutral faces as expressing disgust or fear. There was also a correlation found between accuracy of emotion recognition and negative symptoms of schizophrenia, especially for the identification of fearful expressions. Evidence for reduced accuracy in recognizing emotional expressions with a negative valence has also been reported by van't Wout et al. (2007), who tested patients with schizophrenia and healthy controls on various tasks including an emotion labeling task with two expressions displaying a negative valence (fear and anger) and two expressions displaying a non-negative valence (happy and neutral). They found that people with schizophrenia were less accurate in identifying displays with a negative valence (namely fear), but not for non-negative expressions. Furthermore, they reported that the reduced accuracy to recognize fear expressions was worse in people with greater negative symptoms. Brüne (2005) also found that basic emotion recognition deficits in those with schizophrenia were more evident in emotions with a negative valence, and that the deficits could not be fully explained by more general deficits in executive functioning.

Deficits in recognizing negative emotional expressions may even be a trait feature of schizophrenia, as there is evidence for emotion identification deficits in undiagnosed siblings of people with schizophrenia. Leppänen et al. (2008) found that siblings of people with schizophrenia showed reduced accuracy for identifying emotional expressions with a negative valence versus positive valence, suggesting that difficulties in recognizing negative expressions may potentially be a genetic trait of the disorder. It has also been reported that emotion recognition deficits are evident in people with early-onset and first-episode psychosis, with the most consistent findings involving deficits in the recognition of negative basic emotions (Barkl, Lah, Harris, & Williams, 2014). Difficulties in recognizing negative expressions of emotion have also been linked to differences in social functioning. For example, Abram et al. (2014) looked at the identification of five basic emotions and neutral expressions in 59 adults with schizophrenia and 41 healthy controls and reported that greater accuracy for negative valence emotions was related to better everyday life skills, which included financial and communication abilities people typically carry out in daily life. A further investigation found that the ability to recognize expressions with a negative valence predicted the functional status and outcome of people with schizophrenia, after controlling for symptomology (Brittain, Ffytche, & Surguladze, 2012).

The deficits for recognizing negative basic emotions seen in schizophrenia implicate atypical functioning of the primary emotion processing system centered on the amygdala (Holt & Phillips, 2009). There is evidence from fMRI tasks showing reduced activity in the amygdala in those suffering schizophrenia when asked to read emotional expressions from faces (Gur et al., 2002). This is a similar story to that seen in autism, which is interesting as there are many behavioral overlaps between the disorders, especially with the negative symptoms of schizophrenia.

In conclusion, the findings have revealed reduced accuracy in recognizing emotional expressions in schizophrenia. The deficits in perceiving emotions in others are particularly evident for expressions with a negative valence, which implicates dysfunction of the amygdala-based emotion system in the brain involving hypo-activation to facial expressions.

Recognition of emotional expressions in depression

Depression is a disorder characterized by consistent feelings of sadness accompanied with a loss of interest in normal activities (APA, 2013). People with depression suffer altered social functioning, which often includes reduced social interactions and less enjoyment from spending time with family and friends. Key theories about depression propose a cognitive bias that involves interpreting information in an overly negative manner (Beck, 1967). This negative bias reinforces a negative view of the world which further diminishes interactions with others. Since reading the emotional states of others is important for guiding behavior and interactions, difficulties in accurately perceiving emotional expressions in depression would likely lead to significant social problems in everyday life. Given theories of a negative bias, this would be particularly important if the emotion reading difficulties involve a reduced perception of positive affect or enhanced perception of negative affect in others.

The findings about the accuracy for recognizing emotional expressions in depression have been mixed to date, making the current picture unclear. Methodological differences between studies and the nature of samples included may contribute to mixed results. Some reviews and meta-analyses have reported there are general deficits in recognizing emotional expressions in depression (Demenescu, Kortekaas, den Boer, & Aleman, 2010; Kohler et al., 2011). For example, Persad and Polivy (1993) used a set of photographs of people expressing the six basic emotions and found that depressed people had deficits in accurately reading the emotional expressions, regardless of the specific emotion category. However, many of the studies in depression have included participants who are on medications, and the use of medication is thought to produce more general impairments in recognizing emotional expressions.

Others have suggested there are differences in accuracy for specific basic emotions in depression, including reduced accuracy for happy expressions alongside intact, or even enhanced, accuracy for sadness (Bourke, Douglas, & Porter, 2010; Gur et al., 1992). Enhanced accuracy for sadness was shown in a study by Gollan, McCloskey, Hoxha, and Coccaro (2010), who presented to non-medicated adults with depression and healthy controls photographs of emotional expressions portraying the six basic emotions across different intensity levels ranging from 10% to 80% of the full expression. They found that participants with depression showed greater accuracy to identify sad expressions than healthy controls, but no group differences in identifying other basic emotion expressions. This finding is consistent with a recent meta-analysis of emotion recognition studies in depression that reported that emotion recognition deficits were evident for all the basic emotions, apart from sadness (Dalili, Penton-Voak, Harmer, & Munafo, 2014). People with depression are reported to show a negative bias in emotion recognition, with enhanced accuracy for sadness alongside misattributions of happy, neutral, and ambiguous expressions, which are perceived in a more negative way (Bourke et al., 2010). Research has shown evidence for this, as depressed people are more likely than controls to misattribute sadness to faces displaying neutral expressions (Leppänen, Milders, Bell, Terriere, & Hietanen, 2004), and are more likely to report perceiving sadness in facial displays expressing other emotions (Dalili et al., 2014; Hale, 1998).

A negative response bias for emotional expressions is generally consistent with key theories of cognitive biases in depression proposing they interpret information in an overly negative manner. A negative bias in perceiving emotional expression in depression suggests atypical functioning in the amygdala-based emotion system in the brain, although the dysfunction appears different in depression compared to autism and schizophrenia. A review of the literature of neuroimaging studies in depression reported evidence for hyper-activation of the amygdala in response to negative emotion expression stimuli, alongside hypo-activation for positive valence expressions (Stuhrmann, Suslow, & Dannlowski, 2011). Therefore, greater activation of the amygdala-based emotion system in depression may lead to enhanced perception of negative emotional expressions, such that affected individuals perceive negative expressions in faces that are not actually displaying a negative emotion.

In conclusion, depression is thought to involve a negative bias in recognizing facial expressions of others, with enhanced negative and reduced positive perception. The negative bias in emotion recognition is consistent with key theories about depression proposing they view the world in an overly negative manner, which may involve amygdala hyperactivity.

Recognition of emotional expressions in social anxiety disorder

Another disorder involving deficits in social cognition is social anxiety disorder, also referred to as social phobia. Social anxiety is a crippling social disorder where affected people have a persistent fear or anxiety about social or performance situations that are beyond the nature of the actual threat involved (APA, 2013). People with social anxiety disorder avoid social interactions and experience excessive anxiety during interactions because they fear potential scrutiny and embarrassment that may emerge from what they say or how they behave. Cognitive theories propose this fear of negative evaluation by others in social anxiety disorder is associated with enhanced vigilance toward negative cues in others, and in potential signs of embarrassment in themselves (Rapee & Heimberg, 1997). Studies assessing social attention have reported vigilance for negative stimuli in people with social anxiety, particularly for expressions of anger, which might reflect an obvious sign of negative evaluation by others. This negative bias in social anxiety is thought to arise from a lower threshold in detecting threat, which may relate to the hyperactivity in the amygdala.

Research has found that people with social anxiety are more likely to perceive negative valence in faces displaying neutral expressions. Joormann and Gotlib (2006) presented groups of adult participants with social anxiety disorder, depression, and typical controls morphed sequences of facial emotions that changed from neutral to one of a number of emotions including sadness, anger, fear, and happiness. While the participants with social anxiety did not show difficulties overall in recognizing emotional expressions in comparison to the control or depressed groups, they needed less facial information in order to accurately identify expressions of anger. This finding reinforces the idea that people with social anxiety have an attention bias toward threatening facial information that leads them to identify such negative expressions easier than controls, which may be due to a greater sensitivity for perceiving threatening information. A bias to more readily perceive anger has also been shown by Bell et al. (2011), who presented social anxiety patients and controls with a series of static images of faces that displayed happy, angry, fearful, sad, or disgusted expressions. They reported no overall differences in recognition accuracy between groups, but the errors in recognition made by those with social anxiety disorder were more likely to involve misattributing other basic emotions as expressing anger and misperceive neutral as being angry. Together, these results demonstrate a negative bias in social anxiety, with a greater sensitivity to perceive anger in faces, even when little or no expression of anger is displayed.

A bias for enhanced perception of emotional expressions has also been reported in people from the general population with varying degrees of social anxiety traits. A study by Button, Lewis, Penton-Voak, and Munafò (2013) measured social anxiety traits in 868 female participants from the general population and allocated those with the top and bottom scores within the normal range into high- and low-social anxiety groups. Facial emotion recognition accuracy was measured by showing them a series of images of emotional expressions displaying happiness, anger, fear, sadness, or disgust, and asking them to choose which of those emotions or neutral best described the expression seen. Various intensities of expression were shown for each emotion, ranging from 9% to 65% intensity for the degree of emotion displayed. They also included a social cost task where they had participants rate the images of emotional expressions on what it would be like interacting with them, from very bad to very good. Those with higher social anxiety didn't show reduced overall accuracy in labeling emotions. However, they were more likely than the low-anxious group to label the low-intensity expressions as emotional, but these were more often labeled as the wrong emotions. They were also more likely than the low-social anxiety group to rate interacting with emotional expressions as being very bad for them, and this effect was particularly pronounced for negative emotions. This pattern of emotional expression perception could help cause and maintain social anxiety (Rapee & Heimberg, 1997).

The increased tendency to misattribute emotions as having greater negative valence or association in social anxiety disorder may be related to the amygdala-based emotion system. Brain-imaging studies have generally shown increased amygdala activity in those with social anxiety when reading facial expressions, particularly faces with threat expressions (Phan, Fitzgerald, Nathan, & Tancer, 2006). Interestingly, the amygdala in those with social anxiety disorder is reported to show heightened activation when perceiving neutral faces compared to control participants (Cooney, Atlas, Joormann, Eugène, & Gotlib, 2006). The heightened amygdala response could be related to the negative bias in perception for seeing emotional expressions in faces not actually displaying emotions.

In conclusion, research has generally found reduced accuracy in recognizing emotional expressions in social anxiety disorder, which may involve biases with greater sensitivity toward some emotional expressions and for misattributing emotional expressions inaccurately to faces. When biases in emotion recognition exist in social anxiety disorder, they tend to involve expressions with a negative valence over other types of emotions. Similar to depression, the biases for misattributing emotions in facial expressions in social anxiety disorder may involve heightened activity of the amygdala.

Recognition of emotional expressions in borderline personality disorder

Borderline personality disorder (BPD) is a condition that involves instability in impulse control, emotion regulation, and self-perception and is characterized by impairments in psychosocial functioning (APA, 2013). These often lead to difficulties in maintaining relationships with others, which is a key aspect of the disorder. People with BPD often feel strong bonds or reliance on significant others, which can quickly turn to anger or intense feelings of disdain toward that same individual. It has been suggested that this pattern of behavior could result from dysfunctions in cognitive and emotional processing, including difficulties in accurately recognizing the feelings of others and biases in empathy (Harari, Shamay-Tsoory, Ravid, & Levkovitz, 2010). Many theories about BPD emphasize the difficulties in emotion regulation and that they may become more problematic with stressful environmental conditions. The mildest triggers may set a person with BPD into crisis and produce increased sensitivity to negative social cues and a bias toward negative appraisals.

An early study investigating emotion recognition reported increased accuracy for recognizing emotional expressions in those with BPD compared to controls, rather than deficits (Wagner & Linehan, 1999). They asked participants to describe the emotional expression seen in the face and found greater accuracy in the descriptions for fearful expressions, and further that those with BPD often misattributed neutral faces as looking fearful. This finding is further supported by the results of a study that used the "Reading the Mind in the Eyes" task to assess the recognition of mental and emotional states in those with BPD from only the eye regions of faces (Fertuck et al., 2009), which reported enhanced accuracy in patients with BPD compared to controls in the ability to read the mental and emotional states of others from the eyes. Domes et al. (2008) investigated enhanced sensitivity to emotional expressions in BPD using morphed stimuli, which involved facial expressions of emotion that changed from neutral expressions to an emotional expression in a series of steps. Participants responded when they detected an emotion and also reported which emotion they detected. Patients with BPD did not show enhanced sensitivity in their perception for the emergence of emotional expressions overall, as would have been predicted from the results of earlier research (e.g., Wagner & Linehan, 1999). Instead, those with

BPD showed a learning effect for emotional expressions, where the accuracy to detect the emergence of an emotion significantly improved over the course of the study. The same learning effect was not found in the control participants. Together, the results suggest that those with BPD might show greater accuracy than normal to recognize emotional expressions, perhaps from heightened sensitivity toward the emotions of others.

The perceptual bias in BPD may also extend to recognition with neutral and mild expressions. A study by Dyck et al. (2009) had participants discriminate whether an emotional expression they saw was fearful, angry, or neutral, and found that BPD participants were more likely than controls to perceive neutral emotions as having a negative valence. Interestingly, the same participants with BPD showed similar accuracy for identifying emotional expressions as controls when given unlimited timing in a forced choice emotion recognition task, suggesting the negative bias is most evident under conditions requiring faster emotion processing. A meta-analysis of ten studies suggested that the differences in facial emotion recognition tasks seen in those with BPD were not accounted for by the difference in the speed of presentation (Daros, Zakzanis, & Ruocco, 2013). Another emotion recognition study also reported the same pattern of results for attributing emotional expressions to neutral displays in those with BPD, as well as perceiving mildly sad expressions as more intensely sad than controls (Daros, Uliaszek, & Ruocco, 2014). Together, these results suggest a bias in perceiving emotional expressions in BPD, characterized by increased accuracy and a tendency to perceive emotions in faces displaying little or no emotional expressions.

Biases in emotion recognition may relate to altered activity in the amygdala-centered emotion system. In a review of the literature, Domes, Schulze, and Herpertz (2009) highlighted the evidence for increased activation in the amygdala in response to emotional faces in those with BPD. A further review and meta-analysis reported a large effect size for research demonstrating a heightened amygdala response during emotion recognition, and that this was seen across a range of neuroimaging studies (Mitchell, Dickens, & Picchioni, 2014).

In conclusion, findings from research provide evidence for a bias in emotion perception in BPD, characterized by enhanced accuracy to identify emotional expressions and a tendency to perceive emotions in faces displaying little or no expressions. The biases in emotion processing in BPD may involve hyper-activation of the amygdala.

General conclusions

This chapter has reviewed evidence about the accuracy in recognizing facial expressions of emotion within psychopathology. Comparisons are difficult in the field of emotion recognition given the differences in the paradigms and stimuli used. However, some commonalities and differences were noted in the ability to read facial expressions across the disorders reviewed. In general, two different patterns emerged within the findings presented. First, the research in autism and schizophrenia revealed reduced accuracy for perceiving facial expressions of emotion that tended to be most evident for emotions with a negative valence. Second, a different pattern was noted for the disorders of depression, social anxiety, and BPD, where evidence suggested enhanced accuracy for some emotional expressions alongside misattributions in the perception of facial expressions characterized by the tendency to perceive emotions in faces displaying little or no emotional expressions. Together, the research in psychopathology showed differences in perception tended to involve emotions with a negative valence, with disorders differentiated by either reduced or enhanced perception in reading negative emotions.

Interestingly, there was evidence that the two different general patterns in the perception of emotional expressions also related to differential activation in the primary emotion system centered on the amygdala in the brain for the same disorders. Autism and schizophrenia were generally associated with reduced amygdala activation during emotion perception, while depression, social anxiety, and BPD were associated more with hyperactivity of the amygdala. Since the amygdala is important for perceiving facial signals of threat and danger (Adolphs et al., 1999), reduced ability to accurately read negative facial expressions of others in autism and schizophrenia may involve reduced awareness about the mental states and intentions of others related to potential negative affect. In contrast, depression, social anxiety, and BPD may involve enhanced awareness about the negative affect in others, and misperceptions leading to reading negative emotions in others when they are not actually experiencing it. These difficulties and differences in perceiving the facial expressions of others in psychopathology would likely cause problems in everyday life and be intertwined with their clinical characteristics.

However, little is currently known about how differences in emotion recognition accuracy within each disorder may relate to social functioning and to the clinical characteristics. Atypical emotion recognition in depression may relate most to expressions of sadness, which might represent a mood congruence effect. Social anxiety might involve differences in processing anger, which is a signal of hostility in others and might be associated with the fear of negative evaluation that is central to the disorder. In contrast, deficits in processing negative emotional expressions characterize both autism and schizophrenia despite the differences in behavioral symptoms. It is unlikely that differences in emotion recognition represent primary causal mechanisms for mental disorders, since emotion recognition deficits are not the primary symptom, given the heterogeneous nature of the various disorders. It might seem more likely that differences in processing emotional expression could be a more general consequence of having a mental disorder, and that atypical social cognition is an epiphenomenon within psychopathology, which includes atypical processing of emotional expressions and in relating to others. Given how common difficulties in social cognition are in psychopathology, further research is needed to uncover the exact nature of the emotion recognition deficits to better understand the etiology of mental disorders and for the development of effective treatments.

Future emotion recognition research in psychopathology should emphasize replication studies involving consistent tasks and stimuli so that data can be compared and pooled within disorders. The use of similar methods across studies and the inclusion of relevant psychiatric control groups would allow more direct comparisons to be made across disorders, in order to see similarities and differences in emotion recognition accuracy in psychopathology. Future directions in research should also involve the development of more sensitive measures of emotion recognition ability in order to see both reduced and enhanced accuracy that are often difficult to uncover in lab-based methods. These measures should produce emotion accuracy scores that are suitably below ceiling and above chance levels, aspects that have hampered results and interpretations of some previous research in the field.

Future studies should also more commonly control relevant factors related to the disorders and emotion recognition. For example, medications are commonly used in disorders such as schizophrenia and depression but not always controlled for in the research, and these medications may affect emotion recognition and influence the results. Many studies do not include additional tasks not involving recognizing emotions from expressions, in order to test the specificity of the findings to accuracy for judging facial emotions. This is important given that many disorders involve more general deficits that could affect multiple cognitive and emotional processes, such as difficulties in executive functioning. Other factors such as the severity of the illness, current treatment, comorbidity with other disorders, and age are often not controlled for in the research, and may impact the results.

References

Abram, S. V., Karpouzian, T. M., Reilly, J. L., Derntl, B., Habel, U., & Smith, M. J. (2014). Accurate perception of negative emotions predicts functional capacity in schizophrenia. *Psychiatry Research*, 216, 6–11.

Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature*, 372, 669–672.

Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. (1995). Fear and the human amygdala. *The Journal of Neuroscience*, 15, 5879–5891.

Adolphs, R., Tranel, D., Hamann, S., Young, A. W., Calder, A. J., Phelps, E. A., Anderson, A., Lee, G., & Damasio, A. R. (1999). Recognition of facial emotion in nine individuals with bilateral amygdala damage. *Neuropsychologia*, 37, 1111–1117.

American Psychological Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: American Psychological Association.

Anderson, A. K., & Phelps, E. A. (2000). Expression without recognition: contributions of the human amygdala to emotional communication. *Psychological Science*, 11, 106–111.

Argyle, M. (1972). Nonverbal communication in human social interaction. In R. A. Hinde (Ed.), *Nonverbal communication* (pp. 243–269). Cambridge, UK: Cambridge University Press.

Ashwin, C., Chapman, E., Colle, L., & Baron-Cohen, S. (2006). Impaired recognition of negative basic emotions in autism: A test of the amygdala theory. *Social Neuroscience*, 1, 349–363.

Ashwin, C., Baron-Cohen, S., Wheelwright, S., O'Riordan, M., & Bullmore, E. T. (2007). Differential activation of the amygdala and the "social brain" during fearful face-processing in Asperger Syndrome. *Neuropsychologia*, 45, 2–14.

Barkl, S. J., Lah, S., Harris, A. W., & Williams, L. M. (2014). Facial emotion identification in early-onset and first-episode psychosis: A systematic review with meta-analysis. *Schizophrenia Research*, 159, 62–69.

Baron-Cohen, S. (2010). Empathizing, systemizing, and the extreme male brain theory of autism. *Progress in Brain Research*, 186, 167–175.

Baron-Cohen, S., Ring, H., Bullmore, E., Wheelwright, S., Ashwin, C., & Williams, S. (2000). The amygdala theory of autism. *Neuroscience and Biobehavioral Reviews*, 24, 355–364.

Beck, A. T. (1967). *Depression: Clinical, experimental, and theoretical aspects*. Philadelphia: University of Pennsylvania Press.

Bell, C., Bourke, C., Colhoun, H., Carter, F., Frampton, C., & Porter, R. (2011). The misclassification of facial expressions in generalised social phobia. *Journal of Anxiety Disorders*, 25, 278–83.

Blair, R. J. (2003). Facial expressions, their communicatory functions and neuro-cognitive substrates. *Philosophical Transactions of the Royal Society of London: B Biological Sciences*, 358, 561–572.

Bormann-Kischkel, C., Vilsmeier, M., & Baude, B. (1995). The development of emotional concepts in autism. *Journal of Child Psychololgy and Psychiatry*, 36, 1243–1259.

Bourke, C., Douglas, K., & Porter, R. (2010). Processing of facial emotion expression in major depression: a review. *Australian and New Zealand Journal of Psychiatry*, 44(8), 681–696.

Brittain, P. J., Ffytche, D. F., & Surguladze, S. A. (2012). Emotion perception and functional outcome in schizophrenia: The importance of negative valence and fear. *Psychiatry Research*, 200, 208–213.

Brüne, M. (2005). Emotion recognition, "theory of mind," and social behavior in schizophrenia. *Psychiatry Research*, 133, 135–147.

Buitelaar, J. K., Van Der Wees, M., Swaab-Barneveld, H., & Van Der Gaag, R. J. (1999). Theory of mind and emotion-recognition functioning in autistic spectrum disorders and in psychiatric control and normal children. *Development and Psychopathology*, 11, 39–58.

Button, K., Lewis, G., Penton-Voak, I., & Munafò, M. (2013). Social anxiety is associated with general but not specific biases in emotion recognition. *Psychiatry Research*, 210, 199–207.

Calder, A. J. (1996). Facial emotion recognition after bilateral amygdala damage: Differentially severe impairment of fear. *Cognitive Neuropsychology*, 13, 699–745.

Celani, G., Battacchi, M. W., & Arcidiacono, L. (1999). The understanding of the emotional meaning of facial expressions in people with autism. *Journal of Autism and Developmental Disorders*, 29, 57–66.

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends in Cognitive Sciences*, 16, 231–239.

Cooney, R. E., Atlas, L. Y., Joormann, J., Eugène, F., & Gotlib, I. H. (2006). Amygdala activation in the processing of neutral faces in social anxiety disorder: Is neutral really neutral? *Psychiatry Research: Neuroimaging*, 148, 55–59.

Dalili, M. N., Penton-Voak, I. S., Harmer, C. J., & Munafo, M. R. (2014). Meta-analysis of emotion recognition deficits in major depressive disorder. *Psychological Medicine*, 6, 1–10.

Daros, A. R., Zakzanis, K. K., & Ruocco, A. C. (2013). Facial emotion recognition in borderline personality disorder. *Psychological Medicine*, 43, 1953–1963.

Daros, A. R., Uliaszek, A. A., & Ruocco, A. C. (2014). Perceptual biases in facial emotion recognition in borderline personality disorder. *Personality Disorders*, 5, 79–87.

Darwin, C. (1872). *The expression of the emotions in man and animals*. London: Harper Collins.

Demenescu, L. R., Kortekaas, R., den Boer, J., & Aleman, A. (2010). Impaired attribution of emotion to facial expressions in anxiety and major depression. *PLoS One*, 5, e15058.

Domes, G., Czieschnek, D., Weidler, F., Berger, C., Fast, K., & Herpertz, S. C. (2008). Recognition of facial affect in borderline personality disorder. *Journal of Personality Disorders*, 22, 135–147.

Domes, G., Schulze, L., & Herpertz, S. (2009). Emotion recognition in borderline personality disorder – a review of the literature. *Journal of Personality Disorders*, 23, 6–19.

Dondaine, T., Robert, G., Péron, J., Grandjean, D., Vérin, M., Drapier, D., & Millet, B. (2014). Biases in facial and vocal emotion recognition in chronic schizophrenia. *Frontiers in Psychology*, 5, 900.

Dyck, M., Habel, U., Slodczyk, J., Schlummer, J., Backes, V., Schneider, F., & Reske, M. (2009). Negative bias in fast emotion discrimination in borderline personality disorder. *Psychological Medicine*, 39, 855–864.

Ekman, P., & Friesen, W. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17, 124–129.

Ekman, P., & Friesen, W. (1975). *Unmasking the face*. Oxford, UK: Prentice-Hall.

Feldman, R., Mcgee, G. G., Mann, L. J., & Strain, P. S. (1993). Nonverbal affective decoding ability in children with Autism and in typical preschoolers. *Journal of Early Intervention*, 17, 341–350.

Fertuck, E. A., Jekal, A., Song, I., Wyman, B., Morris, M. C., Brodsky, B. S., & Stanley, B. (2009). Enhanced "Reading the Mind in the Eyes" in borderline personality disorder compared to healthy controls. *Psychological Medicine*, 39, 1979–1988.

Flack, W. F., & Laird J. D. (1998). *Emotions in psychopathology*. Oxford: Oxford Press.

Gaigg, S. B. (2012). The interplay between emotion and cognition in autism spectrum disorder: implications for developmental theory. *Frontiers in Integrative Neuroscience*, 6, 113–152.

Gollan, J. K., McCloskey, M., Hoxha, D., & Coccaro, E. F. (2010). How do depressed and healthy adults interpret nuanced facial expressions? *Journal of Abnormal Psychology*, 119, 804–810.

Gur, R. C., Erwin, R. J., Gur, R. E., Zwil, A. S., Heimberg, C., & Kraemer, H. C. (1992). Facial emotion discrimination: II. Behavioral findings in depression. *Psychiatry Research*, 42(3), 241–251.

Gur, R. E., Mcgrath, C., Chan, R. M., Schroeder, L., Turner, T., Turetsky, B. I., Kohler, C., Alsop, D., Maldjian, J., & Ragland, J. D. (2002). An fMRI study of facial emotion processing in patients with schizophrenia. *American Journal of Psychiatry*, 159, 1992–1999.

Hale, W. W. (1998). Judgment of facial expressions and depression persistence. *Psychiatry Research*, 80, 265–274.

Harari, H., Shamay-Tsoory, S. G., Ravid, M., & Levkovitz, Y. (2010). Double dissociation between cognitive and affective empathy in borderline personality disorder. *Psychiatry Research*, 175, 277–279.

Harms, M. B., Martin, A., & Wallace, G. L. (2010). Facial emotion recognition in autism spectrum disorders: A review of behavioral and neuroimaging studies. *Neuropsychology Review*, 20, 290–322.

Hobson, R. P. (1986). The autistic child's appraisal of expressions of emotion: A further study. *Journal of Child Psychology and Psychiatry*, 27, 671–680.

Holt, D. J., & Phillips, M. L. (2009). The human amygdala in schizophrenia. In P. Whalen & E. Phelps (Eds.), *The human amygdala*. New York: The Guildford Press.

Howard, M. A., Cowell, P. E., Boucher, J., Broks, P., Mayes, A., Farrant, A., & Roberts, N. (2000). Convergent neuroanatomical and behavioural evidence of an amygdala hypothesis of autism. *Neuroreport*, 11, 2931–2935.

Joormann, J., & Gotlib, I. H. (2006). Is this happiness I see? Biases in the identification of emotional facial expressions in depression and social phobia. *Journal of Abnormal Psychology*, 115, 705–714.

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child*, 2, 217–250.

Kohler, C. G., Hoffman, L. J., Eastman, L. B., Healey, K., & Moberg, P. J. (2011). Facial emotion perception in depression and bipolar disorder: a quantitative review. *Psychiatry Research*, 188, 303–309.

Kohler, C. G., Turner, T. H., Bilker, W. B., Brensinger, C. M., Siegel, S. J., Kanes, S. J., Gur, R. E., & Gur, R. C. (2003). Facial emotion recognition in schizophrenia: intensity effects and error pattern. *American Journal of Psychiatry*, 160, 1768–1774.

Kohler, C. G., Walker, J. B., Martin, E. A., Healey, K. M., & Moberg, P. J. (2009). Facial emotion perception in schizophrenia: A meta-analytic review. *Schizophrenia Bulletin*, 36, 1009–1019.

Kring, A. M., & Bachorowski, J. A. (1999). Emotion and psychopathology. *Cognition and Emotion*, 13, 575–600.

Kring, A. M., & Elis, O. (2013). Emotion deficits in people with schizophrenia. *Annual Review of Clinical Psychology*, 9, 409–433.

Law Smith, M. J., Montagne, B., Perrett, D. I., Gill, M., & Gallagher, L. (2010). Detecting subtle facial emotion recognition deficits in high-functioning Autism using dynamic stimuli of varying intensities. *Neuropsychologia*, 48, 2777–2781.

LeDoux, J. E. (1996). *The emotional brain*. New York: Simon and Schuster.

Leppänen, J. M., Milders, M., Bell, J. S., Terriere, E., & Hietanen, J. K. (2004). Depression biases the recognition of emotionally neutral faces. *Psychiatry Research*, 128, 123–133.

Leppänen, J. M., Niehaus, D. J., Koen, L., Du Toit, E., Schoeman, R., & Emsley, R. (2008). Deficits in facial affect recognition in unaffected siblings of schizophrenia patients: evidence for a neurocognitive endophenotype. *Schizophrenia Research*, 99, 270–273.

Loveland, K. A., Tunali-Kotoski, B., Chen, Y. R., Ortegon, J., Pearson, D. A., Brelsford, K. A., & Gibbs, M. C. (1997). Emotion recognition in autism: verbal and nonverbal information. *Development and Psychopathology*, 9, 579–593.

Mitchell, A. E., Dickens, G. L., & Picchioni, M. M. (2014). Facial emotion processing in borderline personality disorder: A systematic review and meta-analysis. *Neuropsychological Reviews*, 24, 166–184.

Morris, R. W., Weickert, C. S., & Loughland, C. M. (2009). Emotional face processing in schizophrenia. *Current Opinion in Psychiatry*, 22, 140–146.

Morrison, R. L., Bellack, A. S., & Mueser, K. T. (1988). Deficits in facial-affect recognition and schizophrenia. *Schizophrenia Bulletin*, 14, 67–83.

Persad, S. M., & Polivy, J. (1993). Differences between depressed and nondepressed individuals in the recognition of and response to facial emotional cues. *Journal of Abnormal Psychology*, 102, 358.

Phan, K. L., Fitzgerald, D. A., Nathan, P. J., & Tancer, M. E. (2006). Association between amygdala hyperactivity to harsh faces and severity of social anxiety in generalized social phobia. *Biological Psychiatry*, 59, 424–429.

Rapee, R. M., & Heimberg, R. G. (1997). A cognitive-behavioral model of anxiety in social phobia. *Behaviour Research and Therapy*, 35, 741–756.

Shah, A., & Frith, U. (1983). An islet of ability in autistic children: A research note. *Journal of Child Psychology and Psychiatry*, 24, 613–620.

Stuhrmann, A., Suslow, T., & Dannlowski, U. (2011). Facial emotion processing in major depression: a systematic review of neuroimaging findings. *Biology of Mood and Anxiety Disorders*, 1, 10.

Tremeau, F. (2006). A review of emotion deficits in schizophrenia. *Dialogues in Clinical Neuroscience*, 8, 59–70.

Uljarevic, M., & Hamilton, A. (2013). Recognition of emotions in autism: A formal meta-analysis. *Journal of Autism and Developmental Disorders*, 43, 1517–1526.

Van't Wout, M., Aleman, A., Bermond, B., & Kahn, R. S. (2007). No words for feelings: alexithymia in schizophrenia patients and first-degree relatives. *Comprehensive Psychiatry*, 48, 27–33.

Wagner, A. W., & Linehan, M. M. (1999). Facial expression recognition ability among women with borderline personality disorder: implications for emotion regulation? *Journal of Personality Disorders*, 13, 329–344.

10 A lifespan developmental perspective on interpersonal accuracy

Derek M. Isaacowitz, Ishabel M. Vicaria, and Matthew W. E. Murry

Abstract
In this chapter, we consider evidence for changes in emotion perception accuracy and social perception accuracy across the lifespan. Perhaps not surprisingly, the focus in the child development literature is on how and when accuracy improves with age, whereas the focus in adult development and aging is on how these processes may decline at older ages. Descriptive work on these questions is more advanced than research on specific mechanisms, and there has been some attention to individual differences as well. Despite discontinuities between the two literatures in methods and constructs, evidence to date suggests emotion perception increases in childhood and declines into old age, but social perception may not show aging-related declines to the same degree.

There are many reasons to be curious about age differences and similarities across the lifespan in the ability to accurately perceive other people's emotional experiences, referred to as *emotion perception*, as well as other individual or dyadic attributes (such as competence, health, rapport), referred to as *social perception*. First, we may be interested in to what degree these abilities develop in concert with general cognitive ability, and to what extent accuracy in emotion and social perception decline at the same time as age-related decrements occur in mental functioning. Second, we may wonder whether accuracy of emotion and social perception varies as a function of experience – people with a longer history of exposure to certain types of experiences might be better at making accurate judgments regarding attributes that are relevant to those experiences. Some of these experiences may be age-graded, such that individuals of some ages are more likely to have them in their everyday lives than are members of other age groups. For example, children may more frequently make judgments about teachers, whereas older adults may more frequently make judgments about doctors. In other cases, we might expect the effects of experience to be additive over time, so more years of life would suggest more experience making judgments and perhaps getting feedback about their accuracy. In yet other cases, we might expect idiosyncratic experiences to guide some individuals toward more accuracy and others not, so development might serve to increase variability in a population. These are just some of the reasons why it is interesting and useful to consider age differences in the accuracy of emotion and social perception.

In this chapter, we consider the current state of knowledge on accuracy in judgments of emotional expressions (emotion perception, relatively more studied in the developmental literature) and in judgments of other personal or dyadic attributes (social perception, relatively less studied developmentally). Because these literatures tend to have quite distinct features, we consider them separately. Perhaps more surprising, though, is the observation that there is no true lifespan developmental literature on accuracy in either emotion or social perception; instead, for each, there is some child development work and some usually unrelated adult development and aging work. The constructs and methods tend to vary between child development research, on the one hand, and adult development and aging research on the other. At the end of the chapter, we consider what might be gained from considering emotion and social perception using a truly lifespan perspective, but for the purpose of reviewing past work we will consider child and adult development work separately.

Accuracy in emotion perception

Children

The starting point for any consideration of emotion perception across the lifespan must be how it develops in childhood. While some studies focus on the development of accurate perception of positive vs. negative states, other work looks specifically at accuracy in perception of discrete emotional states. Research suggests three stages in the development of accurate emotion perception in children: discrimination, categorization, and labeling. These three components create a foundation for the development of accurate emotion perception across the lifespan, and we consider them in turn below.

Emotional expression discrimination. Children are born with a range of visual abilities that aid in the development of emotion perception, such as pattern recognition (Walker-Andrews, 1997). Newborns can only discern a blurry face with a hairline, eyes, nose, and mouth (Banks & Ginsburg, 1985). In the earliest stages of emotion perception, the child must discriminate emotional expressions before he or she can categorize them as sad or happy. Discrimination is important because for a child to identify an emotion accurately, they must recognize the differences between emotions. Because newborns and infants cannot articulate the perceived emotion, researchers test discrimination with visual preference and discrimination paradigms. In the visual preference paradigm, two or more expressions (e.g., happiness and sadness) are presented simultaneously, and the amount of time the child looks at each is measured. If the child looks at one expression longer than the other he or she is said to be discriminating between them (e.g., LaBarbera, Izard, Vietze, & Parisi, 1976). In the discrimination paradigm, a child is shown a facial expression (e.g., happiness) until his or her looking time decreases below a certain criterion (e.g., two seconds), and then the child is shown a novel expression (e.g., surprise). If the child looks longer at the novel expression, discrimination between the emotions is assumed (e.g., Field, Woodson, Greenberg, & Cohen, 1982). The visual preference paradigm assesses which expressions attract attention, but it is susceptible to motivation. If there is no difference in looking time, it could be because newborns are unable to discriminate or because they do not have a preference between the expressions. In habituation, a difference in looking time suggests that the newborn notices a difference between the expressions, but this paradigm is susceptible to previous experience, such as a developed preference for a specific emotion. Both methods are descriptive and neither can explain the underlying mechanisms that cause emotion discrimination to develop.

Using the habituation paradigm and using a within-subject design, newborns appear to discriminate happiness, sadness, and surprise expressions (Field et al., 1982). However, the use of live models for expressions raises questions of consistency and intensity of the expressions. Studies using a control group, where the control group only viewed one expression to prevent regression to the mean (Young-Browne, Rosenfeld, & Horowitz, 1978) or a visual preference task that measures looking time between two expressions presented simultaneously (Barrera & Maurer, 1981; LaBarbera et al., 1976) provide more support for discrimination among emotions.

According to findings with older infants, assessment of discrimination is dependent on the habituating emotion. Three-month-old children may discriminate happiness from surprise (Young-Browne et al., 1978), and anger (Barrera & Maurer, 1981), but not from sadness (Young-Browne et al., 1978). By five months, infants tend to discriminate expressions of sadness from fear, anger from sadness, and anger from fear (Schwartz, Izard, & Ansul, 1985). Finally, by seven months, infants tend to discriminate expressions of fear from those of happiness. On the surface, it appears that happiness discrimination ability, from both negative and positive emotions, is the first to form; however, the results tend to be reliant on stimuli order. Nevertheless, by seven months of age, infants seem to possess the ability to discriminate a few emotions from others.

Emotion expression categorization. Discrimination ability measures the recognition of two differing expressions, whereas the recognition that two similar expressions are different from others is categorization. To test the categorization ability, researchers use a variation of the habituation paradigm. During the habituation phase, instead of one exemplar, participants view multiple exemplars of the same expression of emotion (e.g., happiness), then in the dishabituation phase they view exemplars of novel (e.g., anger) and familiar (e.g., happiness) expressions. If the participants look longer at the novel expression compared to the familiar one, he or she is said to have the categorization ability for those emotions. Some evidence suggests that as early as four-month-old infants begin categorizing happiness and anger (Serrano, Iglesias, & Loeches, 1992); however, another study demonstrated that seven-month-old infants could not categorize fear, but they could categorize happiness using one target (Nelson, Morse, & Leavitt, 1979), multiple targets (Nelson & Dolgin, 1985), and varying intensities (Ludemann & Nelson, 1988). In addition, the majority of studies found that seven-month-old infants could not categorize fear or surprise (e.g., Nelson & Dolgin, 1985), but one study found support for four- to six-month-old infants categorizing fear, anger, and surprise (Serrano et al., 1992). Finally, further evidence suggests that infants cannot categorize fear or anger until ten months when habituated with positive emotions first but not even ten-month olds can categorize anger and fear when habituated with blended positive emotions suggesting that even by this age, categorization is not fully formed and depends on the paradigm (Ludemann, 1991). Although infants may begin categorizing emotions at four months and improve by seven months, research suggests that categorization of specific emotions may not develop until later, around ten months and beyond.

Emotion labeling. Categorization is the ability to recognize similarity between expressions, but emotion labeling is the ability to verbalize the categories in which these similar emotional expressions fit. The way emotion labeling in children and emotion perception accuracy in adults are operationalized makes it possible to consider them variations of the same construct. In addition to being the final component of the development of the ability to accurately perceive emotions, labeling also allows a child to converse with others about emotions, making it an important component of emotion perception accuracy. In order to test the emotion-labeling ability in children, researchers use a free-labeling paradigm instead of a forced-choice paradigm traditionally used in research with adults. In a free-labeling paradigm, children are presented with facial expressions of emotion and asked to provide an emotion label for each expression. Generally, they are allowed to view the expression until a label is provided. Before the task, children are primed with emotion labels (e.g., happy, sad, mad, scared, yucky/disgusted) through stories. During analysis, the labels given by the participant are grouped into discrete emotion categories. Accuracy is measured within these categories (e.g., Widen & Russell, 2003).

Children seem to start with two broad labels (positive and negative), and then they create more labels for each emotion gradually (Widen & Russell, 2010). One model for the development of verbal labeling of emotional facial expressions in children (Widen & Russell, 2003) proposes that children acquire one label at a time and they vary on the exact label acquired at each age. Before 30 months, children tend to lack the ability to label the six emotion labels defined by researchers (happy, sad, angry, fear, disgust, surprise), lagging behind emotion language development which begins around 20 months (Bretherton & Beeghly, 1982). The first label to emerge is happiness at around 36 months, then followed by anger or sadness at 39 months, but all three labels are not typically formed until 45 months. At 58 months, either scared or surprise labels emerge, then both labels are acquired. Finally, sometime shortly after, disgust emerges as the last of the six basic emotion labels (Widen & Russell, 2010). However, recent work suggests that even until eight years old, the majority of children label disgusted faces as anger (Widen & Russell, 2013). Notably, this timeline was formed using the mean age at which a child develops different emotion labels; consequently, the timeframe varies somewhat across children, but the process remains similar.

As children begin to acquire emotion labels, they continue to improve their emotion perception accuracy; moreover, this improvement continues to some extent into adulthood. There are overlaps between labeling in the child literature and accuracy in the adult literature. However, the child literature investigates an accuracy threshold (scores greater than chance), whereas the adult literature uses variability in accuracy as the primary outcome. For example, in the child literature, researchers might test if four-month olds label happiness at an above-chance level when viewing expressions and define accuracy as passing that threshold. Individual differences in accuracy would entail different ages at which the threshold is reached. In contrast, in the adult literature, everyone is assumed to have a threshold of performance and variability in accuracy is simply the differences among individuals in their performance in the accuracy task.

Using the Diagnostic Analysis of Nonverbal Accuracy (DANVA; Baum & Nowicki, 1998) with children as young as four (Nowicki & Duke, 2001) through adults as old as 100 (Roberts, Nowicki, & McClure, 1998), researchers have found that as children age and through young adulthood, they become more accurate at perceiving facial and vocal expressions of children (Demertizis & Nowicki, 1997; Nowicki & Duke, 2001) and adults (Baum & Nowicki, 1998). These findings suggest a trend of higher emotion perception accuracy with age. Further research, using implicit and explicit facial emotional expression tasks held constant across groups, suggests that emotion perception accuracy increases throughout childhood and into early adulthood (Williams et al., 2009). The implicit emotion perception task is not influenced by an age difference in vocabulary or exposure duration and participants are not aware they are being asked to perceive emotions. All these findings support a trend of increasing emotion perception accuracy through childhood and into early adulthood, though the trajectory may differ for later adulthood and is discussed below.

Individual differences in emotion perception accuracy development. In tandem with visual system development, according to previously stated research, children first learn to discriminate between emotions, then categorize them, and finally label them. However, there appear to be individual differences in the timing of the development of accurate emotion perception in children. Parental socialization research gives insight into how individual differences in accuracy emerge during the development of accurate emotion perception. For example, the work of A.G. Halberstadt and colleagues (1999) suggests that children with highly expressive families are more accurate when asked to perceive facial expressions of emotion compared to those with less expressive families. One interpretation is that children who view emotions regularly have more opportunities to "practice" emotion perception, thus increasing accuracy; however, other findings demonstrate that by adulthood, individuals with less-expressive families have higher emotion perception accuracy (Halberstadt & Eaton, 2002). This pattern reversal suggests that even though in early childhood individuals' emotion perception accuracy may benefit from frequent displays of emotion, at some point it becomes more beneficial to have less-obvious displays of emotion to decode. In other words, those in less-expressive families may have difficulty early in childhood, but this difficulty improves their emotion perception accuracy by increasing the level of practice long term. Nevertheless, these findings suggest that components of the social environment may underlie some variability in emotion perception accuracy.

Severe social conditions may also constrain or promote different types of emotion perception accuracy. For instance, parents of neglected children display more negative emotions (Herrenkohl, Herrenkohl, Egolf, & Wu, 1991) compared to parents of typically developing children. If the children are "practicing" emotion perception ability with the social environment, these neglected children are "practicing" with more negative emotions and have lack of "practice" with positive emotions. Similarly, physically abused children tend to require less sensory input to detect anger in facial expressions, but more input for sadness compared to non-abused children (Pollak & Sinha, 2002). If children are exposed to more frequent negative emotional displays in neglecting or abusive environments, they may become more accurate for negative emotion perception compared to other children; however, if these children are biased toward perceiving negative emotions, this might actually make them less accurate in perceiving other (i.e., positive, neutral) emotions. While these children may just perceive more negative emotion in all expressions, some evidence suggests that physically abused children make similar emotion-specific errors compared to non-abused children, suggesting physically abused children may be more accurate for some emotions and not just biased (Pollak & Sinha, 2002). Maltreated children are exposed to complicated, inconsistent, poorly conveyed, and possibly distressing cues and have higher accuracy for certain emotions. Based on the emotions presented in their daily lives, their emotion perception skills may be influenced by the decoding difficulty rather than just the frequency and salience of expressions. This influence is most likely true for children generally.

Further individual differences have been specified in research on correlates of emotion perception accuracy in children. For example, lower scores on the DANVA assessing adult and child facial expressions have been correlated with lower social competence in children, as assessed with sociometric ratings (e.g., Verbeek, 1996), teachers' ratings (e.g., Collins, 1996), and parents' ratings (e.g., McClanahan, 1996). Lower emotion perception accuracy has also been associated with greater depression rates in boys and lower self-esteem in girls (Nowicki & Carton, 1997). Emotion perception accuracy of vocal expressions shows a similar pattern with lower accuracy correlated with lower social competence (e.g., Verbeek, 1996). Furthermore, lower accuracy of vocal expressions has been correlated with lower school performance (e.g., Collins, 1996).

Together these findings demonstrate a possible link between emotion perception accuracy in children and their levels of social competence.

Adults

The trend of increasing emotion perception accuracy continues into early adulthood. In adults, emotion perception accuracy is generally tested using static images of facial expressions of emotion displayed for various times. Using this paradigm and holding the task constant, researchers tend to find that younger adults (18–29 years old) have generally higher accuracy compared to children (10 years old: e.g., Mondloch, Robbins, & Maurer, 2010). If this trend continued, older adults would have the highest accuracy across all age groups because intuitively they have the most experience; however, older adults, over 60 years are generally worse at emotion perception compared to younger adults when tested between age groups, with conditions held constant (see meta-analysis by Ruffman, Henry, Livingstone, & Phillips, 2008). In one large study of a lifespan sample from early childhood to late adulthood, a curvilinear relationship between age and emotion perception accuracy emerged, with middle adults being the most accurate (Williams et al., 2009).

The effects of aging on emotion perception accuracy may be even more complex, though, because the age deficit in emotion perception accuracy of facial expressions varies somewhat in magnitude by emotion. Compared to younger adults, older adults have lower emotion perception accuracy for fearful and sad faces (Calder et al., 2003; Orgeta & Phillips, 2008; Wong, Cronin-Golomb, & Neargarder, 2005), sometimes for angry faces (Calder et al., 2003; Orgeta & Phillips, 2008), and sometimes for neutral faces (McDowell, Harrison, & Demaree, 1994). However, older adults tend to have equal (Orgeta & Phillips, 2008) or higher accuracy for faces of disgust (Calder et al., 2003; Wong et al., 2005). The findings for positive emotions suggest that older adults have slightly lower (Ruffman et al., 2008), equivalent (McDowell et al., 1994; Murphy & Isaacowitz, 2010; Orgeta & Phillips, 2008; Sullivan & Ruffman, 2004), or higher accuracy compared to younger adults (Murphy & Isaacowitz, 2010). Although emotion perception studies rarely use middle-aged adults, research suggests that fear accuracy decreases after age 40 (Williams et al., 2009). However, other negative emotions did not show this trend; furthermore, disgust showed maintenance from middle adulthood into older adulthood (Calder et al., 2003). Although the results are somewhat mixed, older adults tend to be less accurate at perceiving fearful, sad, angry, and neutral faces, but do not show the same deficit with happy or disgusted faces.

Even though the perception of static facial expressions is the most commonly studied modality, other modalities have been investigated, and findings suggest varied patterns. Compared to younger adults, older adults are less accurate when labeling emotional prosody of voices (Brosgole & Weisman, 1995) and bodily expressions (Ruffman, Sullivan, & Dittrich, 2009). This varies somewhat by emotion as well, but across all modalities (face, voice, body), older adults are less accurate at identifying anger and sadness. However, older adults do not seem to show a deficit when asked to distinguish positive and negative emotional valence in dynamic facial and bodily expressions (Krendl & Ambady, 2010). It appears that while a complex relationship exists, there are some forms of decline in emotion perception accuracy with age.

What are the reasons for the age-related deficit in emotion perception accuracy?

Studies investigating potential underlying mechanisms have focused primarily on age differences in neural processes, cognition, motivation, and attention. Several studies have suggested that age-related declines in emotion perception are caused by age-related changes in the brain (Ruffman et al., 2008). For example, when perceiving negative faces, older adults tend to have less activation in the left amygdala (Iidaka et al., 2002), an almond-shaped group of nuclei located deep within the temporal lobe suggested to be associated with emotion processing. Some researchers suggest the left amygdala shows an enhanced neural response to varying intensities of negative expressions, such as sadness (Blair, Morris, Frith, Perrett, & Dolan, 1999). Therefore, if the left amygdala is implicated in the perception of negative emotions, an age-related decrease in activation may result in decreased emotion perception accuracy. However, others have suggested caution when viewing these brain-related findings as participants may have had atypical cognitive decline or preclinical dementia (Burgmans et al., 2009).

General cognitive decline has been suggested as a possible underlying mechanism for age-related decline in emotion perception accuracy. Older adults have well-documented declines in various forms of cognitive functioning (e.g., Salthouse, 1996). For instance, older adults tend to have decreased mental processing speed, which may prohibit completion of cognitive operations, thus degrading cognitive performance. If older adults lack the mental processing speed required to accurately perceive emotions, their performance on perception tasks would suffer. In line with this hypothesis when researchers controlled for working memory, some age differences in emotion perception accuracy were no longer significant (MacPherson, Phillips, & Della Sala, 2002). However, the authors found that only tasks associated with the temporal lobe accounted for the age difference in the perception of sadness, suggesting that specific, rather than general, cognitive decline is responsible for an age difference in accuracy. In addition, the lack of association with accuracy of other emotions suggests that simple age differences in cognitive abilities cannot be said to fully underlie the age deficit in accuracy. Finally, other studies have not been able to statistically eliminate the age deficit by controlling for cognition (Sullivan & Ruffman, 2004; but see Suzuki & Akiyama, 2013). The mixed findings suggest that general cognitive decline may not completely explain the age deficit in emotion perception accuracy.

Other researchers have focused on age differences in motivation as a potential underlying mechanism. Socioemotional selectivity theory (SST) posits that the shortened time perspective of older adults leads them to prioritize emotion-based goals to optimize affect over information-based goals to learn new things, leading older adults to seek out more positive emotional experiences that may explain why adults attend to more positive compared to negative emotional information (e.g., Charles, Mather, & Carstensen, 2003). If older adults direct more attention to positive emotions, they may have more everyday experience with them, thus possibly preserving their emotion perception ability, or at the moment of emotion perception, they may divert their attention away from negative expressions. If older adults interact less with negative emotions compared to younger adults, they may have less "practice" perceiving such emotions. However, older adults may also have a deficit in perceiving positive emotions, albeit smaller than the deficit associated with most negative emotions (Ruffman et al., 2008). (For further discussion on motivation's possible role in accuracy, see Chapter 11).

Differences in gaze patterns, an indicator of internal mental processes, have also been posited as an underlying mechanism for the age deficit in emotion perception accuracy. If older adults attend to different parts of the stimuli (e.g., eyes or mouth), this may result in lower accuracy because different parts of the stimuli may be more indicative for certain emotional displays. For example, more gazing at the mouth has been associated with higher accuracy for certain emotions (Murphy & Isaacowitz, 2010). Furthermore, a link between gaze patterns (e.g., attending more to the mouth) and lower accuracy in older adults has been suggested (Wong et al., 2005). However, gaze patterns do not account fully for age difference in emotion perception (Murphy & Isaacowitz, 2010). So far, none of the previously studied mechanisms fully explain age differences in emotion perception. This leaves multiple lines of research available for future investigations, such as how the social environment influences emotion perception accuracy in older adults and how gaze patterns are associated with stimuli other than static faces.

Accuracy in social perception

Apart from being able to assess emotion expressions, judging other aspects of a person (or a group of people) based on their appearance is an important interpersonal skill. In this chapter, we refer to social perception as forming impressions of others outside of the realm of judging emotional expressions; we will specifically discuss accuracy in judgments of mind states (e.g., intentions) and of attributes (e.g., personality characteristics). Just as in emotion perception, individual characteristics of the perceiver, such as changes in experience and motivation due to age, guide social judgments, especially when minimal information (e.g., one-dimensional stimuli) is available in the given task. In order to discuss lifespan differences in accurate social perception, we will first briefly describe children's development of the skills necessary for social perception and the ways in which accuracy is measured. Then we will review the adult developmental literature and highlight key findings that describe changes in social perception accuracy with age.

Children

Understanding mind states. Like emotion perception, research in the child development literature on accuracy in social perception is often measured in terms of behavior and rests on two different types of methodology: (1) interaction studies that observe infants engaging with others and (2) looking-time paradigms that measure infants' attention and reaction to social stimuli. Using these paradigms, researchers have determined that infants have a good understanding of others' intentional actions by the end of their first year of life, meaning that they can appreciate the connection between the deliberate actions of an agent (e.g., picking up something) and the agent's subsequent experiences (e.g., feeling excited; Baird & Astington, 2005). This basic understanding is a necessary precursor to the development of more complex social awareness and perception.

One of the earliest developed aspects of social cognition is theory of mind (TOM), which is the ability to attribute independent mental states to others in order to explain and predict their behavior (Baron-Cohen, Leslie, & Frith, 1985). False belief understanding is one of the hallmarks of TOM, and it is most notably measured by the following task: children are given a scenario where a protagonist places an object in a location, but when the protagonist leaves the room the object is moved to a different location. Children must answer where they think the protagonist will look for the object when they return to the room (Wimmer & Perner, 1983). Considerable research suggests that TOM development begins as early as infancy, even though most children do not usually pass false-belief tasks (i.e., they think the protagonist will look for the object in its original location) until the age of four. For example, researchers have found that infant dishabituation (i.e., interest in novel events) significantly predicted later false-belief understanding in the same child at age four. This finding comes from measuring infant attention to social stimuli (e.g., a target looking at one of two objects with interest and joy, then selecting one of the objects), and still holds even after IQ, verbal competence, and executive function are controlled (Wellman, Lopez-Duran, LaBounty, & Hamilton, 2008).

By 12–15 months of age, children acquire the ability for joint attention, which is the coordination of attention between interacting partners and objects in the environment. Joint attention is one of the first measures of social awareness, and it is an essential prerequisite for accurate social perception because it paves the way for communication of desires and integration of multiple sources of attention (Bakeman & Adamson, 1984).

Judging attributes. The human face is a window to interpreting inner states, making face perception a crucial milestone early in life (Zebrowitz, 1997). Researchers using habituation paradigms (i.e., longer time spent looking indicates awareness of incongruence) have discovered that newborn infants demonstrate a visual preference for faces over non-face stimuli, their mother's face (Field, Cohen, Garcia, & Greenburg, 1984), and attractive faces (Slater et al., 1998). Using eye tracking, researchers have found that three-month-old infants are able to discern faces from complex scenes (such as movie clips; Frank, Vul, & Johnson, 2009) and even have a preference for own-race faces (Kelly et al., 2005). The development of these face perception abilities is essential for more complex and accurate understanding of social partners later in life.

While face perception research in infants mostly focuses on preferences, researchers have studied social perception in older children in terms of accurate judgments of personality characteristics. For example, 4–7-year-old children matched photographs of adults with dominant looking faces (e.g., lowered brow poses) with stories depicting socially dominant behaviors (e.g., telling others what to do; Keating & Bai, 1986), indicating that children use facial features as cues in determining a person's dominance. Using a similar paradigm, young (3–4-year-old) and older (4.5–6-year-old) children were asked to decide which of a pair of faces was best described by a story (e.g., "point to the person who looks like the leader"). The younger children were able to correctly categorize male but not female faces that were previously judged by adults as "babyish" or "mature," indicating that their impressions of warmth and dominance were influenced by facial characteristics (Montepare & Zebrowitz-McArthur, 1989). Furthermore, 3–10-year-old children were asked to make judgments of trustworthiness ("mean"/"nice"), dominance ("strong"/"not strong"), and competence ("smart"/"not smart") for faces that had been previously validated by adult perceivers. Three-year-old children gave accurate trustworthiness ratings, and children aged seven and older gave judgments that matched those of adults across all domains (Cogsdill, Todorov, Spelke, & Banaji, 2014). In addition, children aged 5–13 observed a pair of faces (which were the official photos of the winner and runner up of a real election) and, after playing a trip planning game, were asked to pick one of the faces to be the captain of their boat. Children's choices matched the results of adults who judged the same faces for competence, which was also predictive of the real election results (see Chapter 6 for further adult data on similar tasks). The results suggest that the leadership selection process between adults and children is not only extremely similar, but also based very heavily on facial features (Antonakis & Dalgas, 2009).

In another perception study, children watched brief videotaped vignettes that were meant to display a child actor's abilities or personality traits (e.g., a child sharing part of his lunch with someone who has nothing to eat) and were asked to make predictions about the actor's behavior in different situations. Younger children (ages 5–8) were able to accurately label the dispositional traits of the actors (e.g., the child was nice), but only the older subjects (ages 9–10 and 18–22) were able to additionally predict consistency of the actors' behaviors across situations, demonstrating the development of this ability with age (Rholes & Ruble, 1984). This paradigm is a good representation of the kinds of judgments made in everyday life. But because most research to date has used still photographs, more research is needed on interpersonal perception skill development in children using more realistic video-based dynamic stimuli. In addition, this paradigm was unique in using visual stimuli depicting child targets, while most of the other research in the literature uses adult faces validated by adult raters. Further research should also investigate potential methods for children's validation of stimuli.

Children's ability to judge attributes of others, especially of adults, is a skill that has real-life implications given the fact that children rely on adults for most of their basic physical and social needs. It is important for children to learn how to accurately judge characteristics that signal approach or avoidance, as well as to make predictions about future actions based on those judgments. The development of this childhood ability lays an important foundation for accurate social perception into adulthood and older age.

Adults

Theory of mind. Accurate perception of others' beliefs and intentions has implications for older adults' social well-being, because a deficiency in this area could lead to miscommunication and harm relationships. A recent meta-analysis found a moderate age difference in TOM overall, with older adults performing worse than younger adults across different types of tasks (e.g., false belief understanding), modalities (e.g., visual vs. verbal), and domains (e.g., cognitive vs. affective; Henry, Phillips, Ruffman, & Bailey 2013). However, a more in-depth look at one domain in particular reveals the benefits of investigating aspects of the judgment strategy (e.g., other skills and preferences) beyond accurate performance on the task. For example, researchers who studied faux pas judgment (i.e., an ability to recognize socially appropriate from inappropriate behavior) found that older adults had worse performance in detecting social gaffes, but poor performance in this task was fully mediated by poor performance in emotion perception (Halberstadt, Ruffman, Murray, Taumoepeau, & Ryan, 2011). In other words, age differences in accurate faux pas discrimination may be caused by age differences in emotion perception tasks, though the cross-sectional design of this study makes this conclusion merely speculative. Furthermore, differences in perceiving social gaffes (e.g., talking about someone while you don't realize they're listening) are related to differences in preferences (e.g., finding humor in these incidents; Stanley, Lohani, & Isaacowitz, 2014). Therefore, investigating characteristics of the perceiver (such as beliefs, preferences, and skills) may reveal how judgment strategies (in addition to accuracy) may differ by age.

Judging attributes. Just as face perception is important in childhood development, accurate perception of facial qualities has important social consequences for aging adults. As discussed earlier in the context of emotion perception, older adults have difficulty perceiving emotions from faces and even have difficulty processing information related to eye gaze (Slessor, Phillips, & Bull, 2008). However, face perception studies that investigate other facial characteristics beyond emotional expressions do not find this age deficit. For example, when younger and older adults judged health and competence from pictures of faces, perceivers' age did not affect the accuracy of these judgments, but the age of the target faces (a quality of the stimulus) did. Perceivers of both age groups were more accurate in judging competence from young faces and health from old faces. This suggests that judging health and competence is an adaptive skill that is preserved in older adulthood (Zebrowitz et al., 2014). Furthermore, when making trait impressions of competence, health, trustworthiness, and aggressiveness from faces, young and older adults have more within age group than between age group agreement of ratings. However, older adults' trait impressions of faces reflected greater positivity (Zebrowitz, Franklin, Hillman, & Boc, 2013). Older adults are also less likely to perceive dangerous qualities in faces that had been previously categorized as highly dangerous by both young and older adult independent raters (Ruffman, Sullivan, & Edge, 2006). This lends support to the idea that older adults prefer to process positive cues over negative cues, consistent with the idea of age-related positivity effects discussed above.

Furthermore, older adults had similar accuracy as younger adults in two first impression tasks based on rating photographs of faces: (1) participants judged faces of CEOs of the highest and lowest ranked US companies on their leadership ability and (2) participants categorized faces of young- and older-adult targets who had self-reported political affiliation as Democrats or Republicans. Importantly, measures of cognitive ability as well as the typically found age difference in emotion perception (i.e., older adults performed worse in judging emotion from faces) was unrelated to these first impression ratings, suggesting that mechanisms for social perception (in this case, judging leadership and social group membership) are unimpaired by aging (Krendl, Rule, & Ambady, 2014).

Similarly, in light of recent arguments questioning the ecological validity of traditional emotion perception tasks especially as applied to the study of aging and age differences (e.g., Isaacowitz & Stanley, 2011), researchers have attempted to create research paradigms that tap into the skills people use in their daily lives by using stimuli of unrehearsed, real-time conversations. These studies typically have young and older adults watch prerecorded interactions and make judgments about the targets in the videos. For example, there is a positive correlation between increasing age and accurate judgments of marital satisfaction from brief videotaped interactions between married couples (Ebling & Levenson, 2003). Moreover, older adults perform better than younger adults (with middle-aged adults falling in between) at judging the affective experiences of married couples discussing important martial topics; this paradigm matches participants' ratings (made continuously with a rating dial) to the actual feelings of the persons being judged because the targets had continuously rated their feelings during the interaction (Sze, Gyurak, Goodkind, & Levenson, 2012). Furthermore, older and younger adults had similar empathic accuracy, defined here as the cognitive understanding of what another person is feeling, when young- and older-adult targets described a topic relevant to older adults (i.e., loss of a loved one), but not when they talked about topics more relevant to younger adults (i.e., moving away to start a new life; Richter & Kunzmann, 2011). However, in both mixed age and same age dyadic interactions, older adults were only worse in empathic accuracy (defined here as judging what their partner was thinking and feeling), when their partners' thoughts and feelings were negative, but there were no age differences in making accurate judgments of positive thoughts and feelings (Blanke, Rauers, & Riediger, 2014). When asked to judge rapport, a dyadic quality emerging from interacting partners' mutual attention, positivity, and coordination, young and older adults have equivalent accuracy when rating videotaped interactions of young-adult targets who had previously provided self-reports of rapport). However, older adults had higher knowledge (i.e., discrimination between valid and invalid cues) and spent more time visually attending to targets' bodies than heads, suggesting that older adults used different sources of information in judging rapport, despite having the same accuracy (Vicaria, Bernieri, & Isaacowitz, 2015). These studies highlight the importance of investigating multiple components of the judgment process in addition to accuracy. Instead of using accuracy as the sole measurement of performance, when researchers take a closer look at the different ways in which older adults use their knowledge-, experience-, and motivation-guided preferences in the service of accurate social perception, they may uncover a better understanding of such phenomena.

Conceptual models for lifespan development of accurate social perception

Because preferential selectivity for certain facial characteristics (e.g., ethnicity) is present in three-month olds but not newborns, the development of perceptual accuracy appears to be learned as a result of exposure during early development (Kelly et al., 2005). For example, three- and four-month-old infants reared by a female primary caregiver had better facial recognition memory of female faces, and infants reared by male primary caregivers had a spontaneous preference for male faces (Slater & Quinn, 2001). Thus, the accurate social perception (at least of attributes depicted by faces) seems to be learned rather than innate.

Three critical components describe the prerequisites for children's accurate social perception; these include comprehension that (1) others' actions are caused by the others themselves, (2) others' actions are guided by intentions, and (3) others' behaviors occur in consistent and predictable ways (Secord & Peevers, 1974). The third component presupposes an understanding of stable dispositions, and is not usually obtained until about eight years of age (Roles & Ruble, 1984).

In general, there are surprisingly few instances of adult age differences in social perception, leading to an overall impression of age similarity, although there is much variation according to the topic and paradigm used (Freund & Isaacowitz, 2014). The idea of age-related positivity effects proposed by SST (Charles et al., 2003) may not fully account for the age differences found in emotion perception tasks (Ruffman, 2011), but it may still provide insight for tasks involving social perception. For example, despite lower overall accuracy in empathic judgments (i.e., identifying others' thoughts and feelings), older adults may have been paying more attention to positive than negative content (Blanke et al., 2014). Furthermore, older adults have more positive impressions of health, hostility, and trustworthiness than young- and middle-aged adults (Zebrowitz et al., 2013). Thus, older adults' decreased interest in negative content may affect their performance because they are biased toward having positive impressions of targets, regardless of the accuracy of those impressions.

In addition, the selective engagement hypothesis (Hess, Leclerc, Swaim, & Weatherbee, 2009) suggests that older people preferentially invest cognitive resources in tasks that hold more meaning and relevance to them. This hypothesis provides a potential explanation for the age-related decline in tasks that lack ecological validity: older adults may feel that the constrained and unrepresentative tasks (i.e., tasks that seem far removed from real-life situations) are less worthy of investing cognitive resources, thereby affecting their accuracy on those tasks.

Whereas age differences are evident in emotion perception tasks, which mostly rely on static stimuli, age similarities abound in tasks that integrate visual and auditory information in dynamic displays, which are typically found in social perception studies. It may be the case that ecologically valid tasks are especially important for older adults because they can draw on real-life experiences rather than waste effort in trying to interpret unrealistic stimuli. Thus, accurate performance in social perception tasks may be reflective of the more ecological approach used in these studies. Furthermore, studies that measure differences in attention and motivation help illustrate the different strategies older adults use in social perception, which in some cases demonstrates how judgment strategies differ by age despite similar performance on accuracy.

Conclusions: future directions, conceptual and empirical, for lifespan study of accuracy

In this chapter, we have reviewed a substantial body of research – both from the study of child development as well from the field of adult development and aging – suggesting age-related differences as well as continuities in emotion perception and social perception. In the child literature, the concern has been mostly with documenting the ages at which various abilities emerge.

In the adult literature, there is strong (albeit methodologically constrained) evidence suggesting an age-related deficit in the ability to accurately perceive emotional expressions, at least from static faces. The evidence is more mixed for dynamic faces. In terms of social perception, we can draw from the title of a recent special section of a journal on aging and social perception: "so far, more similarity than differences." One might reasonably conclude that there is an age-related decrement in emotion perception accuracy but *not* in social perception accuracy.

There are, however, a number of important caveats to this work, which also form suggestions for future needed directions. First, all of the adult development literature reviewed (and much of the child development work) is cross-sectional in nature. This has the usual methodological problems about confounding age and cohort and not permitting conclusions about true developmental change within individuals. It is quite possible that interpersonal accuracy is one ability that is influenced by cohort, to the extent that cohorts may have had different access to information and/or feedback that might train their accuracy, or have had different family experiences that might predict accuracy (as was found in A. G. Halberstadt's work described above). Moreover, without true longitudinal data that consider the development of accuracy (rather than correlates of accuracy) within persons, not only can we not rule out cohort effects, we also cannot determine whether similarity across age groups is the result of a skill that does not change with age, or a cohort that started or became even better at the process but then eventually got worse. Therefore, this is an area where cross-sectional studies leave many open questions, and yet it is also an area where, due to its relative youth as a field, there is only cross-sectional data available and none of the existing large longitudinal datasets – either long term (e.g., Grant Study: Vaillant, 2012) or short term (e.g., MIDUS: Brim, Ryff, & Kessler, 2004) contain interpersonal accuracy measures.

Second, there is a disconnect between the child and the adult literatures, with different constructs and different methods used in the two areas. Even though it is obviously the case that the same individuals move through childhood and adulthood, the research traditions are fairly discontinuous and so it is challenging even conceptually to try to imagine lifespan development of accuracy as opposed to child development of it on one hand, and adult development of it on the other. Stimuli, response options, and empirical questions tend to differ between the two fields. Variability in accuracy tends to be conceptualized as the age at which a threshold is reached in children, but simple differences in mean performance in adults and older adults. The underlying causal mechanisms investigated are also nonoverlapping: there does not appear to be any research to date on individual differences in social environment in middle or late adulthood, and no work testing hypotheses regarding possible motivational processes like positivity in children. To be sure, there are different developmental patterns to be explained: how accuracy increases in childhood, and whether and how it declines or remains stable (despite cognitive changes) in adulthood and old age.

Despite differences between child and adult developmental approaches to accuracy, it is still possible that a unified lifespan field could develop, taking advantage of a full range of creative and ecologically valid stimuli, which would help researchers map out the when, how, and why of lifespan continuity and change in both emotion perception and social perception. This is a topic with great opportunity for the development of truly lifespan mechanistic models – but that remains for future conceptual and empirical work.

References

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

Baird, J., & Astington, J. (2005). The development of the intention concept: From the observable world to the unobservable mind. In R. Hassin, J. Uleman, & J. Bargh (Eds.), *The new unconscious* (pp. 256–276). New York: Oxford University Press.

Bakeman, R., & Adamson, L. B. (1984). Coordinating attention to people and objects in mother infant and peer-infant interaction. *Child Development*, 55, 1278–1289.

Banks, M. S., & Ginsburg, A. P. (1985). Infant visual preferences: A review and new theoretical treatment. In H. W. Reese (Ed.), *Advances in child development and behavior* (pp. 207–246). New York: Academic Press.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, 21, 37–46.

Barrera, M. E., & Maurer, D. (1981). The perception of facial expressions by three-month-olds. *Child Development*, 5, 203–206.

Baum, K., & Nowicki, S., Jr. (1998). Perception of emotion: Measuring decoding accuracy of adult prosodic cues varying in intensity. *Journal of Nonverbal Behavior*, 22, 89–108.

Bretherton, L., & Beeghly, M. (1982). Talking about internal states: The acquisition of an explicit theory of mind. *Developmental Psychology*, 19, 906–921.

Blair, R. J., Morris, J. S., Frith, C. D., Perrett, D. I., & Dolan, R. J. (1999). Dissociable neural responses to facial expressions of sadness and anger. *Brain*, 122, 883–893.

Blanke, E. S., Rauers, A., & Riediger, M. (2014). Nice to meet you: Adult age differences in empathic accuracy for strangers. *Psychology and Aging*, 30, 149–159.

Brim, O. G., Ryff, C. D., & Kessler, R. C. (Eds.) (2004). *How healthy are we?: A national study of well-being at midlife*. Chicago, IL: The University of Chicago Press.

Brosgole, L., & Weisman, J. (1995). Mood recognition across the ages. *International Journal of Neuroscience*, 82, 169–189.

Burgmans, S., van Boxtel, M. P. J., Vuurman, F. P. M., Smeets, F., Gronenschild, E. H. B. M., Uylings, H. B. M., & Jolles, J. (2009). The prevalence of cortical gray matter atrophy may be overestimated in the healthy aging brain. *Neuropsychology*, 23, 541–550.

Calder, A. J., Keane, J., Manly, T., Sprengelmeyer, R., Scott, S., Nimmo-Smith, I., & Young, A. W. (2003). Facial expression recognition across the adult life span. *Neuropsychologia*, 41, 195–202.

Charles, S. T., Mather, M., & Carstensen, L. L. (2003). Aging and emotional memory: The forgettable nature of negative images for older adults. *Journal of Experimental Psychology: General*, 132, 310–324.

Cogsdill, E. J., Todorov, A. T., Spelke, E. S., & Banaji, M. R. (2014). Inferring character from faces: A developmental study. *Psychological Science*, 25, 1132–1139.

Collins, M. (1996). Personality and achievement correlates of nonverbal processing ability in African American children. Unpublished doctoral dissertation, Department of Psychology, Emory University.

Demertizis, A., & Nowicki, S., Jr. (1997, April). *Perception of emotion: Measuring decoding accuracy of child prosodic cues varying in emotion intensity*. Paper presented at meeting of the Southeastern Psychological Association, Mobile, AL.

Ebling, R., & Levenson, R. W. (2003). Who are the marital experts? *Journal of Marriage and Family*, 65, 130–142.

Field, T. M., Cohen, D., Garcia, R., & Greenberg, R. (1984). Mother-stranger face discrimination by the newborn. *Infant Behavior and Development*, 7, 19–25.

Field, T. M., Woodson, R. W., Greenberg, R., & Cohen, C. (1982). Discrimination and imitation of facial expressions by neonates. *Science*, 218, 179–181.

Frank, M. C., Vul, E., & Johnson, S. P. (2009). Development of infants' attention to faces during the first year. *Cognition*, 110, 160–170.

Freund, A. M., & Isaacowitz, D. M. (2014). Aging and social perception: So far, more similarities than differences. *Psychology and Aging*, 29, 451–453.

Halberstadt, A. G., Crisp, V. W., & Eaton, K. L. (1999). Family expressiveness: A retrospective and new directions for research. In P. Philippot, R. S. Feldman, & E. J. Coats (Eds.), *The social context of nonverbal behavior* (pp. 109–155). New York: Cambridge University Press.

Halberstadt, A. G., & Eaton, K. L. (2002). A meta-analysis of family expressiveness and children's emotion expressiveness and understanding. *Marriage and Family Review*, 34, 35–62.

Halberstadt, J., Ruffman, T., Murray, J., Taumoepeau, M., & Ryan, M. (2011). Emotion perception explains age-related differences in the perception of social gaffes. *Psychology and Aging*, 26, 133–136.

Henry, J. D., Phillips, L. H., Ruffman, T., & Bailey, P. E. (2013). A meta-analytic review of age differences in theory of mind. *Psychology and Aging*, 28, 826–839.

Herrenkohl, R., Herrenkohl, E., Egolf, B., & Wu, P. (1991). The developmental consequences of child abuse. In R. Starr & D. Wolfe (Eds.), *The effects of child abuse and neglect* (pp. 57–81). New York: Guilford.Press.

Hess, T. M., Leclerc, C. M., Swaim, E., & Weatherbee, S. R. (2009). Aging and everyday judgments: The impact of motivational and processing resource factors. *Psychology and Aging*, 24, 735–740.

Iidaka, T., Okada, T., Murata, T., Omori, M., Kosaka, H., Sadato, N., & Yonekura, Y. (2002). Age-related differences in the medial temporal lobe responses to emotional faces as revealed by fMRI. *Hippocampus*, 12, 352–362.

Isaacowitz, D. M., & Stanley, J. T. (2011). Bringing an ecological perspective to the study of aging and recognition of emotional facial expressions: Past, current, and future methods. *Journal of Nonverbal Behavior*, 35, 261–278.

Keating, C. F., & Bai, D. L. (1986). Children's attributions of social dominance from facial cues. *Child Development*, 57, 1269–1276.

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Gibson, A., Smith, M., . . . & Pascalis, O. (2005). Three-month-olds, but not newborns, prefer own-race faces. *Developmental Science*, 8, F31–F36.

Krendl, A. C., & Ambady, N. (2010). Older adults' decoding of emotions: Role of dynamic versus static cues and age-related cognitive decline. *Psychology and Aging*, 25, 788–793.

Krendl, A. C., Rule, N. O., & Ambady, N. (2014). Does aging impair first impression accuracy? Differentiating emotion recognition from complex social inferences. *Psychology and Aging*, 29, 482–490.

LaBarbera, V. D., Izard, C. E., Vietze, P., & Parisi, S. A. (1976). Four- and six-month-old infants' visual responses to joy, anger, and neutral expressions. *Child Development*, 47, 535–538.

Ludemann, P. M. (1991). Generalized discrimination of positive facial expressions by seven- and ten-month-old infants. *Child Development*, 62, 55–67.

Ludemann, P. M., & Nelson, C. A. (1988). The categorical representation of facial expressions by 7-month-old infants. *Developmental Psychology*, 24, 492–501.

MacPherson, S. E., Phillips, L. H., & Della Sala, S. (2002). Age, executive function and social decision making: A dorsolateral prefrontal theory of cognitive aging. *Psychology and Aging*, 17, 598–609.

McClanahan, P. (1996). Social competence correlates of children who are 7 and 8 years of age. Unpublished masters thesis, Department of Psychology, Emory University.

McDowell, C. L., Harrison, D. W., & Demaree, H. A. (1994). Is right hemisphere decline in the perception of emotion a function of aging? *International Journal of Neuroscience*, 79, 1–11.

Mondloch, C. J., Robbins, R., & Maurer, D. (2010). Discrimination of facial features by adults, 10-year-olds and cataract-reversal patients. *Perception*, 39, 184–194.

Montepare, J. M., & Zebrowitz-McArthur, L. (1989). Children's perceptions of babyfaced adults. *Perceptual and Motor Skills*, 69, 467–472.

Murphy, N. A., & Isaacowitz, D. M. (2010). Age effects and gaze patterns in recognising emotional expressions: An in-depth look at gaze measures and covariates. *Cognition and Emotion*, 24, 436–452.

Nelson, C. A., & Dolgin, K. (1985). The generalized discrimination of facial expressions by 7-month-old infants. *Child Development*, 56, 58–61.

Nelson, C. A., Morse, P. A., & Leavitt, L. A. (1979). Recognition of facial expressions by 7-month-old infants. *Child Development*, 50, 1239–1242.

Nowicki, S., Jr., & Carton, E. (1997). The relation of nonverbal processing ability of faces and voices and children's feelings of depression and competence. *Journal of Genetic Psychology*, 158, 357–364.

Nowicki, S., Jr., & Duke, M. (2001). Nonverbal receptivity: The diagnostic analysis of nonverbal accuracy (DANVA). In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers, Psychology Press.

Orgeta, V., & Phillips, L. H. (2008). Effects of age and emotional intensity on the recognition of facial emotion. *Experimental Aging Research*, 34, 63–79.

Pollak, S. D., & Sinha, P. (2002). Effects of early experience on children's recognition of facial displays of emotion. *Developmental Psychology*, 38, 784–791.

Rholes, W. S., & Ruble, D. N. (1984). Children's understanding of dispositional characteristics of others. *Child Development*, 55, 550–560.

Richter, D., & Kunzmann, U. (2011). Age differences in three facets of empathy: Performance based evidence. *Psychology and Aging*, 26, 60–70.

Roberts, V. J., Nowicki, S., Jr., & McClure, E. (1998, February). *Emotional prosody recognition and right hemisphere functioning in the elderly*. Paper presented at the 26th Annual Meeting of the International Neuropsychological Society, Honolulu, HI.

Ruffman, T. (2011). Ecological validity and age-related change in emotion recognition. *Journal of Nonverbal Behavior*, 35, 297–304.

Ruffman, T., Henry, J. D., Livingstone, V., & Phillips, L. H. (2008). A meta-analytic review of emotion recognition and aging: Implications for neuropsychological models of aging. *Neuroscience & Biobehavioral Reviews*, 32, 863–881.

Ruffman, T., Sullivan, S., & Edge, N. (2006). Differences in the way older and younger adults rate threat in faces but not situations. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 61, 187–194.

Ruffman, T., Sullivan, S., & Dittrich, W. (2009). Older adults' recognition of bodily and auditory expressions of emotion. *Psychology and Aging*, 24, 614–622.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, 103, 403–428.

Schwartz, G. M., Izard, C. E., & Ansul, S. E. (1985). The 5-month-old's ability to discriminate facial expressions of emotion. *Infant Behavior and Development*, 8, 65–77.

Secord, P., & Peevers, B. (1974). The development and attribution of person concepts. In T. Mischel (Ed.), *Understanding other persons* (pp. 117–142). Oxford: Blackwell.

Serrano, J. M., Iglesias, J., & Leoches, A. (1992). Visual discrimination and recognition of facial expressions of anger, fear, and surprise in four- and six-month-old infants. *Developmental Psychobiology*, 25, 411–425.

Slater, A., & Quinn, P. C. (2001). Face recognition in the newborn infant. *Infant and Child Development*, 10, 21–24.

Slater, A., Von der Schulenburg, C., Brown, E., Badenoch, M., Butterworth, G., Parsons, S., & Samuels, C. (1998). Newborn infants prefer attractive faces. *Infant Behavior and Development*, 21, 345–354.

Slessor, G., Phillips, L. H., & Bull, R. (2008). Age-related declines in basic social perception: Evidence from tasks assessing eye-gaze processing. *Psychology and Aging*, 23, 812–822.

Stanley, J. T., Lohani, M., & Isaacowitz, D. M. (2014). Age-related differences in judgments of inappropriate behavior are related to humor style preferences. *Psychology and Aging*, 29, 528–541.

Sullivan, S., & Ruffman, T. (2004). Emotion recognition deficits in the elderly. *International Journal of Neuroscience*, 114, 94–102.

Suzuki, A., & Akiyama, H. (2013). Cognitive aging explains age-related differences in face-based recognition of basic emotions except for anger and disgust. *Aging, Neuropsychology, and Cognition: A Journal on Normal and Dysfunctional Development*, 20, 253–270.

Sze, J. A., Gyurak, A., Goodkind, M. S., & Levenson, R. W. (2012). Greater emotional empathy and prosocial behavior in late life. *Emotion*, 12, 1129–1140.

Vaillant, G. E. (2012). *Triumphs of experience*. Cambridge: Belknap Press.

Verbeek, P. (1996). Peacemaking in young children. Unpublished doctoral dissertation, Department of Psychology, Emory University.

Vicaria, I. M., Bernieri, F. J., & Isaacowitz, D. M. (2015). Perceptions of rapport across the life span: Gaze patterns and judgment accuracy. *Psychology and Aging*, 30, 396–406.

Walker-Andrews, A. S. (1997). Infants' perception of expressive behaviors: Differentiation of multimodal information. *Psychological Bulletin*, 121, 437–456.

Wellman, H. M., Lopez-Duran, S., LaBounty, J., & Hamilton, B. (2008). Infant attention to intentional action predicts preschool theory of mind. *Developmental Psychology*, 44, 618–623.

Widen, S. C., & Russell, J. A. (2003). A closer look at preschoolers' freely produced labels for facial expressions. *Developmental Psychology*, 39, 114–128.

Widen, S. C., & Russell, J. A. (2010). Differentiation in preschoolers' categories for emotion. *Emotion*, 10, 651–661.

Widen, S. C., & Russell, J. A. (2013). Children's recognition of disgust in others. *Psychological Bulletin*, 139, 271–299.

Williams, L. M., Mathersul, D., Palmer, D. M., Gur, R. C., Gur, R. E., & Gordon, E. (2009). Explicit identification and implicit recognition of facial emotions: 1. Age effects in males and females across 10 decades. *Journal of Clinical and Experimental Neuropsychology*, 31, 257–277.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13, 103–128.

Wong, B., Cronin-Golomb, A., & Neargarder, S. (2005). Patterns of visual scanning as predictors of emotion identification in normal aging. *Neuropsychology*, 19, 739–749.

Young-Browne, G., Rosenfeld, H. M., & Horowitz, F. D. (1978). Infant discrimination of facial expression. *Child Development*, 48, 555–562.

Zebrowitz, L. A. (1997). *Reading faces: Window to the soul?* Boulder, CO: Westview Press.

Zebrowitz, L. A., Franklin, R. G., Jr., Boshyan, J., Luevano, V., Agrigoroaei, S., Milosavljevic, B., & Lachman, M. E. (2014). Older and younger adults' accuracy in discerning health and competence in older and younger faces. *Psychology and Aging*, 29, 454–468.

Zebrowitz, L. A., Franklin, R. G., Jr., Hillman, S., & Boc, H. (2013). Older and younger adults' first impressions from faces: Similar in agreement but different in positivity. *Psychology and Aging*, 28, 202–212.

11 Situational influences on interpersonal accuracy

Petra C. Schmid

Abstract
Forming accurate impressions about others is an important skill that has been associated with better mental health and more successful social interactions. This interpersonal skill varies between people and is also influenced by situational factors. This chapter summarizes the literature on the situational influences that may determine a person's accuracy when forming impressions about others. The first section (motivational factors) focuses on the impact of motivation, as manipulated with task instructions and framing, incentivized performance, and priming of social power and general behavioral tendencies. The second section (affective factors) covers studies on the impact of positive and negative affect on interpersonal accuracy. In the third section (cognitive factors), the influence of deliberate versus automatic processing and local versus global processing on accurate impression formation is discussed. Finally, a fourth section (motor factors) includes research on how mimicry influences accuracy. After reviewing this literature, findings are integrated and general conclusions are drawn. Overall, empirical findings are heterogeneous and seem to depend on both characteristics of the impression formation tasks and specifics of the manipulations.

Situational influences

People constantly form impressions about others. These impressions, accurate or not, may have significant impact, such as in hiring decisions, juridical decisions, and romantic decisions. Forming accurate impressions about others (i.e., being interpersonally accurate) is not only relevant in decision-making contexts, it may also prevent social faux-pas and facilitate social functioning in general (Addington, Saeedi, & Addington, 2006; Bernieri, 2001; Keltner & Kring, 1998). Given the significance of interpersonal accuracy in social contexts, its antecedents have been investigated. Findings suggest that a decoder's ability to judge others correctly is determined by relatively stable factors such as his/her personality traits and gender, as well as by situational factors, such as his/her motivational, cognitive, and affective states. The focus of this chapter lies in research including experimental manipulations of situational factors and their effect on interpersonal accuracy. The influences of relatively stable factors are only considered when they are relevant in conjunction with the manipulated situational factor (e.g., sex is considered in the context of gender-relevant framing effects on interpersonal accuracy tasks, and race in the context of race-dependent memory for others). In this chapter, interpersonal accuracy is defined as the correct decoding of other people's interpersonal signals as well as the accurate recall of other people's behavior and attributes.

Motivational factors

It has often been proposed that motivation increases interpersonal accuracy. For example, the frequently found recognition advantage for same-race faces as opposed to other-race faces is assumed to occur partly because people lack motivation to individuate other-race targets (Young, Hugenberg, Bernstein, & Sacco, 2012). Moreover, it has been posited that social power makes people feel independent from others, which may reduce their motivation to judge others accurately (e.g., Keltner, Gruenfeld, & Anderson, 2003). However, studies that have directly examined the role of motivation in interpersonal accuracy have yielded inconsistent results. The following sections provide an overview of the research on how motivation, manipulated with task instructions (i.e., try-hard and attention-guiding instructions), task framings (i.e., as personally relevant), incentives (i.e., money and affiliation), and priming of social power and general behavioral tendencies (i.e., promotion vs. prevention foci) affect a person's accuracy in judging other people.

Try hard and attention-guiding instructions

The simplest way to encourage people to put more effort into a task is probably by instructing them to try hard or by telling them that performance on the test is important. Indeed, when participants were instructed to try hard, they were better at inferring a target person's personality traits (Biesanz & Human, 2010). In two further studies, participants watched a video of a target person being interviewed for a job and were then tested on how well they could recall the verbal and nonverbal cues of the target person in the video. In one of the two studies, participants who were instructed to try hard were marginally better at recalling the verbal content of this video; the recall of nonverbal behavior was not significantly affected by the try-hard instruction in both studies (Hall et al., 2009). Using similar recall-based interpersonal accuracy paradigms, it was further tested whether accuracy would be increased when participants were forewarned that their memory for the content of this video would be tested. With this forewarning it was attempted to make participants pay more attention to the target person. But neither of these studies produced a significant result (Hall et al., 2009). However, a meta-analysis over the two try-hard instruction and the two forewarning studies revealed a significant but small beneficial effect of instructions on recall for *verbal* cues in the previously seen video (Hall et al., 2009). In other work it was attempted to reduce participants' motivation by telling them that their responses would *not* be evaluated (McLarney-Vesotski, Bernieri, & Rempala, 2011). As expected, when receiving such information, participants were less accurate in interpreting the various social cues (such as guessing whether a target person talks to a man or woman on the phone) in the Interpersonal Perception Task (IPT; Costanzo & Archer, 1989).

Thus, overall research suggests that try-hard and attention-guiding instructions (i.e., forewarning) may be somewhat helpful for accurate interpersonal inference and recall for verbal information. On the other hand, reducing motivation by telling participants that being accurate is not important may decrease their performance.

Personal relevance

Try-hard and attention-guiding instructions may not be the strongest motivators because they do not stress how performing well on the task might be *personally* relevant for perceivers. In an attempt to increase personal motivation to perform well, interpersonal accuracy tasks have been introduced to participants as a measure of cognitive ability supposing that intelligence is a central aspect of the self, and thus personally relevant. The cognitive ability framing had no effects on people's accuracy at recalling a target person's appearance and at inferring a target person's thoughts, personality traits, and status (Hall et al., 2009; Klein & Hodges, 2001); it even had detrimental effects on lie detection (Forrest & Feldman, 2000).

In other studies, interpersonal accuracy tasks were framed in gender-relevant ways assuming that female-relevant framing would increase women's performance, and male-relevant framing would increase men's performance. When framed as "female-relevant," the interpersonal accuracy task was described as a task in which women typically outperform men or as a measure of empathy and interpersonal and conversational skills. When framed as "male-relevant," it was highlighted that the interpersonal accuracy task would measure skills related to leadership, military, or a competitive setting. Such gender-relevant framing boosted judgment accuracy in some studies. For instance, "female-relevant" framing increased women's but not men's accuracy in inferring a person's thoughts and feelings as compared to the control group (Klein & Hodges, 2001), at least when targets were relatively easy to read (Thomas & Maio, 2008). Men also tended to be more accurate on the IPT when it was framed "male-relevant" as opposed to "female-relevant," whereas "male-relevant" framing tended to decrease women's accuracy compared to the control condition. The authors argued that gender-inconsistent framings might evoke a context of negative performance expectations that in turn hurts performance. In further support of this argument, "female-relevant" framing hampered men's performance on the IPT, and this effect occurred because men processed information more deliberately (Koenig & Eagly, 2005). In this study, "female-relevant" framing did not affect women's performance. Effects of gender-relevant framing have not always been replicated; several studies (Hall et al., 2009; Hall & Schmid Mast, 2008) found no effects on the IPT, inferring status and dominance, or judging the various nonverbal cues in the Profile of Nonverbal Sensitivity (PONS; Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979). "Female-relevant" framing even decreased women's accuracy in recalling target people's appearance in one study (Hall et al., 2009).

To summarize, there is no empirical support to date for the idea that framing the task as a measure of cognitive ability would increase interpersonal accuracy. Findings on the impact of gender-relevant task framings were heterogeneous, suggesting that effects may depend on specific aspects of the interpersonal accuracy task and/or the exact wording of the framings. Thus, more research is needed to clarify under which conditions such effects occur.

Monetary incentives

Monetary incentives are known to enhance performance on a diverse range of effort-sensitive tasks (Camerer & Hogarth, 1999). Interpersonal accuracy, however, seems to be largely unaffected by such incentives as suggested by several studies (Hall et al., 2009; Nowicki & Richman, 1985) that included different interpersonal accuracy tasks (i.e., the PONS, the IPT, status and dominance inference, and emotion recognition) and varying amounts of rewards (small amounts or up to $100) and reward expectancy (for performance above average or only for the top scoring participant). Providing monetary incentives even hindered accuracy in a lie detection task (Porter, McCabe, Woodworth, & Peace, 2007). Only one study found that a monetary incentive (for each correct response, up to $8 in total) boosted participants' accuracy, specifically in reading another person's thoughts and feelings (Klein & Hodges, 2001).

Affiliation motivation

The motivation to affiliate is one of the predominant social goals in humans (A. P. Fiske, 1992). It is often assumed that affiliation incentives increase a person's motivation to form accurate judgments; however, it is also likely that it changes the way a person feels and behaves vis-à-vis the target, which in turn might have consequences for accurate impression formation (see Hall, 2014 for more detailed argumentation). Although the exact mechanisms are unclear, research has shown that affiliation motivation influences accuracy. For example, priming heterosexual women with mating goals (vs. control) led to more accurate detection of men's but not of women's sexual orientation (Rule, Rosen, Slepian, & Ambady, 2011). In the same vein, telling male participants that being empathic would help them attract women increased their accuracy in inferring thoughts and feelings (Thomas & Maio, 2008). Further support for the claim that affiliation motivation matters in interpersonal accuracy comes from the literature on social exclusion/inclusion. Individuals whose sense of belonging was threatened (due to a rejection and exclusion manipulation) were better at discriminating between Duchenne smiles and non-Duchenne smiles than both controls and individuals in the inclusion condition (Bernstein, Young, Brown, Sacco, & Claypool, 2008). In fact, after having experienced social rejection by a male in-group member, men showed improved performance on the IPT with "male-relevant" framing as compared to those with "female-relevant" framing or in the no-rejection control condition (Smith & Lewis, 2009). Thus, in support of the argument that affiliation is a very strong motivator in humans, research has consistently demonstrated that the motivation to affiliate boosts interpersonal accuracy.

Social power

Social power has been proposed to influence a person's willingness to form accurate judgments; however, there is an ongoing debate about whether social power should increase or decrease interpersonal accuracy.

It has been argued that high-power individuals are independent and may not therefore need to pay attention to others, whereas low-power individuals depend on other people and thus may be motivated to form accurate impressions (e.g., S. T. Fiske, 1993; Keltner et al., 2003). This suggests that high power decreases interpersonal accuracy compared to low power. Alternatively, it has been postulated that high power increases interpersonal accuracy, potentially indirectly, for example, via its effect on affect and use of cognitive strategies (e.g., Schmid Mast, Jonas, & Hall, 2009). Empirical findings are mixed and to date the direction of the power effect and potential moderating factors are still unclear (see Chapter 13 and also Hall, Schmid Mast, & Latu, 2015 for a recent meta-analysis).

General behavioral tendencies

The previous sections focused on manipulations that are intended and/or assumed to affect the motivation to be accurate. There is the possibility that general motivational states such as behavioral tendencies also relate to interpersonal accuracy. However, empirical research is scarce and it is unclear whether behavioral tendencies should have a direct or indirect effect on accuracy (e.g., by affecting accuracy motivation or information processing). Only one study was found (Sassenrath, Sassenberg, Ray, Scheiter, & Jarodzka, 2014), which demonstrated that induced promotion focus (characterized by a need for growth and a drive to pursue goals) increased emotion recognition accuracy compared to induced prevention focus (defined by a need for security and fulfilling responsibilities). According to the authors, promotion-focused individuals avoid errors of omission, whereas prevention-focused individuals avoid errors of commission, which may affect information processing and by extension, interpersonal accuracy. Specifically, promotion focus may lead to a search of opportunities in the environment, resulting in frequent shifts in attention allocation, and prevention focus may motivate more focused attention and a deliberate examination of each element. Correlational data were consistent with this reasoning such that greater trait promotion focus was associated with shorter fixation times on face stimuli, which in turn correlated positively with emotion recognition accuracy (Sassenrath et al., 2014). Thus, Sassenrath et al. provided initial evidence that behavioral tendencies may play a role in interpersonal accuracy. But clearly, more experimental research is needed to gain a comprehensive picture of how different general behavioral tendencies affect interpersonal accuracy as well as the underlying processes.

Motivational factors: summary and discussion

A relatively large set of studies has addressed the question of whether motivation helps interpersonal accuracy. Results display great inconsistency, which suggests that moderators are at work. For instance, it has been speculated that motivation might result in effortful (i.e., deliberate) processing, facilitating the processing of verbal cues, whereas the processing of nonverbal cues is thought to occur relatively effortlessly (i.e., automatically) (e.g., Ambady & Gray, 2002; Forrest & Feldman, 2000; Tracy & Robins, 2008, see also Hall et al., 2009). Consistent with this reasoning, studies reporting nonsignificant effects of motivation were typically based on nonverbal decoding tasks (e.g., Hall et al., 2009), while studies demonstrating positive effects of motivation on accuracy typically included verbal cues (e.g., Biesanz & Human, 2010; Hall et al., 2009; Klein & Hodges, 2001; Thomas & Maio, 2008). Note, however, that in a lie detection task, motivation increased focus on *nonverbal* cues, which was associated with impeded lie detection (Porter et al., 2007), suggesting that there are additional moderators at work in the links between motivation, information processing, and interpersonal accuracy. One such moderator might be task difficulty. Motivation might be most helpful for relatively easy tasks (i.e., when the perceiver has knowledge about the cues that are diagnostic for accurate judgments; see Hall, 2014 for more detailed argumentation). Consistent with this proposal, Thomas and Maio (2008) found that motivation improved individuals' accuracy in inferring thoughts and feelings of easy-to-read targets but not of hard-to-read targets.

To account for the mixed results, not only task characteristics but also the source of motivation should be considered. Hall (2014) proposed that manipulations intended to increase interpersonal accuracy might also alter various other things such as attention allocation, cognitive load, and worries about performing badly, which in turn might have consequences for accuracy. Thus, clearly, more research is needed that includes the assessment of the processes and mechanisms that might underlie the effect of motivation manipulations on judgment accuracy. It is also important to note that the studies reviewed in this chapter focused on motivational influences at the time the accuracy was measured. These findings do not speak to the question of how motivation contributes to the *acquisition* of the skill. For example, women likely outperform men on interpersonal accuracy tasks (e.g., Hall, 1978, 1984) because they have been socialized to fulfill their gender-stereotypical role that includes caring about others' internal states (e.g., Ickes, Gesn, & Graham, 2000; Snodgrass, 1985). As such, gender might be a long-term motivational influence but not a situational factor (see also Hall, 2014).

Affective factors

Decoding other people often involves making inferences about other people's affective states; however, the decoders may not always encounter a social situation in a neutral state themselves. The following sections review how decoders' affect may influence how accurately they read others and how affect might bias judgments in mood-congruent ways.

Negative affect

Research generally suggests that negative affect hampers interpersonal accuracy on nonverbal tasks and increases accuracy on verbal tasks. For instance, induced sadness and stress both decreased individuals' performance on an emotion recognition paradigm compared to a neutral state (Chepenik, Cornew, & Farah, 2007; Hanggi, 2004). Induced sadness also impaired individuals' accuracy when rating teacher effectiveness (with students' end of semester ratings as criteria) and when assessing the relationship type of dyads from silent videos (Ambady & Gray, 2002; but see Sinclair, 1988). Inconsistent results were found for face recognition tasks. In three studies, the influence of negative affect on the recognition of in-group and out-group faces was examined. Two of these studies produced no effects of negative affect (i.e., fear) versus neutral affect on face recognition accuracy (Johnson & Fredrickson, 2005), and one study showed that negative affect (unpleasant mood) increased face recognition accuracy; target group membership did not moderate these effects. Whereas negative affect generally decreased accuracy on nonverbal inference tasks (except for face recognition), the opposite was found for lie detection for which verbal information may be more diagnostic (Forgas & East, 2008; Reinhard & Schwarz, 2012). Indeed, mediation analyses showed that the effect of sad mood on lie detection accuracy was due to greater self-reported use of verbal versus nonverbal cues (Reinhard & Schwarz, 2012).

There is some evidence that sadness may influence interpersonal accuracy by triggering a local and deliberate processing style. For instance, participants induced with sadness processed facial information more locally (i.e., they focused more on the face features) and less globally (i.e., they focused less on interrelations between different face features) than participants induced with happiness/amusement, which in turn was associated with decreased emotion recognition accuracy (Schmid, Schmid Mast, Bombari, Mast, & Lobmaier, 2011a). Moreover, when sad participants were put under cognitive load, which presumably prevented them from using their usual deliberate processing style, the detrimental effect of sadness on relationship type assessments vanished (Ambady & Gray, 2002).

To summarize, negative affect generally seems to decrease interpersonal accuracy on nonverbal inference tasks and increase lie detection accuracy, possibly because lies are better detected when focusing on verbal cues (e.g., Forgas & East, 2008; Reinhard & Schwarz, 2012). Future research is needed to investigate whether negative affect boosts accuracy in detecting lies specifically, or whether this generalizes to other interpersonal accuracy tasks that are mainly based on verbal content.

Positive affect

Positive affect tends to provoke the opposite findings of negative affect; however, results are somewhat weaker, potentially because positive mood inductions are less effective than negative mood inductions (e.g., Westermann, Spies, Stahl, & Hesse, 1996). For instance, there is some evidence that happiness might increase accuracy on nonverbal decoding tasks. Induced joy enhanced recognition accuracy for other-race faces but did not affect recognition for same-race faces (Johnson & Fredrickson, 2005). The authors speculated that joy led to global/holistic processing, which is favorable for face recognition, and because same-race faces are typically processed more globally/holistically than other-race faces, there might have been a ceiling effect for same-race faces (Johnson & Fredrickson, 2005). No effect of induced happiness was found for accuracy of teacher effectiveness ratings and relationship type assessments (Ambady & Gray, 2002). First evidence that performance on more verbal-based decoding tasks might be hindered by positive affect comes from three studies showing that induced happy mood hampered accuracy in lie detection as compared to induced sad mood and a control condition (Reinhard & Schwarz, 2012). In another study, positive affect had no impact on lie detection (Forgas & East, 2008).

Research on the impact of positive affect on interpersonal accuracy is still scarce. So far, there is some evidence that positive affect might increase interpersonal accuracy on nonverbal tasks, while decreasing accuracy in detecting lies. Existing work primarily focused on the effects of happiness and amusement on interpersonal accuracy – the effects of other positive states such as euphoria, desire, and pride are still largely unknown.

Congruency effects

Effects of positive and negative affect may depend not only on whether interpersonal judgments are based on verbal versus nonverbal cues but also on the valence of these cues. According to mood-congruity theories (e.g., Bower, 1981; Schwarz, 1990; Schwarz & Clore, 1996), affect biases a person's perception and interpretation of affective cues in a mood-congruent way. In support of the mood-congruency hypothesis, happiness facilitated memory of positive descriptive statements (relative to negative statements) about target people, whereas sadness increased memory for negative statements (Forgas & Bower, 1987). Moreover, compared to a neutral state, both induced happiness and sadness hindered the recognition of mood-*incongruent* facial expressions (i.e., sad expressions in happy affective states and happy expressions in sad affective states); the recognition of mood-congruent facial expression was not significantly affected by individuals' affective states (Schmid & Schmid Mast, 2010). Individuals induced with happiness also recalled easy, informal information about a previous interaction better than individuals induced with sadness, who remembered difficult, formal information better (Forgas, Bower, & Krantz, 1984). To the extent that happiness facilitates effortless and sadness effortful information processing (e.g., Schwarz, 1990; Schwarz & Clore, 1996), these findings might also reflect congruency effects. These results suggest that mood effects on interpersonal accuracy may vary as a function of the congruency in terms of valence of the decoding material.

Affective factors: summary and discussion

Overall, the literature suggests that negative affect is detrimental for nonverbal decoding accuracy, whereas positive affect might be beneficial, although its impact is less clear. In contrast, in lie detection tasks that benefit from the processing of verbal information (e.g., Forgas & East, 2008; Reinhard & Schwarz, 2012), negative affect increased and positive affect decreased accuracy. Affect also seems to bias judgments such that positive and negative affect states boost accuracy in decoding mood-congruent and/or hinder accuracy in decoding mood-incongruent information. So far, research on the influences of affective states mostly focused on sadness and happiness, and we know little about the influence of other affective states. It is possible that effects are distinct to a specific emotion (e.g., to sadness but not to anger), rather than to positive versus negative valence – future research might clarify this question.

Cognitive factors

Although rarely shown empirically, it has been frequently proposed that effects of motivational and affective factors on interpersonal accuracy occur due to the use of different information processing styles (e.g., Forrest & Feldman, 2000; Hall et al., 2009; Keltner et al., 2003; Schwarz, 1990). However, the relation between information processing and interpersonal accuracy itself is not very clear, which makes such mediation-based explanations very speculative. In the following section, research on how deliberate versus automatic processing (i.e., effortful vs. non-effortful processing of information) and local versus global processing (i.e., separate vs. holistic processing of features of a cue) relate to interpersonal accuracy will be reviewed.

Deliberate versus automatic processing

Although accurate judgments can be made very quickly (i.e., in fragments of a second) and possibly intuitively (e.g., Matsumoto et al., 2000; Rule & Ambady, 2008), deliberation might help performing in some interpersonal accuracy tasks. Most of the research designed to explore the role of deliberate versus automatic processing in interpersonal accuracy included a cognitive load manipulation in which an additional task (e.g., a working memory task) had to be performed simultaneously with the interpersonal accuracy task. This presumably prevents people from deliberating and induces a more automatic processing style. In other studies, deliberate processing was induced by asking participants to take their time and to make careful and thoughtful judgments, while automatic processing was induced by instructing participants to make judgments based on their gut feeling, or by restricting response time windows.

Several studies showed that cognitive load decreased interpersonal accuracy, as indexed by reduced performance on the PONS (Phillips, Tunstall, & Channon, 2007), the IPT (McLarney-Vesotski et al., 2011, but see Phillips, et al., 2007), and emotion recognition (Phillips, Channon, Tunstall, Hedenstrom, & Lyons, 2008; Tracy & Robins, 2008). A further study revealed that audio/video signal delays in an interaction (which presumably causes cognitive load) *increased* accuracy in inferring the partner's feelings during the first period of the conversation but not during the second one (Powers, Rauh, Henning, Buck, & West, 2011). The authors argued that during the first period, participants were likely highly motivated to form accurate impressions and the delay gave them more time to do so.

In two emotion recognition studies by Tracy and Robins (2008), a cognitive load manipulation was compared to a speeded-response condition, in which the response window was restricted (and thus deliberation hindered), and a deliberate condition, in which participants were instructed to make thoughtful judgments. When comparing the cognitive load condition with the speeded-response condition, no significant effects were found, but both conditions showed decreased accuracy compared to the deliberate condition for the recognition of surprise, fear, sadness, anger, and pride. The recognition of happiness, disgust, contempt, shame, and embarrassment did not differ between conditions. In several emotion recognition studies by Phillips et al. (2008), cognitive load was compared to a no-load control condition. Recognition of all emotions (i.e., happiness, sadness, surprise, anger, and disgust) was similarly hampered by the manipulation. However, the more answer options were provided (two, four, or six categories of emotional expressions), the more performance was decreased by cognitive load; when simply having to discriminate emotion expressions in a same–different task, no effect of cognitive load emerged. Cognitive load might therefore have detrimental effects when task demands are high but not when they are low. In fact, participants who were instructed to look carefully for any specific person-related cues in the IPT (high-task demands) performed worse when they were under cognitive load, whereas participants who were instructed to form a first impression (less task demands) performed better on the IPT when they were under cognitive load (Patterson & Stockbridge, 1998).

To summarize, most research examining the effects of deliberate versus automatic processing included a cognitive load manipulation and findings generally demonstrated that cognitive load reduced interpersonal accuracy. The problem with cognitive load manipulations is that they create a dual-task situation. It is often assumed that performing a secondary task (i.e., the cognitive load task) reduces effort and deliberation in the primary task (i.e., the interpersonal accuracy task). However, alternative strategies can be applied that necessitate effortful executive control rather than automatic processing such as constantly switching attention between tasks or prioritizing one task over the other, whether it be the interpersonal accuracy task or the cognitive load task. It is therefore unclear whether effects of cognitive load on interpersonal accuracy were really due to hindered deliberative processing as they have often been interpreted. The effects of cognitive load may further depend on characteristics of the interpersonal accuracy tasks. For interpersonal accuracy tasks that require fewer resources, cognitive load might make people perform the two tasks in parallel and in a relatively automatic way, whereas for more cognitively demanding tasks, cognitive load might lead to a (controlled) shift in attention away from the interpersonal accuracy task toward the cognitive load task. This proposal is in line with studies demonstrating that cognitive load only reduced interpersonal accuracy when task demands were high but it helped or had no effect when task demands were low (Patterson & Stockbridge, 1998; Phillips et al., 2008). However, additional research is needed to shed light on the exact effects of cognitive load on information processing.

Local versus global processing

Most studies examining the role of local and global information processing in interpersonal accuracy manipulated processing styles by using the letter identification task or the verbal overshadowing paradigm. In the letter identification task (Navon, 1977), a series of big letters composed of small letters are presented and participants either read the small letters (local priming condition) or the big letters (global priming condition). The local versus global focus evoked in this priming task is assumed to transfer to subsequent tasks (e.g., Förster & Dannenberg, 2010). In the verbal overshadowing paradigm (e.g., Schooler & Engstler-Schooler, 1990), participants verbally describe the stimuli prior to the decoding task. This procedure directs individuals' attention to the detailed aspect of the stimuli, which is assumed to prompt a local information processing style.

Findings were mixed. For instance, global letter identification (as compared to local letter identification) decreased emotion recognition (Martin, Slessor, Allen, Phillips, & Darling, 2012, but see Schmid et al., 2011a) as well as performance on the PONS (Schmid, Schmid Mast, Bombari, & Mast, 2011b). Mostly opposite findings were found for face recognition tasks. Local processing decreased face recognition as compared to global processing when induced with the letter identification task (Macrae & Lewis, 2002; Perfect, Dennis, & Snell, 2007; Perfect, 2003, but see Lawson, 2007) as well as with the instruction to verbally describe faces (the so-called verbal overshadowing effect; Schooler & Engstler-Schooler, 1990; for a meta-analysis, see Meissner & Brigham, 2001).

The somewhat inconsistent findings may have emerged due to specifics of the manipulations and their confounding variables. For instance, it has been argued that when letters in the Navon (1977) letter identification task are visualized as having little spaces between the local letters, global letters are read more automatically than local letters, whereas the opposite is the case when the spaces between the small letters are large; it is thus possible that the letter identification task induces deliberate versus automatic processing rather than local versus global processing (Perfect, Weston, Dennis, & Snell, 2008; Weston & Perfect, 2005). Indeed, when there was little space between the local letters, participants reading the global letters recognized faces better as compared to participants reading the local letters, while the opposite was found when spaces between the small letters were large (Perfect et al., 2008). A confound may also occur when instructing participants to verbally describe the task stimuli (i.e., the verbal overshadowing paradigm) – this likely not only makes people focus on local features but also makes them process information more deliberately.

Cognitive factors: summary and discussion

Research on the influence of cognitive factors on interpersonal accuracy mostly suggests that deliberate and local processing increase interpersonal accuracy. However, features of the interpersonal accuracy task might moderate this effect. For example, tasks with more answer options may require a more deliberate processing style, whereas in tasks with fewer answer options, an automatic processing style may be helpful (Phillips et al., 2008).

Furthermore, findings also depend on specifics of the manipulation as well as its interactive effects with task properties. Such seems to be the case for the letter identification task. The construal of the Navon letters (i.e., the small letters close together vs. far apart) may determine whether stimuli are processed deliberately or automatically, with potentially opposite consequences for interpersonal accuracy (Perfect et al., 2008). Moreover, it is often supposed that cognitive load hinders deliberate processing although it is typically not controlled whether this is really the case. Cognitive load might also trigger a controlled shift of attention away or toward the interpersonal accuracy task. Its exact effect may depend on specifics of the task such as how cognitively demanding it is (Patterson & Stockbridge, 1998). Cognitive load should therefore not be seen as a clear-cut manipulation that hinders effortful information processing. These examples illustrate that unintended effects of the manipulations of cognitive factors may at least partly explain the somewhat inconsistent findings. Future research should therefore include rigorous controls of the manipulations.

Motor factors

In the field of embodied cognition, it has been argued that simulating another person's action might help to understand their internal state (Atkinson & Adolphs, 2005; Niedenthal, Brauer, Halberstadt, & Innes-Ker, 2001). Most of the research in this field has focused on the effects of facial mimicry on emotion recognition. Although research has consistently shown that people spontaneously mimic other people's expressions and behavior (e.g., Latkin & Chartrand, 2003; van Baaren, Holland, Kawakami, & van Knippenberg, 2004), it is less clear whether mimicking indeed increases interpersonal accuracy.

Experimental manipulations of mimicry typically intend to prevent mimicry by physically restricting facial movements. For instance, in several studies, participants were instructed to hold a pen in their mouth using their teeth while recognizing emotions. This restriction resulted in less accurate recognition of happiness and disgust but not of fear or sadness (Oberman, Winkielman, & Ramachandran, 2007). In a different study, this manipulation decreased in the recognition of happiness, disgust, and fear but not of anger, sadness, or surprise (Ponari, Conson, D'Amico, Grossi, & Trojano, 2012). When biting on a pen was compared to shoulder movement restrictions, no difference was found in terms of recognition of positive and negative emotions; however, in the bite condition, women (but not men) were slower at recognizing emotions (Stel & van Knippenberg, 2008). According to the authors, this may be due to women's greater expressiveness and thus greater reliance on feedback from their own facial expressions. Other facial restrictions that also focus on the lower half of the face, such as chewing a gum or holding a pen with one's lips, did not affect accuracy in the recognition of happiness, disgust, fear, or sadness (Oberman et al., 2007). Restriction to the upper half of the face (i.e., instruction to draw together two stickers placed on the inner edge of their eyebrows) was associated with decreased recognition accuracy of anger and fear, but not of happiness, disgust, sadness, or surprise (Ponari et al, 2012).

A potential limitation of manipulations that require the participant to use their facial muscles to hold or move an object is that they may be distracting and divert attentional resources away from the interpersonal accuracy task, akin to a cognitive load manipulation (Stel & van Knippenberg, 2008). This leaves the possibility that effects were due to the dual-task situation rather than hindered facial mimicry. Neal and Chartrand (2011) used different methodological approaches to investigate the role of enhanced and reduced facial feedback. In one study, they compared two patient groups that underwent different types of wrinkle-reducing procedures; one group received Botox injections that paralyzed some facial muscles and reduced afferent feedback from these muscles to the brain, and the other group underwent a dermal filler (Restylane) procedure that did not affect feedback from facial muscles. The botox group showed reduced emotion recognition accuracy compared to the restylane group. Assuming that the characteristics of these two patient groups and procedural side effects were comparable, these findings suggest that dampening afferent muscle signal impedes emotion recognition. A second study demonstrated that amplifying afferent muscle signals through proprioceptive feedback increased facial emotion recognition.

As mentioned in the introduction, this chapter focuses on research including experimental manipulations of situational factors and their effect on interpersonal accuracy; however, in the context of mimicry, it is important to note that research based on paralysis-causing neurological disorders and correlational data have revealed that mimicry is not relevant or at least not necessary for accurate emotion recognition. For example, multiple studies showed that patients with bilateral facial paralysis due to neurological disorders (e.g., Moebius syndrome) recognized emotions just as well as healthy controls (Bogart & Matsumoto, 2009; Calder, Keane, Cole, Campbell, & Young, 2010; Keillor, Barrett, Crucian, Kortenkamp, & Heilman, 2002), but two of three patients showed a slight disadvantage when asked to label the predominant emotion in experimentally morphed faces that displayed mixed emotions (Calder et al., 2010). Moreover, a series of unobtrusive studies in which facial mimicry was measured with facial EMG consistently showed that facial expressions were mimicked, but mimicry was not related to emotion recognition accuracy (Blairy, Herrera, & Hess, 1999; Hess & Blairy, 2001, but see Wallbott, 1991). These studies suggest that emotion recognition does not necessitate mimicry.

Motor factors: summary and discussion

Studies that included manipulations of facial mimicry generally suggest that hindering mimicry impedes the recognition of some emotions. In most of these studies, participants were given a task that requires the constant activation of some facial muscles to prevent mimicry (e.g., holding a pen with their mouth) while also performing the emotion recognition task. This creates a dual-task situation, and it is possible that the dual-task situation rather than the facial restriction caused the effects (Stel & van Knippenberg, 2008). The facial restriction manipulations also did not affect all facial muscles (e.g., only those involved when biting on a pen), which may explain at least in part why the recognition of some emotions was affected by such manipulations more than others. It remains unclear whether facial restriction manipulations prevent mimicry or whether they induce an emotional state themselves. For example, the finding that biting on a pen decreased the recognition of happiness seems inconsistent with Strack, Martin, and Steppner's (1988) seminal work, showing that biting on a pen leads to a positivity bias (greater funniness ratings of cartoons), which according to mood-congruity theories should facilitate the recognition of positive emotions.

Importantly, studies of neurological disorders and correlational research found little evidence for a link between mimicry and emotion recognition (e.g., Blairy, Herrera, & Hess, 1999; Hess & Blairy, 2001). Thus, interpersonal accuracy, or emotion recognition specifically, does not seem to necessitate mimicry (see Singer & Lamm, 2009, for a similar point in the context of empathy); however, the interesting question of whether mimicking others supports emotion recognition may require more research.

Conclusions

The literature on situational influences of motivational, affective, cognitive, and motor factors on interpersonal accuracy is characterized by heterogeneous results. This suggests that the link between such situational factors and accuracy in judging others and in recalling other people's attributes is complex and highly moderated. Findings depend on specifics of the task such as target readability (Thomas & Maio, 2008), number of answer options provided (Phillips et al., 2008), and whether the task includes nonverbal or verbal cues (e.g., Hall et al., 2009); these differences may arise because such task features determine whether a task requires deliberate or automatic processing of the cues. It is likely that a variety of other task aspects have similar effects such as variability in answer options (e.g., same vs. different answer options for the different trials of a task), number of target people (e.g., single person vs. group), stimuli homogeneity (e.g., standardized faces vs. a mix of face and full body stimuli), stimuli presentation mode (e.g., still pictures vs. movies), and type of judgment (e.g., global valence judgments vs. inference of specific thoughts and feelings). Clearly, more research is needed in order to gain a more comprehensive understanding of the moderating effects of such task characteristics.

Findings might also appear heterogeneous due to unknown effects of the manipulations. A lot of speculation has been made on the mechanisms through which manipulations intended to alter motivation, affect, and information processing, and in turn, influence interpersonal accuracy. For example, it has been speculated that accuracy motivation shifts people's processing style to a deliberate mode (e.g., Hall et al., 2009), and positive affect has been associated with global and automatic processing and negative affect with local and deliberate processing (e.g., Gasper & Clore, 2002; Schwarz, 1990). Existing work is largely consistent with these assumptions; however, research that directly tests these hypotheses is scarce (see also Hall, 2014). It is also likely that such mediation patterns are further moderated by task characteristics such as those listed above. Although many aspects of interpersonal accuracy require further investigation, the overall understanding is that interpersonal accuracy is malleable, such that the perceivers' own states influence what they infer from the target person.

References

Addington, J., Saeedi, H., & Addington, D. (2006). Facial affect recognition: A mediator between cognitive and social functioning in psychosis? *Schizophrenia Research*, 85, 142–150.

Ambady, N., & Gray, H. M. (2002). On being sad and mistaken: Mood effects on the accuracy of thin-slice judgments. *Journal of Personality and Social Psychology*, 83, 947–961.

Atkinson, A. P., & Adolphs, R. (2005). Visual emotion perception. In L. Feldman Barrett, P. M. Niedenthal, & P. Winkielman (Eds.), *Emotion and consciousness* (pp. 150–184). New York, NY: Guilford Press.

Bernieri, F. J. (2001). Toward a taxonomy of interpersonal sensitivity. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement* (pp. 3–20). Mahwah, NJ: Lawrence Erlbaum Associates.

Bernstein, M. J., Young, S. G., Brown, C. M., Sacco, D. F., & Claypool, H. M. (2008). Adaptive responses to social exclusion: Social rejection improves detection of real and fake smiles. *Psychological Science*, 19, 981–983.

Biesanz, J. C., & Human, L. J. (2010). The cost of forming more accurate impressions: Accuracy-motivated perceivers see the personality of others more distinctively but less normatively than perceivers without an explicit goal. *Psychological Science*, 21, 589–594.

Blairy, S., Herrera, P., & Hess, U. (1999). Mimicry and the judgment of emotional facial expressions. *Journal of Nonverbal Behavior*, 23, 5–41.

Bogart, K. R., & Matsumoto, D. (2009). Facial mimicry is not necessary to recognize emotion: Facial expression recognition by people with Moebius syndrome. *Social Neuroscience*, 5, 241–251.

Bower, G. H. (1981). Mood and memory. *American Psychologist*, 36, 129–148.

Camerer, C. F., & Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty*, 19, 7–42.

Chepenik, L. G., Cornew, L. A., & Farah, M. J. (2007). The influence of sad mood on cognition. *Emotion*, 7, 802–811.

Costanzo, M., & Archer, D. (1989). Interpreting the expressive behavior of others: The interpersonal perception task. *Journal of Nonverbal Behavior*, 13, 225–245.

Fiske, A. P. (1992). The four elementary forms of sociality: Framework for a unified theory of social relations. *Psychological Review*, 99, 689–723.

Fiske., S. T. (1993). Controlling other people: The impact of power on stereotyping. *American Psychologist*, 48, 621–628.

Forgas, J. P., & Bower, G. H. (1987). Mood effects on person perception judgments. *Journal of Personality and Social Psychology*, 53, 53–60.

Forgas, J. P., Bower, G. H., & Krantz, S. E. (1984). The influence of mood on perceptions of social interactions. *Journal of Personality and Social Psychology*, 20, 497–513.

Forgas, J. P., & East, R. (2008). On being happy and gullible: Mood effects on skepticism and the detection of deception. *Journal of Experimental Social Psychology*, 44, 1362–1367.

Forrest, J. A., & Feldman, R. S. (2000). Detecting deception and judge's involvement: Lower task involvement leads to better lie detection. *Personality and Social Psychology Bulletin*, 26, 118–125.

Förster, J., & Dannenberg, L. (2010). GLOMOsys: A systems account of global versus local processing. *Psychological Inquiry*, 21, 175–197.

Gasper, K., & Clore, G. L. (2002). Attending to the big picture: Mood and global versus local processing of visual information. *Psychological Science*, 13, 34–40.

Hall, J. A. (1978). Gender effects in decoding nonverbal cues. *Psychological Bulletin*, 85, 845–857.

Hall, J. A. (1984). *Nonverbal sex differences: Communication accuracy and expressive style*. Baltimore: John Hopkins University Press.

Hall, J. A. (2014). Manipulated motivation and interpersonal accuracy. In J. L. Smith, W. Ickes, J. A. Hall, & S. D. Hodges (Eds.), *Managing interpersonal sensitivity: Knowing when and when not to understand others* (pp. 1–20). New York: Nova Science Publishers, Inc.

Hall, J. A., Blanch, D. C., Horgan, T. G., Murphy N. A., Rosip, J. C., & Schmid Mast, M. (2009). Motivation and interpersonal sensitivity: Does it matter how hard you try? *Motivation and Emotion*, 33, 291–302.

Hall, J. A., & Schmid Mast, M. (2008). Are women always more interpersonally sensitive than men? Impact of goals and content domain. *Personality and Social Psychology Bulletin*, 34, 144–155.

Hall, J. A., Schmid Mast, M., & Latu, I. (2015). The vertical dimension of social relations and accurate interpersonal perception: A meta-analysis. *Journal of Nonverbal Behavior*, 39, 131–163.

Hanggi, Y. (2004). Stress and emotion recognition: An internet experiment. *Swiss Journal of Psychology*, 63, 113–125.

Hess, U. & Blairy, S. (2001). Facial mimicry and emotional contagion to dynamic emotional facial expressions and their influence on decoding accuracy. *International Journal of Psychophysiology*, 40, 129–141.

Ickes, W., Gesn, P. R., & Graham, T. (2000). Gender differences in empathic accuracy: Differential ability or differential motivation? *Personal Relationships*, 7, 95–109.

Johnson, K. J., & Fredrickson, B. L. (2005). "We all look the same to me": Positive emotions eliminate the own-race bias in face recognition. *Psychological Science*, 16, 875–881.

Latkin, J. L., & Chartrand, T. L. (2003). Using nonconscious behavioral mimicry to create affiliation and rapport. *Psychological Science*, 14, 334–339.

Keillor, J. M., Barrett, A. M., Crucian, G. P., Kortenkamp, S., Heilman, K. M. (2002). Emotional experience and perception in the absence of facial feedback. *Journal of International Neuropsychological Society*, 8, 130–135.

Keltner, D., Gruenfeld, D. H., & Anderson, A. (2003). Power, approach, and inhibition. *Psychological Review*, 110, 265–284.

Keltner, D., & Kring, A. (1998). Emotion, social function, and psychopathology. *General Psychological Review*, 2, 320–342.

Klein, K. J. K., & Hodges, S. D. (2001). Gender differences, motivation, and empathic accuracy: When it pays to understand. *Personality and Social Psychology Bulletin*, 27, 720–730.

Koenig, A. M., & Eagly, A. H. (2005). Stereotype threat in men on a test of social sensitivity. *Sex Roles*, 52, 489–496.

Lawson, R. (2007). Local and global processing biases fail to influence face, object, and word recognition. *Visual Cognition*, 15, 710–740.

Macrae, C. N., & Lewis, H. L. (2002). Do I know you? Processing orientation and face recognition. *Psychological Science*, 13, 194–196.

Martin, D., Slessor, G., Allen, R., Phillips, L. H., & Darling, S. (2012). Processing orientation and emotion recognition. *Emotion*, 12, 39–43.

Matsumoto, D., LeRoux, J., Wilson-Cohn, C., Raroque, J., Kooken, K., Ekman, P., . . . Goh, A. (2000). A new test to measure emotion recognition ability: Matsumoto and Ekman's Japanese and Caucasian Brief Affect Recognition Test (JACBART). *Journal of Nonverbal Behavior*, 24, 179–209.

McLarney-Vesotski, A., Bernieri, F., & Rempala, D. (2011). An experimental examination of the "good judge." *Journal of Research in Personality*, 45, 398–400.

Meissner, C. A., & Brigham, J. C. (2001). A meta-analysis of the verbal overshadowing effect in face identification. *Applied Cognitive Psychology*, 15, 603–616.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, 9, 353–383.

Neal, D. T., & Chartrand, T. L. (2011). Embodied emotion perception: Amplifying and dampening facial feedback modulates emotion perception accuracy. *Social Psychological and Personality Science*, 2, 673–678.

Niedenthal, P. M., Brauer, M., Halberstadt, J., & Innes-Ker, A. H. (2001). When did her smile drop? Facial mimicry and the influences of emotional state on the detection of change in emotional expression. *Cognition and Emotion*, 15, 853–864.

Nowicki, S. J., & Richman, D. (1985). The effect of standard, motivation, and strategy instructions on the facial processing accuracy of internal and external subjects. *Journal of Research in Personality*, 19, 254–364.

Oberman, L. M., Winkielman, P., & Ramachandran, V. S. (2007). Face to face: Blocking facial mimicry can selectively impair recognition of emotional expressions. *Social Neuroscience*, 2, 167–178.

Patterson, M. L., & Stockbridge, E. (1998). Effects of cognitive demand and judgment strategy on person perception accuracy. *Journal of Nonverbal Behavior*, 22, 253–263.

Perfect, T. J. (2003). Local processing bias impairs lineup performance. *Psychological Reports*, 93, 393–394.

Perfect, T. J., Dennis, I., & Snell, A. (2007). The effects of local and global processing orientation on eyewitness identification performance. *Memory*, 15, 784–798.

Perfect, T. J., Weston, N. J., Dennis, I., & Snell, A. (2008). The effects of precedence on Navon-induced processing bias in face recognition. *The Quarterly Journal of Experimental Psychology*, 61, 1479–1486.

Phillips, L. H., Channon, S., Tunstall, M., Hedenstrom, A., & Lyons, K. (2008). The role of working memory in decoding emotions. *Emotion*, 8, 184–191.

Phillips, L. H., Tunstall, M., & Channon, S. (2007). Exploring the role of memory in dynamic social cue decoding using dual task methodology. *Journal of Nonverbal Behavior*, 31, 137–152.

Ponari, M., Conson, M., D'Amico, N. P., Grossi, D., Trojano, L. (2012). Mapping correspondence between facial mimicry and emotion recognition in healthy subjects. *Emotion*, 12, 1398–1403.

Porter, S., McCabe, S., Woodworth, M., & Peace, K. A. (2007). "Genius is 1% inspiration and 99% perspiration"... or is it? An investigation of the impact of motivation and feedback on deception detection. *Legal and Criminological Psychology*, 12, 297–309.

Powers, S. R., Rauh, C., Henning, R. A., Buck, R. W., & West, T. V. (2011). The effect of video feedback delay on frustration and emotion communication accuracy. *Computers in Human Behavior*, 27, 1651–1657.

Reinhard, M.-A., & Schwarz, N. (2012). The influence of affective states on the process of lie detection. *Journal of Experimental Psychology: Applied*, 18, 377–389.

Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., & Archer, D. (1979). *Sensitivity to nonverbal communication: The PONS test*. Baltimore: Johns Hopkins University Press.

Rule, N. O., & Ambady, N. (2008). Brief exposures: Male sexual orientation is accurately perceived at 50 ms. *Journal of Experimental Psychology*, 44, 1100–1105.

Rule, N. O., Rosen, K. S., Slepian, M. L., & Ambady, N. (2011). Mating interest improves women's accuracy in judging male sexual orientation. *Psychological Science*, 22, 881–886.

Sassenrath, C., Sassenberg, K., Ray, D. G., Scheiter, K., & Jarodzka, H. (2014). A motivational determinant of facial emotion recognition: Regulatory focus affects recognition of emotions in faces. *PLoS ONE*, 9, e112383.

Schmid Mast, M., Jonas, K., & Hall, J. A. (2009). Give a person power and he or she will show interpersonal sensitivity: The phenomenon and its why and when. *Journal of Personality and Social Psychology*, 97, 835–850.

Schmid, P. C., & Schmid Mast, M. (2010). Mood effects on emotion recognition. *Motivation and Emotion*, 34, 288–292.

Schmid, P. C., Schmid Mast, M., Bombari, D., & Mast, F. W. (2011b). Gender effects in information processing on a nonverbal decoding task. *Sex Roles*, 65, 102–107.

Schmid, P. C., Schmid Mast, M., Bombari, D., Mast, F. W., & Lobmaier, J. S. (2011a). How mood states affect information processing during facial emotion recognition: An eye tracking study. *Swiss Journal of Psychology: Special Issue: Social Cues in Faces*, 70, 223–231.

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology*, 22, 36–71.

Schwarz, N. (1990). Feelings as information: Informational and motivational functions of affective states. In E. T. Higgins and R. Sorrentino (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (Vol. 2, pp. 527–561). New York, NY: Guilford.

Schwarz, N., & Clore, G. L. (1996). Feelings and phenomenal experiences. In E. T. Higgins & A. Kruglanski (Eds.), *Social Psychology: Handbook of basic principles* (pp. 433–465). New York: Guilford Press.

Sinclair, R. C. (1988). Mood, categorization breadth and performance appraisal: The effects of order of information acquisition and affective state on halo, accuracy, information retrieval, and evaluations. *Organizational Behavior and Human Decision Processes*, 42, 22–46.

Singer, T., & Lamm, C. (2009). The social neuroscience of empathy. *Annuals of the New York Academy of Sciences*, 1156, 81–96.

Smith, J. L., & Lewis, K. L. (2009). Men's interpersonal (mis)perception: Fitting in with gender norms following social rejection. *Sex Roles*, 61, 252–264.

Snodgrass, S. E. (1985). Women's intuition: The effect of subordinate role on interpersonal sensitivity. *Journal of Personality and Social Psychology*, 49, 146–155.

Stel, M., & van Knippenberg (2008). The role of facial mimicry in the recognition of affect. *Psychological Science*, 19, 984–985.

Strack, F., Martin, L. L., Stepper, S. (1988). Inhibiting and facilitating conditions of the human smile: A nonobtrusive test of the facial feedback hypothesis. *Journal of Personality and Social Psychology*, 54, 768–777.

Thomas, G., & Maio, G. R. (2008). Man, I feel like a woman: When and how gender-role motivation helps mind-reading. *Journal of Personality and Social Psychology*, 95, 1165–1179.

Tracy, J. L., & Robins, R. W. (2008). The automaticity of emotion recognition. *Emotion*, 8, 81–95.

Van Baaren, R. B., Holland, R. W., Kawakami, K., & van Knippenberg, A. (2004). Mimicry and prosocial behavior. *Psychological Science*, 15, 71–74.

Wallbott, H. G. (1991). Recognition of emotion from facial expression via imitation? Some indirect evidence for an old theory. *British Journal of Social Psychology*, 30, 207–219.

Westermann, R., Spies, K., Stahl, G., & Hesse, F. W. (1996). Relative effectiveness and validity of mood induction procedures: a meta-analysis. *European Journal of Social Psychology*, 26, 557–580.

Weston, N. J., & Perfect, T. J. (2005). Effects of processing bias on the recognition of composite face halves. *Psychonomic Bulletin & Review*, 12, 1038–1042.

Young, S. G., Hugenberg, K., Bernstein, M. J., & Sacco, D. F. (2012). Perception and motivation in face recognition: A critical review of theories of the cross-race effect. *Personality and Social Psychology Review*, 16, 116–142.

12 Training people to be interpersonally accurate

Danielle Blanch-Hartigan, Susan A. Andrzejewski, and Krista M. Hill

Abstract
Given that accurate person perception is a skill associated with a host of positive interpersonal and applied outcomes, a logical extension is to seek to improve the skill. Training person perception involves attempts to improve accuracy of judgments of others' emotions, personality traits, status, and intentions. There is a rich history of training person perception accuracy dating back to Floyd Henry Allport and Arthur Jenness in the 1920s and 1930s. This chapter describes the history of training person perception accuracy and then summarizes a recent meta-analysis, including how training domains and approaches moderate training efficacy. The potential benefits of training and current training research in the applied areas of medicine, law enforcement, and consumer services are presented. Finally, future research needs are proposed to build the evidence base in person perception training and apply these training efforts in real-world contexts by (1) further establishing the benefits of training in applied contexts, (2) developing effective trainings, (3) optimizing training efficacy, and (4) disseminating, implementing, and evaluating training programs.

People make countless judgments of others based on minimal nonverbal information from vocal cues, facial and body movements, physiognomy, dress, and other nonverbal cues. Even when information is fleeting, individuals rely on these cues to draw inferences about others. For example, Will the employee I briefly interviewed be a good organizational fit for our company? Is that person going to try and interact with me? Is the suspect telling the truth? Is my patient in pain? Many inferences derived from nonverbal cues are accurate; however, individuals demonstrate considerable variability in their ability to accurately perceive nonverbal cues. The ability to accurately perceive nonverbal behavior is a type of person perception. Person perception accuracy is the ability to correctly infer the states and traits of others or accurately predict future behavior. This broad definition encompasses judgments and predictions of others' emotions, personality, status, and intentions from verbal and nonverbal cues. Training person perception involves attempts to improve accuracy of these judgments. This chapter will focus specifically on research on training to increase accurate person perception of nonverbal behaviors in nonclinical adult populations.

Benefits of person perception accuracy

Accurate person perception is associated with myriad positive psychosocial characteristics and behaviors (see Hall, Andrzejewski, & Yopchick, 2009 for a review). More accurate perceivers demonstrate fewer depressive symptoms (Carton, Kessler, & Pape, 1999), report less shyness (Schroeder, 1995), report higher social competence (Barnes & Sternberg, 1989), and report having better conversational skills (Miczo, Segrin, & Allspach, 2001). Accurate person perception is also associated with positive social attitudes such as lower prejudice levels (Andrzejewski, Hall, & Salib, 2009; Rule et al., 2015) and less likelihood to use gender stereotyping (Hall & Carter, 1999). Individuals who are more adept at person perception also report higher relationship well-being (Carton et al., 1999) and have more positive attachment styles (Cooley, 2005).

Research in applied settings demonstrates the positive relationship between accurate person perception and positive performance and outcome measures. For example, physicians who accurately recognize the emotional states of their patients receive higher levels of patient satisfaction (DiMatteo, Friedman, & Taranta, 1979; Hall, 2011). In business contexts, individuals who are more accurate at recognizing the emotional states of consumers receive larger salary raises (Byron, Terranova, & Nowicki, 2007) and receive more positive managerial ratings (Byron, 2007). Moreover, employees who are better able to accurately pick up on affective vocal cues rank higher within universities than their less accurate peers (Hall & Halberstadt, 1994).

With decades of research across a wide array of settings demonstrating the association with many positive psychosocial variables and outcomes (Hall et al., 2009), many seek to improve person perception accuracy. A logical extension of this research is to investigate whether or not training improves person perception accuracy, as well as the most effective methods for training person perception accuracy.

There is evidence that certain formative or learning experiences have the potential to increase person perception accuracy. For example, taking music lessons (Thompson, Schellenberg, & Husain, 2004), dance experience (Pitterman & Nowicki, 2004), and having a preverbal toddler at home (Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979) have all been shown to correlate positively with individuals' person perception ability. The potential capacity of such formative experiences to enhance accurate person perception suggests that training may improve the accuracy of inferring the internal states and traits of others.

History of training person perception accuracy

There is a rich history of training programs in person perception accuracy dating back almost a century. The first published study on training person perception accuracy dates back to 1924 by Floyd Henry Allport, considered the father of experimental social psychology and one of the first to focus on individual differences in behaviors and personality (Katz, 1979). In his book *Social Psychology* he describes a study in which 12 young women took a test of facial emotional expressions, studied a chart of facial expressions for 15 minutes, and then repeated the test. Allport reported that eight out of the 12 showed improved performance after practice. Performance improved the most for those who did especially poorly on the first test. Describing improvements in facial emotion recognition he states, "The inferior [judges] are aided by grasping at any clue which will support their meager understanding of the features." Allport's assumption was that development of person perception accuracy and improvements due to training were a result of practice. In 1932, Arthur Jenness, a student of Allport, questioning the aspects of his teacher's original study, set out to replicate and expand on the original research. In two studies, he demonstrated similar improvements in accuracy with both 15 minutes and 45 minutes of practice with larger sample sizes (Jenness, 1932).

In the 1970s, Robert Rosenthal picked up the charge of training person perception accuracy. Rosenthal et al. (1979) summarized 50 years of literature to date, approximately 18 studies from 1924 to 1974 including the work of Allport and Jenness, and concluded that person perception accuracy was a trainable skill. Using the Profile of Nonverbal Sensitivity (PONS), a validated test of accuracy in decoding affect from nonverbal cues of the face, body, and voice, they demonstrated for the first time using a randomized, controlled experimental design that training could improve person perception accuracy. The training included lecturing on the importance of nonverbal communication and practice in judging affect. This work was also groundbreaking because it included judgments of videotaped expressions. Since the work of Rosenthal and colleagues in the 1970s, the field has seen an increase of research in this area as new technologies enable realistic person perception accuracy paradigms. The field has also increasingly focused on training person perception accuracy in various applied settings, including medicine and consumer relations.

Training person perception accuracy in clinical populations

Although the purpose of this chapter is to provide an overview of training in nonclinical populations, it is important to note that there is an extensive literature on training programs for clinical populations. Some of the clinical disorders include deficits in person perception ability (e.g., autism spectrum disorder, Asperger syndrome, and some learning disabilities), and training programs in these clinical populations have demonstrated significant improvements in person perception accuracy (McKenzie, Matheson, McKaskie, Hamilton, & Murray, 2000; Silver & Oakes, 2001). Specifically, research has found that training programs produce significant improvements in person perception accuracy for adults and children on the autism spectrum (Lopata, Thomeer, Volker, & Nida, 2006; Lopata, Thomeer, Volker, Nida, & Lee, 2008; Silver & Oakes, 2001; Solomon, Goodlin-Jones, & Anders, 2004; Statucka & Walder, 2013), with learning disabilities (McKenzie et al., 2000; Wood & Kroese, 2007) and with schizophrenia (Russell, Chu, & Phillips, 2006; Silver, Goodman, Knoll, & Isakov, 2004) (see Chapter 9 for discussion of interpersonal accuracy deficits in clinical populations). The trainings, training environments, and assessments differ across these studies. For example, various techniques are used to train emotion recognition in individuals with autism spectrum disorder, including teaching the basic emotions and allowing individuals to practice emotion recognition, either in group settings or through computer programs (Cappadocia & Weiss, 2011; Faja, Aylward, Bernier, & Dawson, 2007; White, Keonig, & Scahill, 2007). Within the learning disability literature, Wood and Kroese's (2007) review found that a mixture of group and individual trainings was used with various tasks, including group discussions, instruction, practice, and feedback. Statucka and Walder's (2013) review also identified wide variation in training methods developed to improve emotion recognition skills in patients with schizophrenia. Programs ranged from small groups requiring multiple sessions to short, computer-based, individually administered interventions. Some of the interventions focused on teaching patients how to identify and discriminate facial expressions of emotion, while others utilized practice, and others took a more holistic approach that encouraged patients to better understand social situations in general.

Training effectiveness in nonclinical adult populations

The literature in clinical populations clearly demonstrates the effectiveness of training to improve the person perception accuracy. However, nonclinical populations perform better on many person perception tasks. Many individuals can draw very accurate inferences in domains such as detecting deception, affect, and sexual orientation from very thin excerpts of nonverbal information ranging from 3 to 30 seconds (Ambady, Hallahan, & Conner, 1999; Ambady & Rosenthal, 1992). Therefore, one might question whether person perception training would yield positive outcomes for nonclinical adult populations.

To address this concern, the present authors conducted a meta-analysis of 37 independent effect sizes from studies that examined the effectiveness of training on person perception accuracy in nonclinical adult populations by comparing the training group to a control group. In this meta-analysis, training was defined as, “any interventional approach designed to increase accuracy of person perception that was not purely motivational in nature” (Blanch-Hartigan, Andrzejewski, & Hill, 2012, p. 486). This definition included training through instruction, practice, and feedback. Training was significantly effective in improving person perception accuracy (overall effect size $r = .18$). In addition, meta-analyzed results from 21 nonexperimental within-subjects designs also demonstrated that training produced a significant benefit in person perception accuracy ($r = .44$). The effect size for the within-subjects design was likely larger due to this methodology accounting for individual differences in person perception accuracy at baseline. The meta-analysis demonstrated that training can effectively improve person perception accuracy in various domains using a variety of training approaches. The findings also offered insights into practical aspects of effective training programs, including training length and the type of training that may be most effective, which are covered in the next sections.

Training domains

There are many content areas, or domains, to person perception, including judging emotions, attitudes, truthfulness, personality, and social attributes. What unifies these domains under the construct of person perception accuracy is that these judgments all involve the perception of another person (Bernieri, 2001). Training was significantly effective in every person perception domain included in the meta-analysis. However, the meta-analysis revealed differences in effect sizes in moderator analyses comparing domains (Blanch-Hartigan et al., 2012). Training was more effective for improving accurate perceptions of others' comprehension, thoughts and feelings, and truthfulness, than for other person perception domains including judging status.

Compared to other domains, deception detection has the most extensive research literature on training accuracy (Hauch, Sporer, Michael, & Meissner, 2014). Training related to judging others' thoughts and feeling using the empathic accuracy paradigm is also common (Ickes, 2001; also Chapter 3). Interestingly, although much research on person perception is in the domain of emotion recognition, there are very few experimentally validated interventions to improve emotion recognition accuracy in non-clinical adult populations. A recent training program was developed for healthcare providers to improve emotion recognition accuracy involving practice with feedback about performance and discussion of emotion cues (Ruben, Hall, Curtin, Blanch-Hartigan, & Ship, 2015). This training increased accuracy on the Test of Accurate Perception of Patients' Affect (TAPPA), a validated test of accuracy in perception of patient emotions (Hall et al., 2014). Additional examples of emotion recognition training include the Facial Action Coding System (FACS), a comprehensive certification program for recognizing and coding facial movements based on muscular action (Ekman & Rosenberg, 1997) and Paul Ekman's Micro Expressions Training Tool (METT) and Subtle Expressions Training Tool (SETT) (Ekman, 2002). Based on 40 years of Ekman's research, these programs have demonstrated some success at improving accuracy for spotting microexpressions, very brief expressions of emotion in both frontal and profile views of the face that are often consciously or unconsciously concealed (Matsumoto & Hwang, 2011).

Training approaches

Besides domain, the other main distinction in person perception accuracy training is the structure or approach of the training program. Training programs or interventions typically include some or all of the following: practice (familiarity with judgments or exposure to the target being judged), feedback (about overall performance or item-specific performance), and instruction (education on specific accuracy cues or increased awareness about the importance of accurate perception). Moderator analysis results from the meta-analysis revealed that practice with feedback was the most effective training approach (Blanch-Hartigan et al., 2012). Instruction alone was not significantly effective in improving accuracy, even though it was a common training approach.

It appears that providing information about the specific verbal or nonverbal cues to look for may not improve person perception accuracy. The lack of efficacy for instruction alone can yield insights into the mechanism for training effectiveness. One possible explanation is that instruction tries to manipulate a conscious processing of information. Individuals are trained to seek out specific cues and apply what they have been taught. However, many of the automatic, rapid judgments that lead to person perception accuracy may be outside conscious awareness and therefore not amenable to this training approach. Research by Marc-Andre Reinhard and colleagues in the domain of lie detection supports the assumption that person perception may be related to unconscious, as opposed to conscious, processing (Reinhard, Greifeneder, & Scharmach, 2013). In a series of experiments, participants were less accurate in deception detection after actively focusing and deliberating on their judgments compared to others who were distracted with cognitive tasks. In fact, participants making judgments while distracted were significantly more accurate than those who were told to deliberate about their perceptions, and control subjects who were not given specific direction for how to make judgments.

Instruction on its own may not be beneficial, but may increase accuracy when combined with other training approaches (Blanch-Hartigan et al., 2012). Many training interventions employ some combination of training approaches. Results from training studies that tested the influence of the unique components versus a combined approach indicate that a combination of approaches is more effective than a single approach (Blanch-Hartigan, 2012; Costanzo, 1992), with a linear trend of more components yielding greater training efficacy (Ruben et al., 2015). However, few training studies assess individual training components in this way, and no studies to our knowledge have systematically studied the most effective order of training approaches.

Training length

Notably, when we examine characteristics of effective training in person perception accuracy, it appears that the length of training is not a strong determinant of improved accuracy. This finding seems counterintuitive. Indeed, Jenness questioned Allport's results that a short training could improve accuracy (Jenness, 1932). He wrote, "it seems to the writer that one may legitimately question whether 15 minutes of studying such a complicated chart as the one used actually results in a significant improvement in ability to name facial expressions." So after replicating Allport's 15-minute training, he tested a similar 45-minute training. Demonstrating equivalent effects, he concluded, "trebling the length of time for study seems to have little effect on the average gain as a result of training." His conclusion was confirmed by a recent meta-analysis and systematic review (Blanch-Hartigan et al., 2012; Blanch-Hartigan & Ruben, 2013).

Other results of practical importance from the meta-analysis suggest that training is more effectively delivered individually or in small groups than in large group settings and when administered by an instructor as opposed to self-administered on a computer (Blanch-Hartigan et al., 2012).

Training person perception in applied contexts

Person perception accuracy has largely been examined in the basic psychological and communication literatures; however, the potential benefits of person perception accuracy training for applied settings such as medicine, law enforcement, and customer service are numerous.

Medicine

Person perception accuracy is an important skill for healthcare providers. In order to determine appropriate treatments and respond to patients' psychosocial needs, healthcare providers must make accurate perceptions of their patients (Blanch-Hartigan & Ruben, 2013; Hall, 2011; Chapter 14). For example, in routine clinical interactions providers are tasked with assessing patient emotions, preferences for treatment decisions, adherence to medical recommendations, and level of pain. Providers who are better at accurately perceiving their patients have higher levels of patient satisfaction, among other positive clinical outcomes (Hall et al., 2015). Patients often report that their psychosocial needs are not met during clinical care; training providers to recognize patients' emotional cues may lead to more satisfied patients. Training healthcare providers to accurately recognize patients' cues could help providers prescribe the correct amount of pain medication, for example.

Thirteen studies of training in medical contexts have attempted to improve healthcare providers' accuracy in perceiving their patients, in the domains of emotion recognition, patient or caregiver distress, personality assessment, and depression diagnosis (Blanch-Hartigan & Ruben, 2013). They also varied in training length and approach. Ten of 13 studies demonstrated training improved person perception accuracy.

Person perception training may be particularly relevant for providers interacting with patients who have functional barriers to expressing emotions. For example, patients with facial paralysis - as a result of trauma to the face, neurological damage, or conditions including Bell's Palsy and Moebius syndrome, often cannot express emotions through their face. This makes it difficult for others to perceive them accurately and can lead to negative interpersonal perceptions of these individuals, even from healthcare providers who see many patients with these conditions (Bogart, Tickle-Degnen, & Ambady, 2014). In particular, healthcare providers may be biased to judge these individuals as less extraverted. A novel training program was developed to train providers through education about facial paralysis and feedback about their accuracy (Bogart & Tickle-Degnen, 2015). Although the training did not significantly improve overall accuracy for patients with facial paralysis, it did reduce the negativity bias in perceptions of these individuals. Person perception training may reduce stigma toward these individuals. Moreover, accurate person perception may prove valuable for healthcare providers when there are sociocultural barriers to communication such as gender and racial differences, or a perceived status imbalance between provider and patient (Cooper-Patrick et al., 1999).

Law enforcement

Person perception training may assist law enforcement agents to better recognize cues that are diagnostic of deception and relevant emotional states (e.g., fear and anger). The Transportation Security Administration (TSA) established the Screening of Passengers by Observational Techniques (SPOT) program to assist agents in identifying malicious intent in passengers (Transportation Security Administration, 2013). As part of this initiative, the TSA has developed a specialized position, the Behavior Detection Officer (BDO), whose primary task is to view the security screening process and identify behavior patterns that may pose a security risk. As part of their training, BDOs may receive nonverbal training and their experience on the job may also lead them to be more attuned to passengers' nonverbal behavior. Interestingly, when BDOs were exposed to a microexpression person perception training program, their ability to accurately pick up on nonverbal cues was moderated by the amount of previous nonverbal training they received throughout their career. BDOs without any prior nonverbal training improved the most, while BDOs with some prior nonverbal training did not show as strong an effect (Hurley, Anker, Frank, Matsumoto, & Hwang, 2014).

Training person perception accuracy in applied contexts often necessitates multidisciplinary research collaboration. For example, the TSA, Department of Defense, and Federal Bureau of Investigation share an interest in credibility assessment and, thus, often combine efforts from social psychologists and communication experts with other fields, including information systems. Together, researchers are developing computerized automated screening systems to assess credibility (Nunamaker, Derrick, Elkins, Burgoon, & Patton, 2011; Twyman, Elkins, Burgoon, & Nunamaker, 2014). These systems can be used to help train person perception accuracy in criminal investigators or airport security.

Business

In retail or service situations, the ability to make accurate perceptions of others based on first impressions is considered an essential component of marketing communication (Bonoma & Felder, 1977). Nonverbal behaviors, such as facial expression and body movements, communicate more information about what a customer is thinking and feeling than words and provide unique insights into customers' thoughts and feelings (Puccinelli, Motyka, & Grewal, 2010). Tapping into this insight has been linked to increased customer satisfaction (Puccinelli, Andrzejewski, Markos, Noga, & Motyka, 2013). Matsumoto and Hwang (2011) found that microexpression perception training has the ability to enhance the perception of barely perceptible nonverbal behaviors in retail employees. Such training may ultimately have a positive impact on a firm's bottom line as trained retail and service providers may be better able to meet the needs of consumers through accurate person perception.

Accuracy in person perception has also been linked to positive managerial ratings (Byron, 2007), better job performance (Hall et al., 2009; Rosenthal et al., 1979), the ability to accurately determine a candidate's personality in interview settings (Barrick, Patton, & Haugland, 2000), and multiple positive psychosocial traits that are commonly considered important in the business world (e.g., trust and affiliation in negotiations, etc.; Hall et al., 2009).

Research needs in training person perception accuracy

Despite the importance of person perception accuracy and evidence to date for training efficacy and potential benefits in many contexts, we still have a long way to go in building the evidence base for person perception accuracy training. We propose the following research is still needed: (1)

Work to further validate the correlation between accuracy and positive outcomes in various contexts (e.g., medicine and business); (2) Develop effective trainings in multiple domains; (3) Optimize training efficacy for various contexts and audiences; (4) Disseminate, implement, and evaluate training programs.

The first step in the process is to establish the potential benefits of trainings. Although work has been done in this area and the association between person perception accuracy and positive outcomes is well established (Hall et al., 2009), additional research demonstrating the benefits of increased person perception accuracy in particular contexts is needed. Next, researchers must work to develop training programs and demonstrate their reliability and validity of training methods across various domains (e.g., emotion recognition, judging personality, etc.). Ideally, validity should be established through the use of randomly assigned, experimental methods. Once effective trainings have been established, the next step is to maximize efficacy and cost-effectiveness by examining which components or approaches to training this skill are most beneficial. Here, researchers can also work to identify groups of individuals who will benefit most from training. Training approaches can be effectively tailored to specific contexts to fully meet the needs of particular individuals or particular settings, for example, in business, medicine, or law enforcement. Once training is optimized, efforts should turn to dissemination and implementation of training in the particular applied context. This includes research on the particular barriers to scaling up of programs and the fidelity with which experimentally validated training programs are applied in a "real-world" setting. In order to fully understand the effects of person perception accuracy training, we must evaluate the training in the applied contexts. This is looking beyond the primary outcomes that may have established efficacy in the development and optimization phases to include more long term, applied outcomes. For example, an effective training to increase emotion recognition accuracy implemented in a customer service setting should evaluate downstream consequences in customer satisfaction or sales.

Future directions in training person perception accuracy

For many domains of person perception accuracy and in many applied contexts, we are well on our way through this hierarchy of needs. The benefits of person perception accuracy have been established for many domains (Hall et al., 2009) and efficacy trials of promising interventions have been conducted (Blanch-Hartigan et al., 2012).

These efforts are really just a first step in increasing person perception accuracy in the broader context.

It is important for researchers within these fields to further explore the efficacy of training within these specific contexts. Training may benefit some groups more than others, and specific training methods may be more effective for some than they are for others. It is important to identify these differences in order to optimize and individualize trainings for individuals throughout various fields. For example, the majority of research in person perception accuracy training (in nonclinical adult populations) relies on convenience samples, typically college student populations (Blanch-Hartigan et al., 2012). Although a lot can be learned from this research, especially in the early stages of training development and optimization, later stages may require the use of specific populations or community-based samples.

We must not consider the implementation and evaluation of person perception training programs to be outside the scope of person perception research. Failure to disseminate evidence-based training programs can lead to the adoption of unvalidated and perhaps ineffective trainings. An online search for lie detection training, for example, will produce numerous programs that have been implemented without the benefit of either rigorous scientific development or formalized evaluation. A glance at the self-help or psychology shelves at a bookstore displays multiple titles about improving the ability to perceive others with enticing titles like: *The Power of Body Language: How to Succeed in Every Business and Social Encounter*, *How to Read a Client from Across the Room: Win More Business with the Proven Character Code System to Decode Verbal and Nonverbal Communication*, and *What Every BODY is Saying: An Ex-FBI Agent's Guide to Speed-Reading People*.

Another important question that must be addressed is whether evidence-based training can have benefits for interpersonal communication. For example, if a service provider can accurately identify that a customer is upset will the provider also be better at communicating with the customer? Will this then produce real-world benefits such as increased customer satisfaction or sales? In addition, it would be beneficial to know whether training in one domain would also improve accuracy in other domains. For example, research suggests that emotion recognition accuracy is somewhat correlated with personality judgment accuracy and empathic accuracy (Ruben, Hill, & Hall, 2014; see also Chapter 18). Would training to improve emotion recognition also improve personality judgment?

Considering person perception accuracy is associated with many positive outcomes, it seems possible that a training that improves accuracy will subsequently improve these outcomes. However, in most cases, long-term or generalized outcomes have not been studied. Research examining these types of results will not only support adoption of training programs, but also contribute to person perception theory by identifying the directionality of the relationship between person perception accuracy and its positive correlates.

References

Allport, F. H. (1924). *Social psychology*. Cambridge, MA: Riverside Press.

Ambady, N., Hallahan, M., & Conner, B. (1999). Accuracy of judgments of sexual orientation from thin slices of behavior. *Journal of Personality and Social Psychology*, 77, 538–547.

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, 111, 256–274.

Andrzejewski, S. A., Hall, J. A., & Salib, E. R. (2009). Anti-Semitism and identification of Jewish group membership from photographs. *Journal of Nonverbal Behavior*, 33, 149–180.

Barnes, M. L., & Sternberg, R. J. (1989). Social intelligence and decoding of nonverbal cues. *Intelligence*, 13, 263–287.

Barrick, M. R., Patton, G. K., & Haugland, S. N. (2000). Accuracy of interviewer judgments of job applicant personality traits. *Personnel Psychology*, 44, 1–26.

Bernieri, F. J. (2001). Toward a taxonomy of interpersonal sensitivity. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement* (pp. 3–20). Mahwah, NJ: Lawrence Erlbaum Associates.

Blanch-Hartigan, D. (2012). An effective training to increase accurate recognition of patient emotion cues. *Patient Education and Counseling*, 89, 274–280.

Blanch-Hartigan, D., Andrzejewski, S. A., & Hill, K. M. (2012). The effectiveness of training to improve person perception accuracy: A meta-analysis. *Basic and Applied Social Psychology*, 34, 483–498.

Blanch-Hartigan, D., & Ruben, M. A. (2013). Training clinicians to accurately perceive their patients: Current state and future directions. *Patient Education and Counseling*, 92, 328–336.

Bogart, K. R., & Tickle-Degnen, L. (2015). Looking beyond the face: A training to improve perceivers' impressions of people with facial paralysis. *Patient Education and Counseling*, 98, 251–256.

Bogart, K. R., Tickle-Degnen, L., & Ambady, N. (2014). Communicating without the face: Holistic perception of emotions of people with facial paralysis. *Basic and Applied Social Psychology*, 36, 309–320.

Bonoma, T. V. & Felder, L. C. (1977). Nonverbal communication in marketing: Toward a communication analysis. *Journal of Marketing Research*, 14, 169–180.

Byron, K. (2007). Male and female managers' ability to "read" emotions: Relationships with supervisor's performance ratings and subordinates satisfaction ratings. *Journal of Occupational and Organizational Psychology*, 80, 713–733.

Byron, K., Terranova, S., & Nowicki, S. (2007). Nonverbal emotion recognition and salespersons: Linking ability to perceived and actual success. *Journal of Applied Social Psychology*, 27, 2600–2619.

Cappadocia, M. C., & Weiss, J. A. (2011). Review of social skills training groups for youth with Asperger syndrome and high functioning autism. *Research in Autism Spectrum Disorder*, 5, 70–78.

Carton, J. S., Kessler, E. A., & Pape, C. L. (1999). Nonverbal decoding skills and relationship well-being in adults. *Journal of Nonverbal Behavior*, 23, 91–100.

Cooley, E. L., (2005). Attachment style and decoding of nonverbal cues. *North American Journal of Psychology*, 7, 25–34.

Cooper-Patrick, L., Gallo, J. J., Gonzales, J. J., Vu, H. T., Powe, N. R., Nelson, C., & Ford, D. E. (1999). Race, gender, and partnership in the patient-physician relationship. *Journal of the American Medical Association*, 282, 583–589.

Costanzo, M. (1992). Training students to decode verbal and nonverbal cues: Effects on confidence and performance. *Journal of Educational Psychology*, 84, 308–313.

DiMatteo, M. R., Friedman, H. S., & Taranta, A. (1979). Sensitivity to bodily nonverbal communication as a factor in practitioner-patient rapport. *Journal of Nonverbal Behavior*, 4, 18–26.

Ekman, P. (2002). *Micro Expression Training Tool (METT)*. San Francisco: University of California.

Ekman, P., & Rosenberg, E. L. (Eds.). (1997). *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford: Oxford University Press.

Faja, S., Aylward, E., Bernier, R., & Dawson, G. (2007). Becoming a face expert: A computerized face-training program for high-functioning individuals with autism spectrum disorders. *Developmental Neuropsychology*, 33, 1–24.

Hall, J. A., Andrzejewski, S. A., & Yopchick, J. E. (2009). Psychosocial correlates of interpersonal sensitivity: A meta-analysis. *Journal of Nonverbal Behavior*, 33, 149–180.

Hall, J. A., & Carter, J. D. (1999). Gender-stereotype accuracy as an individual difference. *Journal of Personality and Social Psychology*, 77, 350–359.

Hall, J. A., & Halberstadt, A. G. (1994). "Subordination" and sensitivity to nonverbal cues: A study and synthesis of findings based on trait measures. *Sex Roles*, 37, 295–317.

Hall, J. A. (2011). Clinicians' accuracy in perceiving patients: Its relevance for clinical practice and a narrative review of methods and correlates. *Patient Education and Counseling*, 84, 319–324.

Hall, J. A., Ship, A. N., Ruben, M. A., Curtin, E. M., Roter, D. L., Clever, S. L., Smith, C.C., & Pounds, K. (2014). The Test of Accurate Perception of Patients' Affect (TAPPA): An ecologically valid tool for assessing interpersonal perception accuracy in clinicians. *Patient Education and Counseling*, 94, 218–223.

Hall, J. A., Ship, A. N., Ruben, M. A., Curtin, E. M., Roter, D. L., Clever, S. L., Smith, C.C., & Pounds, K. (2015). Clinically relevant correlates of accurate perception of patients' thoughts and feelings. *Health Communication*, 30, 423–429.

Hauch, V., Sporer, S. L., Michael, S. W., & Meissner, C. A. (2014). Does training improve the detection of deception? A meta-analysis. *Communication Research*. Advance online publication.

Hurley, C. M., Anker, A. E., Frank, M. G., Matsumoto, D., & Hwang, H. C. (2014). Background factors predicting accuracy and improvement in micro expression recognition. *Motivation & Emotion*, 38, 700–714.

Ickes, W. (2001). Measuring empathic accuracy. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement* (pp. 219–241). Mahwah, NJ: Lawrence Erlbaum Associates.

Jenness, A. (1932). The effects of coaching subjects in the recognition of facial expression. *Journal of General Psychology*, 7, 163–178.

Katz, D. (1979). Obituary: Floyd H. Allport (1890–1978). *American Psychologist*, 34, 351–353.

Lopata, C., Thomeer, M. L., Volker, M. A., & Nida, R. E. (2006). Effectiveness of a cognitive-behavioral treatment on the social behaviors of children with Asperger disorder. *Focus on Autism and Other Developmental Disabilities*, 21, 237–244.

Lopata, C., Thomeer, M. L., Volker, M. A., Nida, R. E., & Lee, G. K. (2008). Effectiveness of a manualized summer social treatment program for high-functioning children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 38, 890–904.

Matsumoto, D., & Hwang, H. S. (2011). Evidence for training the ability to read microexpressions of emotion. *Motivation & Emotion*, 35, 181–191.

McKenzie, K., Matheson, E., McKaskie, K., Hamilton, L., & Murray, G. C. (2000). Impact of group training on emotion recognition in individuals with a learning disability. *British Journal of Learning Disabilities*, 28, 143–147.

Miczo, N., Segrin, C., & Allspach, L. E. (2001). Relationship between nonverbal sensitivity, encoding, and relational satisfaction. *Communication Reports*, 14, 39–48.

Nunamaker, J. F., Derrick, D. C., Elkins, A. C., Burgoon, J. K., & Patton, M. W. (2011). Embodied conversational agent-based kiosk for automated interviewing. *Journal of Management Information Systems*, 28, 17–48.

Pitterman, H., & Nowicki, S. (2004). A test of the ability to identify emotion in human standing and sitting postures: The Diagnostic Analysis of Nonverbal Accuracy-2 Posture Text (DANVA2-POS). *Genetic, Social, and General Psychology Monographs*, 130, 146–162.

Puccinelli, N., Andrzejewski, S. A., Markos, E., Noga, T., & Motyka, S. (2013). The value of knowing what customers really want: The impact of salesperson ability to read affect on service quality. *Journal of Marketing Management*, 29, 356–373.

Puccinelli, N., Motyka, S., & Grewal, D. (2010). Can you trust a customer's expression? Insights into nonverbal communication in the retail context. *Psychology & Marketing*, 27, 964–988.

Reinhard, M. A., Greifeneder, R., & Scharmach, M. (2013). Unconscious processes improve lie detection. *Journal of Personality and Social Psychology*, 105, 721.

Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., & Archer, D. (1979). *Sensitivity to nonverbal communication: The PONS test*. Baltimore: The Johns Hopkins University Press.

Ruben, M. A., Hall, J. A., Curtin, E. M., Blanch-Hartigan, D., & Ship, A.N. (2015). Discussion increases efficacy when training accurate perception of patients' affect. *Journal of Applied Social Psychology*, 45(6), 355–362.

Ruben, M. A., Hill, K. M., & Hall, J. A. (2014). How women's sexual orientation guides accuracy of interpersonal judgments of other women. *Cognition & Emotion*, 28, 1512–1521.

Rule, N. O., Tskhay, K. O., Brambilla, M., Riva, P., Andrzejewski, S. A., & Krendl, A. C. (2015). The influence of anti-gay prejudice on the categorizations of sexual orientation. *Personality and Individual Differences*, 77, 74–80.

Russell, T. A., Chu, E., & Phillips, M. L. (2006). A pilot study to investigate the effectiveness of emotion recognition remediation in schizophrenia using the micro-expression training tool. *British Journal of Clinical Psychology*, 45, 579–583.

Schroeder, J. E. (1995). Interpersonal perception skills: Self-concept correlates. *Perceptual and Motor Skills*, 80, 51–56.

Silver, M., & Oakes, P. (2001). Evaluation of a new computer intervention to teach people with autism or Asperger syndrome to recognize and predict emotions in others. *Autism*, 5, 299–316.

Silver, H., Goodman, C., Knoll, G., & Isakov, V. (2004). Brief emotion training improves recognition of facial emotions in chronic schizophrenia. A pilot study. *Psychiatry Research*, 128, 147–154.

Solomon, M., Goodlin-Jones, B. L., & Anders, T. F. (2004). A social adjustment enhancement intervention for high functioning Autism, Asperger's syndrome, and pervasive developmental disorder NOS. *Journal of Autism and Developmental Disorders*, 34, 649–668.

Statucka, M., & Walder, D. J. (2013). Efficacy of social cognition remediation programs targeting facial affect recognition deficits in schizophrenia: A review and consideration of high-risk samples and sex differences. *Psychiatry Research*, 206, 125–139.

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help? *Emotion*, 4, 46–64.

Transportation Security Administration (TSA). (2013). *Transportation Security Administration's Screening of Passengers by Observation Techniques*. Retrieved from www.oig.dhs.gov/assets/Mgmt/2013/OIG_13–91_May13.pdf

Twyman, N. W., Elkins, A. C., Burgoon, J. K., & Nunamaker, J. F. (2014). A rigidity detection system for automated credibility assessment. *Journal of Management Information Systems*, 31, 173–202.

White, S. W., Keonig, K., & Scahill, L. (2007). Social skills development in children with autism spectrum disorders: A review of the intervention research. *Journal of Autism and Developmental Disorders*, 37, 1858–1868.

Wood, P. M., & Kroese, B. S. (2007). Enhancing the emotion recognition skills of individuals with learning disabilities: A review of the literature. *Journal of Applied Research in Intellectual Disabilities*, 20, 576–579.

13 Interpersonal accuracy in relation to the workplace, leadership, and hierarchy

Marianne Schmid Mast and Ioana Latu

Abstract

In this chapter, we discuss how interpersonal accuracy – the ability to accurately assess others' states and traits – plays out in hierarchies, particularly those related to workplace and leadership. We begin by discussing the importance of interpersonal accuracy for workplace relationships, such as those with customers, co-workers, and among subordinates and superiors. Overall, the literature on this topic shows that interpersonal accuracy skills are associated with positive outcomes in workplace tasks such as sales and negotiations. Moreover, we summarize research that shows the positive outcomes associated with interpersonally accurate leaders. In the second part of the chapter, we summarize the theoretical issues and empirical research that investigate how power is related to interpersonal accuracy. Finally, we discuss research investigating the extent to which individuals are accurate in assessing others' power or hierarchical positions, as well as the verbal and nonverbal cues they use. Overall, we show that there is mostly a positive relationship between power and interpersonal accuracy in the workplace context, and we highlight the limitations and possible future directions in this line of research.

Interpersonal accuracy describes the extent to which a person is able to accurately assess others with respect to their states and traits (e.g., emotions, intentions, personality). It can be assessed either with standardized tests (testing paradigm) or in person-to-person interactions (in vivo paradigm). In the testing paradigm, different participants view a series of target individuals and are asked to infer these targets' states or traits. The test constructor possesses the information about the targets' *actual* states or traits so that participants' answers (the *inferred* states or traits) can be scored as correct or incorrect. This procedure is similar to how general mental ability is measured in a standardized intelligence test. In the in vivo paradigm, participants' inferences are scored against their interaction partners' self-reported states and traits to obtain a measure of accuracy.

Regardless of the paradigm used and the states or traits assessed, in this chapter, we focus on interpersonal accuracy measured through the comparison of a person's inferences about a target person (*perceived* or *inferred* states or traits) with a criterion or gold standard (*actual* states or traits). Interpersonal accuracy measured through a comparison with a criterion is also called performance-based interpersonal accuracy assessment. The criterion is not always easy to gain (Schmid Mast, Murphy, & Hall, 2006). Sometimes, it is an objective fact such as who is the boss of the other in a work setting. At other times, it is a self-report about how a person actually feels or what his/her personality is. The criterion can also be defined by consensus in that, for instance, third observers assess the target's personality, and the *actual* trait is the one most observers agree upon. Independent of how the criterion is assessed, in performance-based interpersonal accuracy tests, it is always assessed independently of the inference.

Intuitively, we might think that interpersonal accuracy is particularly important for close relationships. We think about a mother or a father who needs to accurately read the nonverbal cues of his/her newborn to be able to provide a fostering and caring environment for optimal upbringing. Or, we might think about how important it is to accurately assess the personality of a dating partner to know whether to pursue the relationship or not. Interpersonal accuracy certainly is crucial for close relationships but research shows that it is also an important factor in relationships that are not primarily characterized by the horizontal dimension in social interactions (i.e., friendliness, caring, agreeableness) but by the vertical dimension (i.e., dominance, power, control) (Moskowitz, 1988; Srivastava & Anderson, 2014). Relationships in the workplace are often hierarchical and the vertical dimension is prevalent. In this chapter, we will focus on the vertical dimension. We will show that interpersonal accuracy can be related to being more successful in the workplace, that interpersonal accuracy plays an important role in the leader–subordinate relationship, and that being interpersonally accurate as a leader might be beneficial. We will also shed light on the question of whether it is the people higher up or lower down in the hierarchy who are more interpersonally accurate and demonstrate that people are, in general, accurate at assessing where on the vertical dimension a target is placed.

More specifically, in the first part of the chapter we will outline the importance of interpersonal accuracy for relationships with customers, co-workers, and among subordinates and superiors. We will summarize the empirical research that suggests that interpersonal accuracy skills are associated with positive outcomes in the workplace, for example, in sales and negotiations. We will then go on to discuss the research that shows the positive outcomes associated with interpersonally accurate leaders. In a following section, we will discuss power and interpersonal accuracy under the perspective of whether more or less powerful individuals are better at assessing others accurately, and of how accurate people are in correctly assessing others' power or hierarchical positions and which cues they use to do so.

Interpersonal accuracy and workplace outcomes

To the extent that interpersonal accuracy makes a person attuned to others' thoughts and feelings and good at reading those thoughts and feelings, for certain tasks or jobs, interpersonal accuracy skills might be an advantage. Interactive service work includes jobs in which social interactions (e.g., with clients) are key for job success and being interpersonally skilled certainly is an advantage for this kind of jobs (Belt, Richardson, & Webster, 2002). Correctly reading a client's cues when presenting him/her with different options, assessing what the client wants or needs even if the client is not explicit about it, and having insight into the client's personality – these skills can help to present the product or treatment that is just right for the person. As a consequence, client satisfaction most likely will increase which, at least for the service industry, entails better job success by selling more products or having the customer return. Research shows that, for instance, interpersonally accurate car salespeople sell more cars per year (Byron, Terranova, & Nowicki, 2007), and interpersonally accurate doctors have more satisfied patients (DiMatteo, Taranta, Friedman, & Prince, 1980) and more returning patients (DiMatteo, Hays, & Prince, 1986; for a more detailed discussion about the link between interpersonal accuracy in a clinician and positive patient outcomes see Chapter 14). Confirming these results, a meta-analysis by Elfenbein, Foo, White, Tan, and Aik (2007) suggests that people with high interpersonal accuracy generally show increased job performance. Similarly, a more recent meta-analysis (Hall, Andrzejewski, & Yopchick, 2009) showed that interpersonal accuracy was positively correlated with workplace effectiveness, defined as performance rated by superiors, peers, or senior staff.

In sum, when job performance depends directly on the quality of the social relationship with the customer or the client, interpersonal accuracy seems to be a plus. To the extent that a person can sense the other correctly and then adapt his/her behavior and communication, the other person most likely is more satisfied with the service or more willing to buy the product. Whether a person's interpersonal accuracy also affects the performance of the people who work with the interpersonally accurate person (e.g., team colleagues or subordinates) seems less clear and is discussed in the remainder of this section.

For a person's interpersonal accuracy to affect colleagues or subordinates, the path might be more indirect in that interpersonal accuracy affects communication, collaboration, and coordination in work teams. Task-relevant information is sometimes expressed through subtle nonverbal cues and correctly picking up on them is important. To illustrate, if a task-competent team member has serious doubts about the appropriateness of a leader's decision but does not dare contradict the superior, his/her disagreement or doubt might still be conveyed through nonverbal cues. If the leader or the other team members do not correctly assess those cues and follow up on them, wrong decisions can mislead whole work teams. Such a dynamic can have serious negative consequences for team performance, especially if the concerns that should have been raised were well-founded. Research seems to support that interpersonal accuracy can be an advantage for dyadic or group outcomes in negotiation tasks. In a seller–buyer negotiation, the seller's but not the buyer's emotion recognition ability was related to better negotiation outcomes for both (Elfenbein et al., 2007). In the same vein, in a recruiter–job applicant negotiation, better emotion recognition accuracy of the recruiter was related to better joint gains (Schlegel, 2013).

Particularly relevant for negotiations is the concept of emotional intelligence that partially overlaps with interpersonal accuracy, in that one aspect of emotional intelligence is correctly assessing other people's emotions. Emotional intelligence is, however, a much broader construct including more than just accurate emotion recognition in others (Mayer, Salovey, & Caruso, 2004). In one study, negotiators with higher levels of emotional intelligence obtained better dyadic outcomes when the outcome was contingent upon the negotiation performance (Kong, Bottom, & Konczak, 2011). Emotional intelligence was also related to the dyadic partners' increased individual negotiation outcomes (Foo, Elfenbein, Tan, & Aik, 2004). It has to be noted, however, that in the last two studies, emotional intelligence was assessed with a self-report questionnaire, which does not correspond to the standard for performance-based assessment of interpersonal accuracy we defined at the outset of the chapter.

Interpersonally accurate leaders

Interpersonal accuracy might be especially important for leaders. Leaders spend much of their working time in interactions with other people (Kotter, 1999) and in particular with their subordinates. This heavy reliance on interpersonal interactions suggests that being interpersonally accurate might play an important role for leaders. To the extent that accuracy influences outcomes not only for the person who is accurate as we have seen in the previous paragraphs, the positive effects of interpersonally accurate leaders likely cascade down to the subordinates.

A leader allocates different tasks to different members of his/her team, and choosing the right person for the right job at the right time seems crucial for a leader to be effective in his/her leadership task. As an example, if a leader accurately assesses that one of his/her subordinates has been particularly irritated and stressed over the past weeks, the leader will most likely not add responsibility for another important task to this particular subordinate. This would not only be in the best of interests for this particular subordinate but also crucial for assuring the optimal performance output from the entire work team.

Mostly, the concept of emotional intelligence (which overlaps some with interpersonal accuracy) has been studied in the realm of leadership (Caruso, Mayer, & Salovey, 2002). Emotional intelligence is often measured by self-report questionnaires and not with performance-based tests (e.g., Palmer, Walls, Burgess, & Stough, 2001). Interpersonal accuracy as a measurable performance has rarely been studied in the leadership context. Adhering to such performance-based interpersonal accuracy tests, we review literature that has postulated or investigated the link between interpersonal accuracy and leadership.

To date, we do not have a very good understanding of whether and how interpersonal accuracy in leaders affects subordinate satisfaction and/or team performance. While there is some evidence suggesting that more interpersonally accurate leaders have more satisfied subordinates, the link between leader interpersonal accuracy and team performance is not well investigated.

With respect to the question of whether interpersonally accurate leaders have more satisfied subordinates, one can argue that when a person is assessed accurately, he/she will most likely feel understood and cared for. Research shows that feeling understood is related to better subjective well-being (Oishi, Krochik, & Akimoto, 2010), which could explain why subordinates with interpersonally accurate bosses are more satisfied. Research indeed shows that interpersonally accurate leaders have more satisfied subordinates. Two studies show supporting evidence. First, participants in a laboratory experiment were paired, and the roles of a leader and a subordinate were randomly assigned to them. These leader–subordinate dyads interacted on a problem-solving task after which the subordinate indicated how satisfied he/she was with the leader. The leader's interpersonal accuracy was measured after the interaction with a standardized test in which the thoughts and feelings of a target person on video had to be inferred (empathic accuracy paradigm, described in more detail in Chapter 3). Results showed that leaders who were more interpersonally accurate had more satisfied subordinates (Schmid Mast, Jonas, Cronauer, & Darioly, 2012). To the extent that collaborator satisfaction is related to better job performance, being interpersonally accurate offers leaders an important advantage. Second, when actual managers were investigated, female managers with better ability to accurately read others' nonverbal expressions had more satisfied subordinates (Byron, 2007).

Although research shows that better subordinate satisfaction is related to better subordinate performance (Petty, McGee, & Cavender, 1984), it would be necessary to investigate whether there is a direct link between leader interpersonal accuracy and leadership effectiveness in terms of group or subordinate performance. On a theoretical level, such a link is often postulated. For instance, transformational leadership behavior is related to better leadership effectiveness and group productivity (Lowe, Kroeck, & Sivasubramaniam, 1996). Transformational leadership includes leader behavior such as showing inspirational motivation (providing a vision and inspiring and motivating the subordinates), intellectual stimulation (fostering innovation and creativity of the subordinates), idealized influence (being a trusted and admired role model), and individualized consideration (having personalized interactions with subordinates, teaching and coaching them) (Bass & Avolio, 1994). The latter, individualized consideration, means that "individual differences in terms of needs and desires are recognized" (Bass & Riggio, 2006, p. 7) in subordinates. The correct recognition of the individual differences in needs and desires among the subordinates is what we understand by interpersonal accuracy. In other words, the leadership style that is currently seen as most effective includes as one dimension the idea of interpersonal accuracy. Moreover, research supports that there is a link between transformational leadership style and certain aspects of interpersonal accuracy (e.g., emotion recognition) (Rubin, Munz, & Bommer, 2005). In addition, female leaders who were better at correctly assessing others' emotions obtained better performance ratings from their superiors (Byron, 2007). Concerning the latter study, subjective performance ratings are affected by liking and therefore the performance ratings might be more of a manifestation of satisfaction than performance. Clearly, there is more research needed to investigate whether teams and subordinates led by interpersonally accurate leaders perform better, and those performance measures should preferably be objective ones. Research shows that dyadic decision making on an objetive performance task (hidden profile task) was better when the superiors engaged in perspective taking because of increased sharing of critial information (Galinsky, Magee, Rus, Rothman, & Todd, 2014). Whether such decision making tasks would profit also from a superior with more interpersonal accuracy remains to be tested.

Hierarchy and interpersonal accuracy

There has been a longstanding debate in the literature regarding whether individuals high or low in power are more interpersonally accurate. In other words, are powerful people or powerless people better at reading others? We base our report on a meta-analysis conducted by Hall, Schmid Mast, and Latu (2015), in which more than 90 studies were analyzed. The studies differed both in their definition of power/hierarchy and the accuracy paradigm that was used.

Power and hierarchy can take different forms including structural power, leadership, status, or dominance. Despite their differences, these terms all indicate the amount of control or influence one has (or strives to have) in a relationship or a larger context such as a company or organization. Structural power is equivalent to formal authority and it is often related to social or occupational positions (Ellyson & Dovidio, 1985). Leadership is a related term, but refers more precisely to the role and ability to influence others in order to achieve a common goal (Bass, 1960). Status as related to hierarchical relations refers to being a member of a specific social group, such as being a man versus a woman, or having high or low socioeconomic status (SES, Pratto, Sidanius, Stallworth, & Malle, 1994), but it can also be derived from the respect an invidual is awarded by group members. Dominance is an individual difference (personality trait) that describes a person's tendency to strive for having control over other people (Ellyson & Dovidio, 1985). Dominance can also be seen as a behavior that is aimed at social control (Schmid Mast, 2010).

In terms of the accuracy paradigm, researchers have used either a *testing paradigm*, in which participants infer others' emotions or thoughts based on videos, photographs, or other recorded material, or an *in vivo paradigm* in which participants infer others' emotions or thoughts after actually interacting with them. Testing paradigms often use tests such as the Profile of Nonverbal Sensitivity (PONS; Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979) or the Diagnostic Analysis of Nonverbal Accuracy (DANVA; Nowicki & Duke, 1994). In the testing paradigms, accuracy is scored against a correct, established criterion. In the in vivo paradigm, accuracy is scored against the partner's self-ratings of his/her states or traits. Importantly, accuracy paradigms also differ in whether they focus on how people *infer* the states of others or how people *recall* information about others.

Although we focus primarily on the direction of the relationship between power and interpersonal accuracy, the current section will also include a discussion of other theoretical issues related to the literature, such as causality and locus of accuracy.

Causality

Although significant correlations between power and accuracy are reported in the literature, it is difficult to establish the causal direction of the relationship. Is interpersonal accuracy a skill that develops once a person becomes a leader or do people become leaders because they are interpersonally accurate, or both? Possessing high interpersonal accuracy may help people rise to the top. For example, Johnson and Bechler (1998) showed that group members who had effective listening skills (operationalized by interest and attention to other people in the group) were more likely to emerge as leaders of their small groups during a decision making discussion. Walter, Cole, van der Vegt, Rubin, and Bommer (2012) showed that individuals with high levels of emotion recognition ability combined with high extraversion were more likely to emerge as leaders in groups. The two aspects, extraversion and interpersonal accuracy, facilitate task coordination in groups, which explains why extraverted and interpersonally accurate individuals emerge as leaders. Thus, interpersonal accuracy may cause people to achieve power in groups. However, the opposite may also be true – that granting someone power may determine him/her to be more or less interpersonally accurate, which is consistent with the social psychological view of power as a situational force.

Locus of accuracy

A second theoretical issue is that of the locus of accuracy. If power does influence interpersonal accuracy, is it having power or lacking power that makes people more or less accurate? For example, if we focus on the view that power and accuracy are positively related, is it because high-power people (leaders) have an advantage in terms of accuracy or is it because low-power people (subordinates) are disadvantaged in terms of their accuracy?

Is it the experience of having or lacking power that makes people more or less interpersonally accurate? Both of these possibilities are discussed below.

Direction of the power–interpersonal accuracy relationship

The analysis of the existing literature revealed that there is no clear advantage for individuals high or low in power in terms of interpersonal accuracy. In fact, studies found support for both hypotheses. We discuss the theoretical explanations and empirical findings for both points of view before presenting the results of the meta-analysis.

Powerless people are more accurate than powerful people. One may predict that low power people possess more interpersonal accuracy for adaptive purposes (Henley, 1977; Thomas, Franks, & Calonico, 1972). For example, subordinates would want to discern their superiors' states (intention, moods, desires), because by having this knowledge subordinates can adapt their behaviors in order to please their superior or simply to make interactions with the superior smoother.

It can also be that superiors are especially low in accuracy, because of motivational and cognitive factors. For example, given that superiors control resources in their environment, they may not be motivated to learn their subordinate's thoughts or feelings. In addition, given the high cognitive demands imposed by their role, they may not have the necessary cognitive resources to do so (Fiske, 1993; Russell & Fiske, 2010).

The view that low-power individuals are more interpersonally accurate than high-power individuals was supported by several studies that conceptualized power as situationally induced power (Galinsky, Magee, Inesi, & Gruenfeld, 2006), personality dominance (Moeller, Ewing Lee, & Robinson, 2011), and SES (Kraus, Côté, & Keltner, 2010).

Powerful people are more accurate than powerless people. On the other hand, high-power people may be more accurate than low-power people, because accuracy is an important skill in leading other people. In fact, Schmid Mast, Jonas, and Hall (2009) showed that high-power people were more accurate because they felt more pride and respect, suggesting another-person orientation. This is consistent with Hall and Halberstadt (1997) and Hall, Halberstadt, and O'Brien (1997), who suggested that superiors needed to be accurate in order to maintain the respect and support of their followers. In fact, in the leadership domain (Riggio, 2001), it is proposed that it is this interpersonal skill that actually leads people to rise in the leadership ranks.

The point of view that high-power people are more accurate than the low-power people was supported by several studies by Schmid Mast et al. (2009), which found across different manipulations of situationally induced power and across different interpersonal accuracy tests that higher power led to more interpersonal accuracy. This finding was replicated for power as actual rank within an organization: in one study conducted in a factory setting, managers were more interpersonally accurate compared to factory workers (Zhong, Zhang, & Chen, 2013). In the same vein, Schmid Mast and Darioly (2014) found that leaders in different organizations and domains scored higher on a standardized emotion recognition accuracy task than their subordinates.

Meta-analysis findings. Overall, individual studies do not offer a conclusive picture of the relationship between the position in the hierarchy and interpersonal accuracy. The meta-analysis by Hall and colleagues (2015) offers some clarification regarding certain patterns. First, meta-analytical findings showed a significant relationship between SES and the accuracy of making inferences, such that higher SES was related to higher accuracy. A second pattern emerged for the effects of experimentally manipulated power on recall of interpersonal information (mostly spoken words or written utterances) in the testing paradigm. Put differently, high-power people were better at recalling other people's words.

The meta-analysis also showed great heterogeneity across studies, which indicates the presence of moderators. Indeed, given the lack of consistency in the findings, it may be that the relationship between power/hierarchy and interpersonal accuracy is moderated by several situational and individual factors. For example, Schmid Mast et al. (2009) found that people with an empathic mindset were more accurate than those with an egoistic mindset. Meta-analysis and moderation analysis confirmed this idea and showed that the type of personality dominance (prosocial/responsible versus egoistic/aggressive) was a significant moderator. Prosocial/responsible personality dominance was related to higher interpersonal accuracy in the testing paradigm, whereas egoistic/aggressive personality dominance was related to lower interpersonal accuracy in the testing paradigm.

The meta-analytic result for the in vivo studies showed low-power people to be more accurate than high-power people. However, it is difficult to interpret this finding given that accuracy in this paradigm is, in most studies, confounded between the perceiver's perceptivity and the target-partner's expressivity. For example, the higher accuracy of subordinates may be not only due to their capacity to decode their superiors but also due to the superiors' ability to express their feelings and intentions in a clear, recognizable way.

Several studies directly investigated this possibility by controlling for the target's expressivity. For example, Alkire, Collum, Kaswan, and Love

(1968) found that lower-power individuals (sorority pledges) were less accurate than high-power individuals (sorority members), but this was due to high-power individuals being low in expressive clarity. Studies that found the opposite pattern but controlled for targets' expressivity suggested the same: when low-power individuals showed an accuracy advantage, this was accounted for by the expressive clarity of the partner (Hall, Rosip, Smith LeBeau, Horgan, & Carter, 2006; Snodgrass, Hecht, & Ploutz-Snyder, 1998).

Overall, few clear patterns emerged even when analyzing studies within a meta-analysis. Moreover, with some exceptions (e.g., Kraus et al. 2010, Study 1; Zhong et al., 2013, Study 2), the studies that fit the inclusion criteria for this meta-analysis did not investigate the relationship between accuracy and power in the workplace setting. We suggest that future studies should investigate this relationship directly in the workplace domain, while investigating both mediators and possible moderators of the power-interpersonal accuracy link.

Accurately assessing hierarchies and power

When considering power and interpersonal accuracy, there is not only the question as to whether it is the powerful or the powerless who are better at accurately assessing others, but one can also ask whether people in general are able to accurately judge others' power. Can we tell who is the boss when we enter a room of people whom we do not know? Research suggests that we can. For example, individuals are able to accurately tell which of two people in a photograph is the other's boss when the leader poses with his/her subordinate side by side (Barnes & Sternberg, 1989). In addition, when university employees with an actual status difference among them (e.g., the professor and his/her postdoctoral student) were photographed at a random moment during their interaction, outside observers were able to accurately judge their status difference (Schmid Mast & Hall, 2004), and observers were able to assess targets' assertiveness in videotaped get-acquainted interactions at better than chance level (Schmid Mast, Hall, Murphy, & Colvin, 2003). Moreover, individuals were able to accurately identify emergent leaders in a hierarchy based on records of the group meetings containing only verbal content, nonverbal content, or both (Stein, 1975).

Indeed, in order to assess power in unknown others, people use a number of different verbal and nonverbal cues. Hall and colleagues (2005) conducted a meta-analysis looking at which nonverbal cues observers use when judging others' status or power. Individuals were perceived as more dominant or higher status when they showed more gazing, lowered eyebrows, a more expressive face, more nodding, less self-touch, more other touch, more gestures, more bodily openness, more erect or tense posture, more body or leg shifts, smaller interpersonal distance, a more variable voice, a louder voice, more interruptions, less pausing, a faster speech rate, a lower voice pitch, and more vocal relaxation. Moreover, there is a strong positive relation between speaking time and perceived status (Schmid Mast, 2002), and observers use the visual dominance ratio (the percentage of looking while speaking relative to the percentage of looking while listening) as an indicator of high status (Dovidio, Ellyson, Keating, Heltman, & Brown, 1988).

In their meta-analysis, Hall et al. (2005) also investigated which nonverbal cues were the ones related to actual power or status. Only a fraction of the aforementioned cues were actually indicative of people's actual standing on the vertical axis. High-power people use more open body positions, have closer interpersonal distances to others, are more facially expressive, speak more loudly, engage in more successful interruptions, and have less vocal variability compared to lower-power people. They also talk more and use more visual dominance (Dovidio et al., 1988; Schmid Mast, 2002).

Interestingly, despite the only partial overlap of how perceivers use others' cues to assess their dominance and how these cues are related to actual dominance, people are on average accurate at assessing dominance in others. To assess relative status in photographs, observers used different cues for assessing status in women than in men, and different cues were indicative of actual status (Schmid Mast & Hall, 2004). For women, perceivers used downward head tilt to infer high power (among other cues) and this cue was also indicative of the women's actual power. For men, observers used formal dress as cue (among other cues) to infer power and this cue again was related to actual power in men.

If power is accurately inferred in social interactions, this has important implications for the workplace and leadership domains. It can help to prevent social faux-pas and make social interactions smoother and it serves to maintain hierarchies which – in certain situations – can be beneficial for team productivity (Frauendorfer, Schmid Mast, Sanchez-Cortes, & Gatica-Perez, 2015). In fact, work on motivated accuracy suggests that people are motivated to be accurate about their own position in the hierarchy because otherwise they would face severe social costs (Srivastava & Anderson, 2014). For example, individuals who overestimated their status within a group (status self-enhancers) were liked less by other group members and, importantly, were paid less for their work (Anderson, Ames, & Gosling, 2008).

Conclusions

Interpersonal accuracy is undoubtedly an asset in the workplace. Empirical research suggests that performance-based interpersonal accuracy predicts many positive outcomes in the workplace. Interpersonally accurate individuals make better leaders, salespeople, and negotiators. Interpersonally accurate leaders tend to have more satisfied collaborators and subordinates. Moreover, when power and dominance are construed as other-orientation (e.g., transformational leadership or prosocial/responsible personality dominance), it is associated with more interpersonal accuracy, suggesting not only that power in hierarchies may lead people to be more interpersonally accurate (Schmid Mast et al., 2009), but also that interpersonal accuracy may be a skill that helps people rise in the hierarchy (Walter et al., 2012). However, there are still many inconsistencies in the literature on power and interpersonal accuracy (Hall et al., 2015), and too few studies that directly address the relationship between interpersonal accuracy, leadership, and workplace outcomes. We suggest that future studies should directly investigate these relationships both in laboratory and in field studies.

References

Alkire, A. A., Collum, M. E., Kaswan, J., & Love, L. R. (1968). Information exchange and accuracy of verbal communication under social power conditions. *Journal of Personality and Social Psychology*, 9, 301–308.

Anderson, C., Ames, D. R., & Gosling, S. D. (2008). Punishing hubris: The perils of overestimating one's status in a group. *Personality and Social Psychology Bulletin*, 34, 90–101.

Barnes, M. L., & Sternberg, R. J. (1989). Social intelligence and decoding of nonverbal cues. *Intelligence*, 13, 263–287.

Bass, B. M. (1960). *Leadership, psychology, and organizational behavior*. Oxford, UK: Harper.

Bass, B. M., & Avolio, B. J. (1994). *Improving organizational effectiveness through transformational leadership*. Thousand Oaks, CA: Sage.

Bass, B. M., & Riggio, R. E. (2006). *Transformational leadership*. Mahwah, NJ: Lawrence Erlbaum.

Belt, V., Richardson, R., & Webster, J. (2002). Women, social skill and interactive service work in telephone call centres. *New Technology, Work and Employment*, 17, 20–34.

Byron, K. (2007). Male and female managers' ability to read emotions: Relationships with supervisor's performance ratings and subordinates' satisfaction ratings. *Journal of Occupational and Organizational Psychology*, 80, 713–733.

Byron, K., Terranova, S., & Nowicki, S. J. (2007). Nonverbal emotion recognition and salespersons: Linking ability to perceived and actual success. *Journal of Applied Social Psychology*, 37, 2600–2619.

Caruso, D. R., Mayer, J. D., & Salovey, P. (2002). Emotional intelligence and emotional leadership. In R. E. Riggio, S. E. Murphy, & F. J. Pirozzolo (Eds.), *Multiple intelligences and leadership. LEA's organization and management series* (pp. 55–74). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

DiMatteo, M. R., Hays, R. D., & Prince, L. M. (1986). Relationship of physicians' nonverbal communication skills to patient satisfaction, appointment noncompliance, and physician workload. *Health Psychology*, 5, 581–594.

DiMatteo, M. R., Taranta, A., Friedman, H. S., & Prince, L. M. (1980). Predicting patient satisfaction from physicians' nonverbal communication skills. *Medical Care*, 18, 376–387.

Dovidio, J. F., Ellyson, S. L., Keating, C. F., Heltman, K., & Brown, C. E. (1988). The relationship of social power to visual displays of dominance between men and women. *Journal of Personality and Social Psychology*, 54, 233–242.

Elfenbein, H. A., Foo, M. D., White, J., Tan, H. H., & Aik, V. C. (2007). Reading your counterpart: The benefit of emotion recognition accuracy for effectiveness in negotiation. *Journal of Nonverbal Behavior*, 31, 205–223.

Ellyson, S. L., & Dovidio, J. F. (1985). Power, dominance, and nonverbal behavior: Basic concepts and issues. In S. L. Ellyson & J. F. Dovidio (Eds.), *Power, dominance, and nonverbal behavior* (pp. 1–27). New York: Springer.

Fiske, S. T. (1993). Controlling other people: The impact of power on stereotyping. *American Psychologist*, 48, 621–628.

Foo, M. D., Elfenbein, H. A., Tan, H. H., & Aik, V. C. (2004). Emotional intelligence and negotiation: The tension between creating and claiming value. *International Journal of Conflict Management*, 15, 411–429.

Frauendorfer, D., Schmid Mast, M., Sanchez-Cortes, D., & Gatica-Perez, D. (2015). Emergent power hierarchies and group performance. *International Journal of Psychology*. Manuscript submitted for publication.

Galinsky, A. D., Magee, J. C., Inesi, M. E., & Gruenfeld, D. H. (2006). Power and perspectives not taken. *Psychological Science*, 17, 1068–1074.

Galinsky, A. D., Magee, J. C., Rus, D., Rothman, N. B., & Todd, A. R. (2014). Acceleration with steering: The synergistic benefits of combining power and perspective-taking. *Social Psychological and Personality Science*, 5, 627–635.

Hall, J. A., Andrzejewski, S. A., & Yopchick, J. E. (2009). Psychosocial correlates of interpersonal sensitivity: A meta-analysis. *Journal of Nonverbal Behavior*, 33, 149–180.

Hall, J. A., Coats, E. J., & Smith LeBeau, L. (2005). Nonverbal behavior and the vertical dimension of social relations: A meta-analysis. *Psychological Bulletin*, 131, 898–924.

Hall, J. A., & Halberstadt, A. G. (1997). Subordination and nonverbal sensitivity: A hypothesis in search of support. In M. R. Walsh (Ed.), *Women, men, and gender: Ongoing debates* (pp. 120–133). New Haven: Yale University Press.

Hall, J. A., Halberstadt, A. G., & O'Brien, C. E. (1997). "Subordination" and nonverbal sensitivity: A study and synthesis of findings based on trait measures. *Sex Roles*, 37, 295–317.

Hall, J. A., Rosip, J. C., Smith LeBeau, L., Horgan, T. G., & Carter, J. D. (2006). Attributing the sources of accuracy in unequal-power dyadic communication: Who is better and why? *Journal of Experimental Social Psychology*, 42, 18–27.

Hall, J. A., Schmid Mast, M., & Latu, I. (2015). The vertical dimension of social relations and accurate interpersonal perception: A meta-analysis. *Journal of Nonverbal Behavior*, 39, 131–163.

Henley, N. M. (1977). *Body politics: Power, sex, and nonverbal communication*. Englewood Cliffs, NJ: Prentice-Hall.

Kong, D., Bottom, W. P., & Konczak, L. (2011, July). *Re-examining the role of emotional intelligence in negotiations*. Paper presented at the IACM 24TH Annual Conference.

Kotter, J. P. (1999). *On what leaders really do*. Cambridge, MA: Harvard Business Press.

Kraus, M. W., Côté, S., & Keltner, D. (2010). Social class, contextualism, and empathic accuracy. *Psychological Science*, 21, 1716–1723.

Lowe, K. B., Kroeck, K. G., & Sivasubramaniam, N. (1996). Effectiveness correlates of transformational and transactional leadership: A meta-analytic review of the MLQ literature. *The Leadership Quarterly*, 7, 385–425.

Mayer, J. D., Salovey, P., & Caruso, D. R. (2004). Emotional intelligence: Theory, findings, and implications. *Psychological Inquiry*, 15, 197–215.

Moeller, S. K., Ewing Lee, E. A., & Robinson, M. D. (2011). You never think about my feelings: Interpersonal dominance as a predictor of emotion decoding accuracy. *Emotion*, 11, 816–824.

Moskowitz, D. S. (1988). Cross-situational generality in the laboratory: Dominance and friendliness. *Journal of Personality and Social Psychology*, 54, 829.

Nowicki, S., Jr., & Duke, M. (1994). Individual differences in the nonverbal communication of affect: The Diagnostic Analysis of Nonverbal Accuracy Scale. *Journal of Nonverbal Behavior*, 18, 9–35.

Oishi, S., Krochik, M., & Akimoto, S. (2010). Felt understanding as a bridge between close relationships and subjective well-being: Antecedents and consequences across individuals and cultures. *Social and Personality Psychology Compass*, 4, 403–416.

Palmer, B., Walls, M., Burgess, Z., & Stough, C. (2001). Emotional intelligence and effective leadership. *Leadership & Organization Development Journal*, 22, 5–10.

Petty, M. M., McGee, G. W., & Cavender, J. W. (1984). A meta-analysis of the relationships between individual job satisfaction and individual performance. *Academy of Management Review*, 9, 712–721.

Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of Personality and Social Psychology*, 67, 741–763.

Riggio, R. E. (2001). Interpersonal sensitivity research and organizational psychology: Theoretical and methodological applications. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement* (pp. 305–317). Mahwah, NJ: Lawrence Erlbaum Associates.

Rubin, R. S., Munz, D. C., & Bommer, W. H. (2005). Leading from within: The effects of emotion recognition and personality on transformational leadership behavior. *Academy of Management Journal*, 48, 845–858.

Russell, A. M., & Fiske, S. T. (2010). Power and social perception. In A. Guinote & T. K. Vescio (Eds.), *The social psychology of power* (pp. 231–250). New York: Guilford.

Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., & Archer, D. (1979). *Sensitivity to nonverbal communication: The PONS test*. Baltimore, MD: The Johns Hopkins University Press.

Schlegel, K. (2013). *Improving the measurement of emotion recognition ability*. (Unpublished doctoral dissertation). University of Geneva, Switzerland.

Schmid Mast, M. (2002). Dominance as expressed and inferred through speaking time: A meta-analysis. *Human Communication Research*, 28, 420–450.

Schmid Mast, M. (2010). Interpersonal behaviour and social perception in a hierarchy: The interpersonal power and behaviour model. *European Review of Social Psychology*, 21, 1–33.

Schmid Mast, M., & Darioly, A. (2014). Emotion recognition accuracy in hierarchical relationships. *Swiss Journal of Psychology*, 73, 69–75.

Schmid Mast, M., & Hall, J. A. (2004). Who is the boss and who is not? Accuracy of judging status. *Journal of Nonverbal Behavior*, 28, 145–165.

Schmid Mast, M., Murphy, N. A., & Hall, J. A. (2006). A brief review of interpersonal sensitivity: Measuring accuracy in perceiving others. In D. Chadee & J. Young (Eds.), *Current themes in social psychology* (pp. 163–185). Trinidad, CO: University of the West Indies Press.

Schmid Mast, M., Hall, J. A., Murphy, N. A., & Colvin, C. R. (2003). Judging assertiveness in female and male targets. *Facta Universitatis*, 2, 731–743.

Schmid Mast, M., Jonas, K., Cronauer, C. K., & Darioly, A. (2012). On the importance of the superior's interpersonal sensitivity for good leadership. *Journal of Applied Social Psychology*, 42, 1043–1068.

Schmid Mast, M., Jonas, K., & Hall, J. A. (2009). Give a person power and he or she will show interpersonal sensitivity: The phenomenon and its why and when. *Journal of Personality and Social Psychology*, 97, 835–850.

Snodgrass, S. E., Hecht, M. A., & Ploutz-Snyder, R. (1998). Interpersonal sensitivity: Expressivity or perceptivity? *Journal of Personality and Social Psychology*, 74, 238–249.

Srivastava, S., & Anderson, C. (2011). Accurate when it counts: Perceiving power and status in social groups. In J. L. Smith, W. Ickes, J. A. Hall, & S. D. Hodges (Eds.), *Managing interpersonal sensitivity: Knowing when – and when not – to understand others* (pp. 41–58). New York: Novinka/Nova Science.

Stein, R. T. (1975). Identifying emergent leaders from verbal and nonverbal communications. *Journal of Personality and Social Psychology*, 32, 125.

Thomas, D. L., Franks, D. D., & Calonico, J. M. (1972). Role-taking and power in social psychology. *American Sociological Review*, 37, 605–614.

Walter, F., Cole, M. S., van der Vegt, G. S., Rubin, R. S., & Bommer, W. H. (2012). Emotion recognition and emergent leadership: Unraveling mediating mechanisms and boundary conditions. *The Leadership Quarterly*, 23, 977–991.

Zhong, Y., Zhang, S., & Chen, Y. (2013). Differences of interpersonal sensitivity between high-power people and low-power people. *Journal of Clinical Psychology*, 21, 62–65.

14 Interpersonal accuracy in the clinical setting

Mollie A. Ruben

Abstract

Interpersonal accuracy in the clinical setting is the ability of clinicians to accurately assess their patients. Patients can be accurately assessed on a wide range of states and traits including but not limited to patients' psychological and physical states (e.g., anxiety and pain). There are many ways to measure interpersonal accuracy and methods for quantifying interpersonal accuracy that will be discussed. Interpersonal accuracy among clinicians predicts several positive patient outcomes such as increased patient satisfaction and adherence. Interpersonal accuracy is an important clinical skill, which builds trust within the doctor–patient relationship. Interpersonal accuracy allows clinicians not only to build trusting relationships with their patients, but also to better diagnose and treat their patients' symptoms, improving both patients' perceptions of care and their overall health.

Interpersonal accuracy in the clinical setting will be discussed in this chapter as the ability of clinicians – nurses, doctors, clinicians, therapists, and caregivers – to accurately assess their patients' states and traits. Although patients can and do accurately assess their physicians' states and traits, this chapter is focused solely on the abilities of clinicians. This approach is in line with the patient-centered model of care, in which it is the patient's perspective that matters most (Institute of Medicine Committee on Quality Health Care in America, 2001; McWhinney, 1989). Clinicians are increasingly expected to really listen to their patients' needs and concerns and respond to patients' emotions accordingly. However, empirical research on interpersonal accuracy in clinical encounters has not been conducted in proportion to its acknowledged importance as a skill.

This chapter will discuss:

(1) What can clinicians be accurate about?

(2) How is interpersonal accuracy measured in clinical contexts?

(3) What types of interpersonal accuracy have been researched in the clinical context?
(4) What makes a clinician more or less interpersonally accurate?
(5) What patient outcomes are associated with having an interpersonally accurate clinician?

What can clinicians be accurate about?

Clinicians can be accurate at interpreting patients on a wide range of states and traits, including emotions, desires, truthfulness, intentions, needs, pain, fatigue, personality, attitudes, beliefs, and values. Clinicians in medicine obviously also need to be good at noticing and interpreting the signs that are relevant to the diagnosis of mental and physical disease. This chapter will not be concerned with the accuracy of clinical diagnosis except as it pertains to the recognition of psychosocial and emotional states.

Clinicians can base their interpersonal accuracy on many kinds of cues. The cues can be visible, such as the patient's smiling, posture, hand movements, or self-touching. The cues can be heard, either in vocal quality (rate, pitch, loudness, dysfluencies, pauses, etc.) or perceived through linguistic cues both blatant and subtle. Patients' needs, desires, emotions, etc. may not be conveyed clearly through their words, but may require the clinician to ferret out the meaning by "reading between the lines," picking up on dropped hints, and so forth.

There are two basic types of interpersonal accuracy. The first type is simply to notice (and, relatedly, remember) the other person's appearance, words, or nonverbal behavior. This type of interpersonal accuracy has not been studied much in the clinical context. One study found that female doctors shown standardized videotaped vignettes of actors portraying patients with coronary heart disease recalled more patient cues overall compared to male doctors (Adams et al., 2008). The second kind of interpersonal accuracy involves accuracy in interpreting cues. Most of this chapter will focus on the latter type of interpersonal accuracy.

How is interpersonal accuracy measured in the clinical context?

To measure interpersonal accuracy, it is crucial to know whether a clinician has perceived the abovementioned characteristics of their patients. To measure this, researchers must ascertain what clinicians have perceived. Research in the clinical context on understanding patients often omits the perceiving aspect and only measures the behavioral aspect, that is, how clinicians behaviorally respond to their patients. For example, a patient could be distressed because they were just laid off from their job and they are experiencing flu-like symptoms. If the clinician was not aware of the distress associated with their unemployment, they may act in a compassionate manner around the patient's flu symptoms, but miss a crucial opportunity to address and respond to another stressor that could compound the impact of the flu. Thus, when measuring interpersonal accuracy it is important to make a distinction between clinicians' noticing patient cues and interpreting them accurately, on the one hand, and responding behaviorally to patient cues, on the other. This distinction becomes especially relevant when the clinician's behavioral response is nonoptimal – for example, when the patient drops hints of emotional distress but the clinician lets this "empathic opportunity" pass without responding (e.g., Blanch-Hartigan, 2013; Zimmermann, Del Piccolo, & Finset, 2007). In that case, it is important to know whether the clinician failed to perceive or interpret accurately, or instead was accurate but did not know how to respond or willfully chose not to respond. Only by separately investigating perception and behavior can we fully understand the communication process and ultimately important outcomes such as satisfaction, adherence, and health itself.

Interpersonal accuracy in the medical context can be assessed in many ways; however, there are limitations of each method that should be acknowledged.

Self-assessment of interpersonal accuracy. One method for measuring interpersonal accuracy is asking clinicians how accurate they think they are. Unfortunately, perceptions of one's own interpersonal accuracy are often not related to actual accuracy. In two meta-analyses on the correlation between self-assessments of interpersonal accuracy and accuracy as measured using psychometric tests, there were significant positive correlations but the effect sizes were fairly weak (weighted mean effect size r = .21 for clinicians and weighted mean effect size r = .13 for unspecified population), meaning that self-assessments cannot be substituted for empirically assessed skill (Blanch-Hartigan, 2011a; Hall, Andrzejewski, & Yopchick, 2009). In one study that tested the efficacy of various training components on emotion recognition accuracy using a standardized test of emotion recognition, the Patient Emotion Cue Test (PECT; Blanch-Hartigan, 2011b), participants were asked to rate how well the following statements described themselves, "I am good at recognizing the emotions of others," "I am confident in my ability to recognize the emotions of others," and "I found the tasks today to be fairly easy." Results showed no relationship between the objective measure of accuracy (i.e., participants' scores on the emotion recognition test) and their self-reported confidence in their ability to detect emotions (Blanch-Hartigan, 2012). In other research, physicians and medical students are generally poor at self-assessing, especially when the self-assessment involves communication skills (Blanch-Hartigan, 2011a; Davis et al., 2006). In a meta-analysis, medical students were more likely to overestimate their performance on communication skills than they were to overestimate their performance on knowledge-based measures testing their technical skill or expertise (Blanch-Hartigan, 2011a).

In vivo method. The in vivo method assesses accuracy between two interacting people usually immediately after an interaction (Gulbrandsen et al., 2012; Hall, Rosip, Smith LeBeau, Horgan, & Carter, 2006; Snodgrass, Hecht, & Ploutz-Snyder, 1998; Street & Haidet, 2010). The in vivo method is more ecologically valid than other methods as it assesses accuracy within a live interaction rather than using a standardized test or asking participants to estimate their accuracy. Most in vivo measurements ask the clinician and the patient identical questions about the patient's feelings, attitudes, or some other attribute related to health. There are several ways to analyze this type of data that are discussed in a later section.

There are two major limitations of using the in vivo method. First, measuring in vivo interpersonal accuracy can intrude on any type of interaction but especially a medical interaction. Minimally, an in vivo accuracy study would require both the clinician and the patient to fill in questionnaires in order to gather each party's perspective on the patient's states or traits.

Second, researchers must disambiguate the ability to accurately perceive an interaction partner from that interaction partner being a good expressor of whatever state or trait is being measured (Snodgrass et al., 1998). For example, if a patient is distressed from being laid off recently, and the clinician does not pick up on the level of distress, researchers must determine whether this occurred because the clinician is not a good detector of distress or because the patient was not emitting cues associated with their distressed state. Disambiguation typically involves recording the clinical encounter, so that the patient's overall expressivity can be assessed by independent observers (Hall et al., 2006; Noller, 1980). Thus, the intrusion into the clinical visit is compounded by the presence of recording devices and furthermore the assessment of the patient's expressivity is itself a laborious task for the researcher. These are impediments to the implementation of this method. Finally, there is the question of how being recorded might change behavior. Most patients who take part in medical research using audiotapes or videotapes report forgetting that the audiotape or videotape was in the room after a few minutes, and their behavior typically does not change as a function of a recording device (Penner et al., 2007). One study examined clinicians' behavior during surgical consultations when they were aware and not aware of being videotaped. The study findings provided no evidence that awareness of being video recorded altered clinicians' behavior (Pringle & Stewart-Evans, 1990). However, it is not impossible that sometimes the audio- or videotape could cause some patients or clinicians to feel uncomfortable and could alter their behavior; how this would affect the clinician's accuracy is unknown.

Standardized tests. The most common method to assess interpersonal accuracy is to administer tests of judging the meaning of cues. A number of established tests are available to researchers, which mainly measure accuracy of interpreting affect cues [e.g., Diagnostic Analysis of Nonverbal Accuracy (DANVA); Nowicki & Duke, 1994; Emotion Recognition Index (ERI); Scherer & Scherer, 2011; Geneva Emotion Recognition Test (GERT); Schlegel, Grandjean, & Scherer, 2014; Interpersonal Perception Task (IPT); Costanzo & Archer, 1989; Japanese and Caucasian Facial Expressions of Emotion (JACFEE); Matsumoto et al., 2000; Multimodal Emotion Recognition Test (MERT); Bänziger, Grandjean, & Scherer, 2009; Pictures of Facial Affect (POFA); Ekman & Friesen, 1976; Profile of Nonverbal Sensitivity (PONS); Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979; Reading the Mind in the Eyes; Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001]. The test stimuli can consist of only nonverbal cues such as facial expressions or combined nonverbal and linguistic cues. Typically, the answer options are in multiple-choice format. For example, one test of judging emotions asks test takers to decide between happy, sad, angry, or fearful emotion after viewing facial expressions (DANVA-2 Adult Faces; Nowicki & Duke, 1994). These types of tests vary in how long they are (from 3 min to 45 min) and also in presentation modality (photographs versus videos), in what channels are presented (voice, face, eyes, postures, alone or in combination), and in their content domain.

Only a few tests have been developed specifically for clinicians. One such test is the Patient Emotion Cue Test (PECT; Blanch-Hartigan, 2011b), which presents video clips of an actress portraying real patient statements that convey confusion, anger, happiness, sadness, and anxiety, plus neutral statements and nonverbal cues. Each video clip varies in affective intensity in both the nonverbal and the verbal modalities. The PECT is a reliable and valid tool for assessing emotion detection.

Another videotape test using clinically relevant stimuli is the Test of Accurate Perception of Patients' Affect (TAPPA; Hall et al., 2014a).

The TAPPA tests the ability to infer thoughts and feelings of patients and is the first known test using real doctor–patient stimuli to assess interpersonal accuracy. The TAPPA was developed using videotaped medical interactions of patients talking to their doctors in their routine medical visits. After the medical interaction, patients watched the videotaped medical interaction and stopped the videotape every time they remembered having a thought or feeling. They recorded what their thought or feeling was and proceeded until they watched the entire duration of their medical encounter. This method was adapted from Ickes (Ickes, Stinson, Bissonnette, & Garcia, 1990) and is often called the empathic accuracy paradigm (see Chapter 3). The information was used to extract short audiovisual clips for which the correct answer was the patient's report of the thought or feeling associated with that clip. Alternative plausible but incorrect options were also developed for each clip. The TAPPA contains 48 audiovisual clips, each of about 30-s duration, with a multiple-choice–answer format for each clip. The TAPPA shows good psychometric properties and convergent validity with other tests of emotion recognition, especially with the PECT, which was also developed for clinician test takers.

Behavioral coding of a medical interaction. Another measure of interpersonal skill in the clinical context involves videotaping or audio recording doctor–patient interactions and asking reliable observers for their impressions of the interaction using a coding scheme. This method does not rely on asking clinicians what cues they noticed but rather interpreting from clinicians' responses or clinicians' or patients' behavior that they were responding to identified patient cues. Although not a precise measurement of actual clinician interpersonal accuracy because of the inferences that are made, this method does not require much interaction or intrusion for the patients or clinicians, except for their permission to audiotape or videotape the medical interaction. However, coding these behaviors requires a trained and reliable coder, which can be costly in terms of time and effort.

Several valid and reliable behavioral coding schemes can be used to measure interpersonal skill in this way. For example, the Roter Interaction Analysis System (RIAS; Roter, 1995) provides a tool for measuring the dynamic interaction of patients and clinicians as they exchange resources (e.g., about health, disease, symptoms, and lifestyle choices) and other relevant information during a medical interaction. The RIAS can be used to measure interpersonal perception by coding categories of emotional talk by the clinician such as how well clinicians address concerns, exhibit reassurance, empathy, and partnership. The Four Habits Coding Scheme (4HCS; Krupat, Frankel, Stein, & Irish,

2006) is another coding instrument that allows for the description and evaluation of clinician behavior. In particular, the 4HCS codes four habits or categories of clinician behavior – invest in beginning, elicits patient's perspective, demonstrate empathy, and invest in end. Most relevant to accuracy is demonstrating empathy, which includes the use of encouraging emotional expressions, accepting feelings, identifying feelings, and showing good nonverbal behavior, all of which can contribute to good interpersonal skills.

These behavioral coding systems provide great utility when interested in general categories of empathy and interpersonal skill, but they do not measure precisely clinicians' ability to perceive their patients' nonverbal behavior or thoughts and feelings. However, such behaviors, along with others that might be coded in the clinician and patient, may be associated with interpersonal perception skill. In a sample of medical students who took the TAPPA, higher scores on that test correlated with the students' higher engagement during a standardized patient clinical interaction (Hall et al., 2014b). Although engagement is a valued patient-centered characteristic for clinicians to display toward their patients, it is not interpersonal accuracy as we define it – accurately assessing patients.

It is important to distinguish interpersonal accuracy from other interpersonal skills or empathy. In research involving doctor–patient communication, interpersonal skills and in particular clinician empathy have become a large topic of study that is undoubtedly important to study but encompasses many types of interpersonal skills. Thus it is important for researchers to distinguish exactly what skill they are interested in and what method is best to measure that skill as using a coding scheme of exhibiting reassurance (for example) may result in different outcomes and may not even be correlated with accurately recalling patient behavior or accurately inferring patients' emotions.

Quantifying interpersonal accuracy. In addition to various ways to measure interpersonal accuracy, there are various ways to calculate accuracy, each of which has its strengths and limitations. We will discuss three forms of quantifying accuracy: (1) correlational accuracy, (2) mean-level bias, and (3) signal detection accuracy. These three methods do not necessarily produce the same results (Hall, Stein, Roter, & Rieser, 1999), and failure to distinguish among these can produce confusing conclusions (Blanch-Hartigan, 2011a).

Correlational accuracy. When researchers are interested in, for example, whether clinicians can discriminate higher intensities of patient pain from lower intensities of patient pain or patients who are experiencing more negative emotions from patients who are experiencing less negative emotions, they often employ the correlational accuracy approach. Accuracy correlations are calculated by correlating the patients' self-reported state or trait (e.g., patient's self-reported pain intensity) with the inferred state or trait rating made by the clinician (e.g., inferred pain intensity of the patient), for a sample of patients. Each clinician then has an accuracy correlation. Higher accuracy correlations signify more covariation between the patients' actual state/trait and the clinician's inferred state/trait. In the judgment of pain example, a higher accuracy correlation means the clinician can discriminate higher intensity levels of pain from lower intensity levels of pain. The correlational approach results in a measure of generality, that is, having an ability to detect where a patient stands in relation to other patients on a given characteristic.

Mean-level bias. Another form of calculating accuracy is mean-level bias also known as direction bias (West & Kenny, 2011). Mean-level bias is calculated when a researcher is interested in whether clinicians have a general tendency to over- or underestimate patients' traits or states, for example, pain or satisfaction. Measuring the underestimation or overestimation of a state or trait requires difference scores that are derived from subtracting clinicians' inferences about the patients' states or traits from patients' actual self-reported states or traits. The difference scores signify the concordance or discordance between the clinician and patient on the dimension of interest. Mean-level bias results in a measure of specificity that describes in which direction and by how much clinicians have a tendency to estimate their patients on a given characteristic.

These two methods of measurement have some obvious and important implications for patients. The distinction between specificity and generality is rarely discussed but is important to note in order to understand what each assessment is actually measuring. Importantly, these two kinds of accuracy are statistically independent of each other. According to the correlational accuracy method or generality, clinicians who score higher on correlations with their patients tend to know which patients are (for example) in more pain than others. However, if a physician has high correlational accuracy, they could have high or low specificity. If a clinician has low specificity, this would mean that they always have a tendency to systematically over- or underestimate their patients on the dimension of interest. For example, a clinician with low specificity in judging pain would always underestimate how much pain their patients were experiencing and therefore never estimate the exact intensity of pain for any one patient. This clinician might give relatively correct doses of pain medication but absolutely the wrong amount for everyone (i.e., no one gets as much as they need). Although never studied, from this rationale, patients who have clinicians who tend to underestimate cues in general and/or have poor correlational accuracy may be at higher risk for poorer health outcomes. In the example above, they would give higher doses of pain medication to the patients who needed more medicine, but all of their patients would still not be receiving enough pain medication.

Signal detection accuracy. The last approach that will be discussed is the signal detection accuracy method, which measures both of the previous kinds of accuracy at the same time. Signal detection theory (SDT) is based on the premise that there is some uncertainty in all decision making. SDT was first developed for use in sensory experiments but has been adapted for use in other disciplines including social psychology and clinical settings. Imagine a clinician is interacting with a patient looking for symptoms of depression. The patient is depressed (signal present) or is not depressed (signal absent), and the clinician diagnoses depression (they respond “yes”) or does not diagnose depression (they respond “no”). There are four possible outcomes: hit (depression present, clinician says yes), miss (depression present, clinician says no), false alarm (depression absent, clinician says yes), or correction rejection (depression absent, clinician says no). Hits in combination with correct rejections lead to higher accuracy, or in signal detection language, higher sensitivity. Sensitivity reflects the ability to discriminate the state or trait from a bias for selecting that state or trait. In the previous example, sensitivity reflects the ability to discriminate clinically depressed patients from nonclinically depressed patients independent of the clinician’s bias for selecting depressed. Clinicians can also have high and low criteria, also known as response bias. A low criterion would mean that the clinician would respond “yes” to depression for nearly every patient, leading to high-hit rates but also high-false alarm rates. A high criterion would mean that the clinician would respond “no” to depression for every patient, leading to high-correct rejection rates but also high miss rates.

SDT is a useful statistical tool for separating sensitivity from bias; however, there are several limitations using SDT in this type of research. For example, SDT requires the researcher to make several assumptions about the underlying shape and distribution of the data that can have important implications for the interpretation of results (e.g., that the noise or uncertainty distribution is normal). Testing these assumptions is crucial when using SDT in order to make strong arguments about sensitivity and bias. Reliability is also hurt when there is a small number of trials, an uneven number of trial types, or when errors in clinicians’ responses are rare. See Wickens (2001) for more details on the strengths and weaknesses of SDT.

What types of interpersonal accuracy have been researched in the clinical context?

The most commonly researched type of interpersonal accuracy in clinical and nonclinical settings is detecting emotions (see Chapter 2). For example, is the patient upset? Nervous about an upcoming procedure? Distressed from a recent diagnosis? Coping well with life stressors? There are many other dimensions of patients that clinicians can and should be accurate about. For example, diagnosing dementia or cognitive impairment, psychiatric disturbances, physiological or physical states such as how much pain a patient is experiencing, diagnosing alcohol disorders, or other patient characteristics such as their personality, values, culture, expectations, whether they're being truthful or trying to deceive their clinician, and whether they like their clinician or are satisfied with care.

Pain. An important interpersonal accuracy skill for clinicians is judging the level of pain that patients are experiencing. Inaccurate assessments of pain have serious implications for patients. Underestimation of pain can lead to improper management of pain and prolonged suffering, while overestimation of pain can lead to overtreatment with potentially addictive pain medication. For those who cannot self-report their pain, such as infants and dementia patients, observer judgments of pain are the primary assessment of pain and they guide treatment and medication.

There is a large literature on pain assessment accuracy, or the ability to accurately judge pain, that has highlighted some important individual and group differences in both correlational accuracy and mean-level bias. Most of the literature on pain assessment accuracy has focused on nurses' assessments of patients' pain. Correlations between nurse and patient ratings tend to be significantly positive but in the low to moderate range (average $r = .38$) (Teske, Daut, & Cleeland, 1983; Van der Does, 1989). That is, nurses can moderately discriminate one patient's pain intensity from another patient's pain intensity.

In terms of mean-level bias in judging pain, most physicians and nurses underestimate the intensity of the pain experience by an average of 2.0–2.5 cm on the Visual Analogue Scale, a scale in which patients are asked to place an "X" on a line 100 cm long anchored at "no pain at all" to "the most intense pain imaginable." In one study, patients experiencing acute pain and their family practice physicians were asked to estimate their pain at the time of the medical appointment. Physicians significantly underestimated their patients' pain ($p < .001$) (Sutherland et al., 1988). Although nurses show this general pattern of underestimating patient pain, it is not usually as drastic as found in physicians (Heikkinen, Salantera, Kettu, & Taittonen, 2005). Some studies have also provided evidence that as providers gain healthcare experience, they become less accurate at judging pain and underestimate pain even more than providers with minimal experience (Choiniere, Melzack, Girard, Rondeau, & Paquin, 1990; Mason, 1981).

Researchers have proposed many possibilities for why healthcare professionals have an inability to accurately judge their patients' pain. One possibility is that clinicians are under time constraints during medical appointments, which leads them to miss patient cues that are pertinent to pain assessment accuracy (Schafheutle, Cantrill, & Noyce, 2001). Another possibility is that clinicians develop a defense mechanism that protects them from continued exposure to patients' suffering or negative emotions and causes them to minimize the pain that they observe (Choiniere et al., 1990; Goubert, Craig, & Buysse, 2009). Short trainings focused on the nonverbal facial expressions of pain have shown some promise in improving pain assessment accuracy in nonclinical settings (Hill & Craig, 2004). In Chapter 12, studies of training healthcare providers on their interpersonal accuracy abilities are reviewed. Although research on clinical outcomes associated with providers' pain assessment accuracy has not been conducted, two studies examined the role of caregivers' pain assessment accuracy on patients' mental and physical health outcomes. Miaskowski, Zimmer, Barrett, Dibble, and Wallhagen (1997) found that when caregivers had lower levels of pain assessment accuracy, patients had more mood disturbance and poorer quality of life. Riemsma, Taal, and Rasker (2000) similarly found that both over- and underestimation of pain by partners was related to patients' poorer mental health status. The expressivity confound that was discussed earlier is relevant to these studies due to the dyadic nature of these studies. For example, we do not know whether these caregivers had lower pain assessment accuracy or whether patients with poorer mental health status conveyed their pain less accurately to their caregiver.

Depression. Diagnosing or at least identifying symptoms associated with mental health conditions is an important attribute of any good clinician, not just psychiatrists and clinical psychologists. Most published data on correctly diagnosing patients' mental disturbances (e.g., depression) use a specificity and sensitivity approach. A perfectly accurate clinician would have a sensitivity and specificity score of 100 percent, correctly identifying all depressed patients and nondepressed patients.

Diagnosing depression has been studied extensively through meta-analysis. Mitchell and Kakkadasam (2011) conducted a meta-analysis to examine the identification of depression by practice and community nurses. On average, nurses correctly identified 26.3% of people with depression (sensitivity) and correctly identified 94.8% of nondepressed patients (specificity). In a different review, primary care physicians showed a slightly higher sensitivity (47.3%) compared to nurses but a slightly lower specificity (81.3%) (Mitchell, Vaze, & Rao, 2009).

In addition, research has shown that nurses have a tendency to underestimate the level of depressive symptoms in patients who are severely depressed and use invalid cues of depression such as crying and depressed mood to detect depression (McDonald et al., 1999). There may also be differences in the patient that make it easier or harder to detect illnesses like depression. For example, in another meta-analysis, primary care physicians were less successful in identifying depression in older patients than in younger patients (Mitchell, Rao, & Vaze, 2010). There are several possible reasons for the low identification rates of depression in older adults. First, older patients may be more likely to believe that depressive symptoms are a normal part of aging and may not display the symptoms that clinicians use for diagnosing depression in younger patients (Lyness et al., 1995). Similarly, physicians of older adults might perceive depressive symptoms but believe that this is a normal part of aging. Another possibility is that late-life depression presents with vague symptoms or depression without low mood (Lyness et al., 1996). Finally, because older adults suffer from many physical symptoms, clinicians may focus more on the physical complaints, missing or ignoring the somatic complaints associated with depression (O'Connor, Rosewarne, & Bruce, 2001). Clinicians working with this population should be targeted for interventions to improve the recognition of the valid cues of depression for older adults especially because the prevalence of depression increases with age (Thielke, Diehr, & Unutzer, 2010).

Along with patient age affecting recognition of depression, patient race and ethnic disparities exist in the recognition and diagnosis of depression. For example, non-Hispanic white patients were significantly more likely to be diagnosed with depression following a stroke compared to other racial/ethnic groups, even after adjusting for sociodemographic and clinical risk factors (Jia et al., 2010). It is unclear whether the differences in diagnosis are due to racial or ethnic differences in symptom expression by patients or in recognition of symptoms by providers.

Suicide. Karver, Tarquini, and Caporino (2010) examined whether helpline counselors' judgments of risk accurately predicted suicide risk behavior in a 6-month period following intake. Counselors were quite accurate at judging clients' risk for future suicide risk behavior. Counselors' sensitivity at correctly identifying suicide risk was 66.6% while their specificity at identifying youth not at risk for suicide behavior was 93%. It was much more difficult for counselors to accurately identify low risk compared to no risk or moderate to high risk cases.

Distress. Many argue that the focus of detection should not be on the psychiatric disorder, such as depression, but rather on identifying people who are suffering from distress because it is a marker for many psychiatric disorders (Mulder, 2008; Saraceno, Laviola, Sternai, Terzian, & Tagnoni, 1994). Patients can complete single item Visual Analogue Scales of Distress (NCCN, 2007), or longer self-report measures of distress such as the General Health Questionnaire (GHQ; Goldberg & Hillier, 1979), the Hopkins Symptom Checklist (Derogatis, 1977), and the Kessler Psychological Distress Scale K10 (Kessler et al., 2002).

Mitchell, Rao, and Vaze (2011) meta-analyzed the ability of general practitioners (GPs) to identify distress in studies of sensitivity and specificity. GPs correctly identified 48.4% of people with distress (sensitivity) and correctly identified non-distressed people 79.4% of the time (specificity). The sensitivity in detecting distress was slightly higher among these GPs than was the sensitivity in detecting depression in their meta-analysis of nurses and primary care physicians. The detection of mild disorders poses a challenge to clinicians because they do not differ greatly from healthy patients (Olfson, Gilbert, Weissman, Blacklow, & Broadhead, 1995; Perez Stable, Miranda, & Munoz, 1990).

Dementia and cognitive impairment. Similar to depression and distress, accurate detection of dementia is challenging for clinicians. Not only is there a low prevalence of dementia, 16% in those aged 70 and older (Olafsdottir, Skoog, & Marcusson, 2000), but many people suffering from dementia or cognitive impairment do not seek help for explicit memory complaints (Iliffe & Pealing, 2010; Mitchell, 2008). Mitchell, Meader, and Pentzek (2011) conducted a meta-analysis on the ability of GPs to recognize cognitive impairment (mild to severe dementia). GPs had generally poor sensitivity for recognizing impairment (63%) but showed high specificity (93%). Mild cognitive impairment (45% sensitivity) was more difficult to recognize compared to dementia (75% sensitivity). The authors concluded that about one in four cases of dementia remains undetected even in those seeking help, while even higher rates are undetected for those with mild cognitive impairment (one in two).

Alcohol disorders. Clinicians also have difficulty identifying and helping patients with alcohol problems. Diagnosis is usually made by clinical judgment without the use of scales, blood tests, or reference to diagnostic criteria (Aalto & Seppa, 2005; Berner et al., 2006). Preliminary research suggests that only about one-third of individuals with alcohol problems are detected by their GPs (Rydon, Redman, Sanson-Fisher, & Reid, 1992). In a meta-analysis by Mitchell, Meader, Bird, and Rizzo (2012), health professionals' sensitivity in correctly identifying alcohol problems was 50% (mental health professionals' sensitivity was 54.7%, GPs' sensitivity was 41.7%).

Adherence. Adherence is defined as the extent to which a patient's behavior coincides with medical or health instructions (Haynes, McDonald, & Garg, 2002). Accurate detection of patient adherence is an important skill for clinicians because it allows for intervention at an early stage by addressing the issues that may lead to poor adherence (e.g., patient attitudes, patients' frustration, misunderstanding of their illness, and taking medication). Yet the literature provides much evidence that physicians are inaccurate judges of their patients' adherence to treatment (Wagner et al., 2001; Zeller, Taegtmeyer, Martina, Battegary, & Tschudi, 2008). Phillips, Leventhal, and Leventhal (2011) examined the factors related to the accuracy of physicians' adherence predictions. Immediately after routine medical encounters, physicians completed questionnaires that assessed their perceptions of patients' prospective adherence in the next month and factors that impacted their predictions. Patients were asked immediately after their visit about their agreement with the physicians about illness and treatment and their perceptions of their physician's discussion about their illness and treatment. Physicians' prospective prediction of patient adherence was significantly and positively correlated with patients' self-reported adherence but weakly.

Patient satisfaction and quality of care. An important predictor of health-seeking behavior and health in general is patients' satisfaction (Pascoe, 1983; Sitzia & Wood, 1997). Merkel (1984) examined whether advanced family practice residents could accurately predict their patients' level of satisfaction immediately after a medical interaction with that resident. Patient satisfaction ratings were only related to how much residents liked their patients. In general, residents could not accurately distinguish more satisfied patients from less-satisfied patients, though there were individual differences in this skill as two residents did accurately predict their patients' satisfaction. Clinicians' accuracy in judging patient satisfaction could be low due to a restriction of range. Patients tend to rate satisfaction with their physicians on the high end of the scale, creating a ceiling effect in patients' ratings of satisfaction, which makes it harder for providers to accurately detect satisfaction.

Similarly, Hall et al. (1999) examined primary care physicians' awareness of their patients' rated emotions, satisfaction, and opinion of the quality of their communication. Both physicians and patients filled out a questionnaire following a routine medical appointment. Accuracy correlations between patients' and physicians' views of patients' emotions and satisfaction were weak to moderate. Examination of mean ratings (bias) showed that physicians tended to estimate their patients' responses as more negative than they actually were. Physicians were slightly better at judging emotions than satisfaction. The authors reasoned that patients may be more likely to show their emotions than their evaluations of their physician in outward cues, making it easier to infer emotions.

Another study examined physicians' accuracy at judging patients' liking for their physician after a routine medical appointment. Both physicians and patients were asked to rate how much they liked each other and how much they perceived the other to like them. Accuracy of perceived liking was significant and positive and equal in magnitude for both patients ("doctor likes me" and "I like patient") and physicians ("patient likes me" and "I like doctor"), though the correlations were rather weak (Hall, Horgan, Stein, & Roter, 2002).

How do clinicians compare to nonclinicians in interpersonal accuracy?

The foregoing evidence suggests room for improvement in clinicians' accuracy and that, as a group, people who choose to enter clinical professions may not be especially gifted in interpersonal perception. The few studies that allow such comparisons support this hypothesis. Rosenthal et al. (1979) found that clinical psychologists in training were no better in interpreting nonverbal cues of affect than were unselected norm groups, and the same was found for medical students when compared with published norms for the same tests (Evans, Coman, & Stanley, 1988; Hall, Roter, Blanch, & Frankel, 2009).

What makes a clinician interpersonally accurate?

The literature on correlates of clinicians' interpersonal perception accuracy is relatively small and remains correlational in nature, meaning that causal paths to and from such accuracy cannot be specified with confidence. However, the studies that exist have produced tantalizing results.

Many studies outside of the clinical realm find that women perform better than men on tests of interpersonal accuracy (see Chapter 15). Similarly, women in medicine also outperform their male counterparts (Hall, Roter et al., 2009; Hall et al., 2014a).

The TAPPA test of interpreting patients' thoughts and feelings was used to examine several other variables that might be causal antecedents of accuracy. Accuracy on the TAPPA was significantly positively correlated with nursing students' clinical course experience, suggesting that more experience with actual patients may build greater accuracy in perceiving them. In addition, in medical or nursing trainees, tested interpersonal accuracy was significantly correlated with having more favorable attitudes to psychosocial discussion (Hall, Roter et al., 2009, Hall et al., 2014a).

What patient outcomes are associated with having an interpersonally accurate clinician?

Interpersonal accuracy among clinicians has been correlated with several positive patient outcomes. This includes patient satisfaction (DiMatteo, Taranta, Friedman, & Prince, 1980), a better rate of appointment keeping in patients (DiMatteo, Hays, & Prince, 1986), higher ratings of clinical skill as a psychotherapist made by their supervisors (Rosenthal et al., 1979), more vigilance for anxiety and depression in medical patients and marginally more accuracy in detecting these states (Robbins, Kirmayer, Cathébras, Yaffe, & Dworkind, 1994), better performance by occupational therapy students on their clinical fieldwork examinations (Tickle-Degnen, 1998), standardized patients' warmth and engagement when interacting with medical students and higher observer ratings of liking and satisfaction (Hall, Roter et al., 2009), and higher ratings of medical students' interpersonal skill by standardized patients after a simulated clinical visit (Hall et al., 2014b).

Conclusion

Interpersonal accuracy is important for the clinician–patient relationship. Not only does it allow clinicians to better diagnose and treat their patients' symptoms, but it also appears to build better relationships with patients. The studies reviewed here, plus many from the nonclinical literature showing that interpersonal accuracy is correlated with prosociality, open-mindedness, leadership, persuasion, and negotiation success, yield a strong case for the importance of this skill in the clinical context (Hall et al., 2009; see also Chapters 13 and 17).

References

Aalto, M., & Seppa, K. (2005). Use of laboratory markers and the audit questionnaire by primary care physicians to detect alcohol abuse by patients. *Alcohol and Alcoholism*, 40, 520–523.

Adams, A., Buckingham, C. D., Lindenmeyer, A., McKinlay, J. B., Link, C., Marceau, L., & Arber, S. (2008). The influence of patient and doctor gender on diagnosing coronary heart disease. *Sociology of Health and Illness*, 30, 1–18.

Bänziger, T., Grandjean, D., & Scherer, K. R. (2009). Emotion recognition from expressions in face, voice, and body: The Multimodal Emotion Recognition Test (MERT). *Emotion*, 9, 691–704.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry*, 42, 241–251.

Berner, M. M., Zeidler, C., Kriston, L., Mundle, G., Lorenz, G., & Harter, M. (2006). Diagnostic and treatment approaches to alcohol-related disorders. *Fortschritte der Neurologie*, 74, 157–164.

Blanch-Hartigan, D. (2011a). Medical students' self-assessment of performance: Results from three meta-analyses. *Patient Education and Counseling*, 84, 3–9.

Blanch-Hartigan, D. (2011b). Measuring providers' verbal and nonverbal emotion recognition ability: Reliability and validity of the Patient Emotion Cue Test (PECT). *Patient Education and Counseling*, 82, 370–376.

Blanch-Hartigan, D. (2012). An effective training to increase accurate recognition of patient emotion cues. *Patient education and counseling*, 89, 274–280.

Blanch-Hartigan, D. (2013). Patient satisfaction with physician errors in detecting and identifying patient emotion cues. *Patient Education and Counseling*, 93, 56–62.

Choiniere, M., Melzack, R., Girard, N., Rondeau, J., & Paquin, M. J. (1990). Comparisons between patients' and nurses' assessment of pain and medication efficacy in severe burn injuries. *Pain*, 40, 143–152.

Costanzo, M., & Archer, D. (1989). Interpreting the expressive behavior of others: The Interpersonal Perception Task. *Journal of Nonverbal Behavior*, 13, 225–245.

Davis, D. A., Mazmanian, P. E., Fordis, M., Van Harrison, R., Thorpe, K. E., & Perrier, L. (2006). Accuracy of physician self-assessment compared with observed measures of competence: A systematic review. *The Journal of the American Medical Association*, 296, 1094–1102.

Derogatis, L. R. (1977). *SCL-90-R, Administration, Scoring, and Procedures Manual 1*. Baltimore: Clinical Psychometric Research.

DiMatteo, M. R., Hays, R. D., & Prince, L. M. (1986). Relationship of physicians' nonverbal communication skills to patient satisfaction, appointment noncompliance, and physician workload. *Health Psychology*, 5, 581–595.

DiMatteo, M. R., Taranta, A., Friedman, H. S., & Prince, L. M. (1980). Predicting patient satisfaction from physicians' nonverbal communication skills. *Medical Care*, 18, 376–387.

Ekman, P., & Friesen, W. V. (1976). *Pictures of facial affect*. Palo Alto, CA: Consulting Psychologists Press.

Evans, B. J., Coman, G. J., & Stanley, R. O. (1988). Scores on the Profile of Nonverbal Sensitivity: A sample of Australian medical students. *Psychological Reports*, 62, 903–906.

Goldberg, D. P., & Hillier, V. F. (1979). A scaled version of the General Health Questionnaire. *Psychological Medicine*, 9, 139–145.

Goubert, L., Craig, K. D., & Buysse, A. (2009). Perceiving others in pain: Experimental and clinical evidence on the role of empathy. In W. Ickes & J. Decety (Eds.), *The social neuroscience of empathy* (pp. 153–165). Cambridge, MA: The MIT Press.

Gulbrandsen, P., Benth, J. S., Dahl, F. A., Fossli Jensen, B., Finset, A., & Hall, J. A. (2012). Specialist physicians' sensitivity to patient affect and satisfaction. *Medical Care*, 50, 290–293.

Hall, J. A., Andrzejewski, S. A., & Yopchick, J. E. (2009). Psychosocial correlates of interpersonal sensitivity: A meta-analysis. *Journal of Nonverbal Behavior*, 33, 149–180.

Hall, J. A., Horgan, T. G., Stein, T. S., & Roter, D. L. (2002). Liking in the physician-patient relationship. *Patient Education and Counseling*, 48, 69–77.

Hall, J. A., Rosip, J. C., Smith LeBeau, L., Horgan, T. G., & Carter, J. D. (2006). Attributing the sources of accuracy in unequal-power dyadic communication: Who is better and why? *Journal of Experimental Social Psychology*, 42, 18–27.

Hall, J. A., Roter, D. L., Blanch, D. C., & Frankel, R. M. (2009). Nonverbal sensitivity in medical students: Implications for clinical interactions. *Journal of General Internal Medicine*, 24, 1217–1222.

Hall, J. A., Ship, A. N., Ruben, M. A., Curtin, E. M., Clever, S., Roter, D. L., Smith, C., & Pounds, K. (2014a). The Test of Accurate Perception of Patients' Affect (TAPPA): An ecologically valid tool for assessing interpersonal perception accuracy in clinicians. *Patient Education and Counseling*, 94, 218–223.

Hall, J. A., Ship, A. N., Ruben, M. A., Curtin, E. M., Roter, D. L., Clever, S., Smith, C., & Pounds, K. (2014b). Clinically relevant correlates of accurate perception of patients' thoughts and feelings. *Health Communication*, 20, 1–7.

Hall, J. A., Stein, T. S., Roter, D. L., & Rieser, N. (1999). Inaccuracies in physicians' perceptions of their patients. *Medical Care*, 37, 1164–1168.

Haynes, R. B., McDonald, H. P., & Garg, A. X. (2002). Helping patients follow prescribed treatment. *Journal of American Medical Association*, 288, 2880–2883.

Heikkinen, K., Salantera, S., Kettu, M., & Taittonen, M. (2005). Prostatectomy patients' postoperative pain assessment in the recovery room. *Journal of Advanced Nursing*, 52, 592–600.

Hill, M. L., & Craig, K. D. (2004). Detecting deception in facial expressions of pain: Accuracy and training. *The Clinical Journal of Pain*, 20, 415–422.

Ickes, W., Stinson, L., Bissonnette, V., & Garcia, S. (1990). Naturalistic social cognition: empathic accuracy in mixed-sex dyads. *Journal of Personality and Social Psychology*, 59, 730–742.

Iliffe, S, & Pealing, L. (2010). Subjective memory problems. *British Medical Journal*, 340, c1425.

Institute of Medicine Committee on Quality Health Care in America. (2001). *Crossing the quality chasm: a new health system for the 21st century*. Washington, DC: National Academy Press.

Jia, H., Chumbler, N. R., Wang, X., Chuang, H. C., Damush, T. M., Cameon, R., & Williams, L. S. (2010). Racial and ethnic disparities in post-stroke depression detection. *International Journal of Geriatric Psychiatry*, 25, 298–304.

Karver, M. S., Tarquini, S. J., & Caporino, N. E. (2010). The judgment of future suicide-related behavior: Helpline counselors' accuracy and agreement. *Crisis*, 31, 272–280.

Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S. L., Walters, E. E., & Zaslavsky, A. M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine*, 32, 959–976.

Krupat, E., Frankel, R., Stein, T., & Irish, J. (2006). The Four Habits Coding Scheme: Validation of an instrument to assess clinicians' communication behavior. *Patient Education and Counseling*, 62, 38–45.

Lyness, J. M., Bruce, M. L., Koenig, H. G., Parmelee, P A., Schulz, R., Lawton, M. P, & Reynolds, C. F. (1996). Depression and medical illness in late life: Report of a symposium. *Journal of the American Geriatrics Society*, 44, 198–203.

Lyness, J. M., Cox, C., Curry, J., Conwell, Y., King, D. A., & Caine, E. D. (1995). Older age and the underreporting of depressive symptoms. *Journal of the American Geriatrics Society*, 43, 216–221.

Mason, D. J. (1981). An investigation of the influences of selected factors on nurses' inferences of patient suffering. *International Journal of Nursing Studies*, 18, 251–259.

Matsumoto, D., LeRoux, J., Wilson-Cohn, C., Raroque, J., Kooken, K., Ekman, P., Yrizarry, N., Loewinger, S., Uchida, H., Yee, A., Amo, L., & Goh, A. (2000). A new test to measure emotion recognition ability: Matsumoto and Ekman's Japanese and Caucasian Brief Affect Recognition Test (JACBART). *Journal of Nonverbal Behavior*, 24, 179–209.

McDonald, M. V., Passik, S. D., Dugan, W., Rosenfeld, B., Theobald, D. E., & Edgerton, S. (1999). Nurses' recognition of depression in their patients with cancer. *Oncology Nursing Forum*, 26, 593–599.

McWhinney, I. (1989). The need for a transformed clinical method. In M. Stewart & D. Roter (Eds.), *Communicating with medical patients* (pp. 25–42). London: Sage.

Merkel, W. (1984). Physician perception of patient satisfaction: Do doctors know which patients are satisfied? *Medical Care*, 22, 453–459.

Miaskowski, C., Zimmer, E. F., Barrett, S. L., Dibble, M., & Wallhagen, M. (1997). Differences in patients' and family caregivers' perceptions of the pain experience influence patient and caregiver outcomes. *Pain*, 72, 217–226.

Mitchell, A. J. (2008). The clinical significance of subjective memory complaints in the diagnosis of mild cognitive impairment and dementia: A meta-analysis. *International Journal of Geriatric Psychiatry*, 23, 1191–1202.

Mitchell, A. J., & Kakkadasam, V. (2011). Ability of nurses to identify depression in primary care, secondary care, and nursing homes – A meta-analysis of routine clinical accuracy. *International Journal of Nursing Studies*, 48, 359–368.

Mitchell, A. J., Meader, N., Bird, V., & Rizzo, M. (2012). Clinical recognition and recording of alcohol disorders by clinicians in primary and secondary care: A meta-analysis. *The British Journal of Psychiatry*, 201, 93–100.

Mitchell, A. J., Meader, N., & Pentzek, M. (2011). Clinical recognition of dementia and cognitive impairment in primary care: A meta-analysis of physician accuracy. *Acta Psychiatrica Scandinavica*, 124, 165–183.

Mitchell, A. J., Rao, S., & Vaze, A. (2010). Do primary care physicians have particular difficulty identifying late-life depression? A meta-analysis stratified by age. *Psychotherapy and Psychosomatics*, 79, 285–294.

Mitchell, A. J., Rao, S., & Vaze, A. (2011). Can general practitioners identify people with distress and mild depression? A meta-analysis of clinical accuracy. *Journal of Affective Disorders*, 130, 26–36.

Mitchell, A. J., Vaze, A., & Rao, S. (2009). Clinical diagnosis of depression in primary care: A meta-analysis. *Lancet*, 374, 609–619.

Mulder, R. T. (2008). An epidemic of depression or medicalization of distress? *Perspectives in Biology and Medicine*, 51, 238–250.

NCCN Clinical Practice Guidelines in Oncology Distress Management V.1.2007. www.nccn.org/professionals/physician_gls/PDF/distress.pdf.

Noller, P. (1980). Misunderstandings in marital communication: A study of couples' nonverbal communication. *Journal of Personality and Social Psychology*, 39, 1135–1148.

Nowicki, S. & Duke, M. P. (1994). Individual differences in the nonverbal communication of affect: the Diagnostic Analysis of Nonverbal Accuracy scale. *Journal of Nonverbal Behavior*, 18, 9–34.

O'Connor, D. W., Rosewarne, R., & Bruce, A. (2001). Depression in primary care 2: General practitioners' recognition of major depression in elderly patients. *International Psychogeriatrics*, 13, 367–374.

Olafsdottir, M., Skoog, I., & Marcusson, J. (2000). Detection of dementia in primary care: The Linkoping Study. *Dementia and Geriatric Cognitive Disorders*, 11, 223–229.

Olfson, M., Gilbert, T., Weissman, M., Blacklow, R. S., & Broadhead, W. E. (1995). Recognition of emotional distress in physically healthy primary care patients who perceive poor physical health. *General Hospital Psychiatry*, 17, 173–180.

Pascoe, G. C. (1983). Patient satisfaction in primary health care: A literature review and analysis. *Evaluation and Program Planning*, 6, 185–210.

Penner, L. A., Orom, H., Albrecht, T. L., Franks, M. M., Foster, T. S., & Ruckdeschel, J. C. (2007). Camera-related behaviors during video recorded medical interactions. *Journal of Nonverbal Behavior*, 31, 99–117.

Perez Stable, E., Miranda, J., & Munoz, R. F. (1990). Depression in medical outpatients: Underrecognition and misdiagnosis. *Archives of Internal Medicine*, 150, 1083–1088.

Phillips, L. A., Leventhal, E. A., & Leventhal, H. (2011). Factors associated with the accuracy of physicians' predictions of patient adherence. *Patient Education and Counseling*, 85, 461–467.

Pringle, M., & Stewart-Evans, C. (1990). Does awareness of being video recorded affect doctors' consultation behaviour? *British Journal of General Practice*, 40, 455–458.

Riemsma, R. P., Taal, E., & Rasker, J. J. (2000). Perceptions about perceived functional disabilities and pain of people with rheumatoid arthritis: differences between patients and their spouses and correlates with well-being. *Arthritis Care and Research*, 13, 255–261.

Robbins, J. M., Kirmayer, L. J., Cathébras, P., Yaffe, M. J., & Dworkind, M. (1994). Physician characteristics and the recognition of depression and anxiety in primary care. *Medical Care*, 32, 795–812.

Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., & Archer, D. (1979). *Sensitivity to nonverbal communication: The PONS test*. Baltimore: The Johns Hopkins University Press.

Roter, D. L. (1995). *The Roter Method of Interaction Process Analysis: RIAS manual*. Baltimore: John Hopkins University.

Rydon, P., Redman, S., Sanson Fisher, R. W., & Reid, A. L. (1992). Detection of alcohol related problems in general practice. *Journal of Studies on Alcohol*, 53, 197–202.

Saraceno, B., Laviola, F., Sternai, E., Terzian, E., & Tagnoni, G. (1994). Consequences of mental distress recognition in general-practice in Italy - A follow-up-study. *Social Science & Medicine*, 39, 789–796.

Schafheutle, E. I., Cantrill, J. A., & Noyce, P. R. (2001). Why is pain management suboptimal on surgical wards? *Journal of Advanced Nursing*, 33, 728–737.

Scherer, K. R., & Scherer, U. (2011). Assessing the ability to recognize facial and vocal expressions of emotion: construction and validation of the Emotion Recognition Index (ERI). *Journal of Nonverbal Behavior*, 35, 305–326.

Schlegel, K., Grandjean, D., & Scherer, K. R. (2014). Introducing the Geneva Emotion Recognition Test: An example of Rasch-based test development. *Psychological Assessment*, 26, 666–672.

Sitzia, J., & Wood, N. (1997). Patient satisfaction: A review of issues and concepts. *Social Science & Medicine*, 45, 1829–1843.

Snodgrass, S. E., Hecht, M. A., & Ploutz-Snyder, R. (1998). Interpersonal sensitivity: Expressivity or perceptivity? *Journal of Personality and Social Psychology*, 74, 238–249.

Street, R. L., & Haidet, P. (2010). How well do doctors know their patients? Factors affecting physician understanding of patients' health beliefs. *Journal of General Internal Medicine*, 26, 21–27.

Sutherland, J. E., Wesley, R. M., Cole, P. M., Nesvacil, L. J., Daly, M. L., & Gepner, G. J. (1988). Differences and similarities between patient and physician perceptions of patient pain. *Family Medicine*, 20, 343–346.

Teske, K., Daut, R. L., & Cleeland, C. S. (1983). Relationships between nurses' observations and patients' self-reports of pain, *Pain*, 16, 289–296.

Thielke, S. M., Diehr, P., & Unutzer, J. (2010). Prevalence, incidence, and persistence of major depressive symptoms in the cardiovascular health study. *Aging Ment Health*, 14(2), 168–176.

Tickle-Degnen, L. (1998). Working well with others: The prediction of students' clinical performance. *American Journal of Occupational Therapy*, 52, 133–142.

Van der Does, A. J. (1989). Patients' and nurses' ratings of pain and anxiety during burn wound care. *Pain*, 39, 95–101.

Wagner, J. H., Justice, A. C., Chesney, M., Sinclair, G., Weissman, S., & Rodriguez-Barradas, M. (2001). Patient and provider reported adherence: toward a clinically relevant useful approach to measuring antiretroviral adherence. *Journal of Clinical Epidemiology*, 54, S91–S98.

West, T. V., & Kenny, D. A. (2011). The truth and bias model of judgment. *Psychological Review*, 118(2), 357–378.

Wickens, T. D. (2001). *Elementary signal detection theory*. Oxford: Oxford University Press.

Zeller, A., Taegtmeyer, A., Martina, B., Battegary, E., & Tschudi, P. (2008). Physicians' ability to predict patients' adherence to antihypertensive medication in primary care. *Hypertension Research*, 31, 1765–1771.

Zimmermann, C., Del Piccolo, L., & Finset, A. (2007). Cues and concerns by patients in medical consultations: a literature review. *Psychological Bulletin*, 133, 438–463.

15 Gender differences in interpersonal accuracy

Judith A. Hall, Sarah D. Gunnery, and Terrence G. Horgan

Abstract
This chapter reviews several traditions of research and theory on gender differences in interpersonal accuracy. Females excel over males in inferring the meanings of affective cues as judged in multiple cue modalities. This difference holds across time, cultures, age groups, and target gender. Females also have greater ability in judging personality, though fewer studies are available. They also excel in remembering others' appearance and nonverbal behavior, they respond more quickly on accuracy tasks, and they have more extensive knowledge of the meanings and usages of nonverbal communication as assessed on a written test. However, little difference in judgment accuracy is seen for lie detection and the judgment of status/dominance, and for the judgment of physical pain there is evidence that males excel over females. Accuracy in other domains is also discussed, including the recognition of flirtation and romantic interest. Finally, the chapter reviews the main theoretical frameworks within which accuracy gender differences have been discussed.

Early in the history of social psychology, Buzby (1924) published a study on accuracy in judging emotions from images of faces. But even more interesting is the fact that, long before it was standard to include women in psychology research, this study highlighted a gender difference: women performed better than men in two groups of university students. In fact, in the 1920s and 1930s, there were at least a dozen published studies that compared men's and women's emotion recognition accuracy. Buzby did not offer any interpretation, but many later authors have theorized about this difference in what grew into an extensive and still-growing literature.

This chapter summarizes the research on gender differences in emotion recognition accuracy and response latency for making emotion judgments, as well as research on accurate inference about a variety of other states and traits, the accurate recall of others' appearance and behavior, and knowledge about nonverbal cues and their correlates. We also discuss moderators of gender differences and briefly review several traditions of theory that attempt to explain the differences. Fortunately, for some questions we can build on, and extend, research summaries that are already published.[1]

Accurate inferences from others' behavior

Throughout the years, some authors have expressed uncertainty about whether men and women differ in accurately inferring states and traits from others' behavior. Some of their uncertainty stemmed from investigating the question with small samples of participants and not obtaining a statistically significant difference. However, effect sizes tell a much more informative and compelling story, and even summaries of significance levels leave no doubt that in some domains of judgment there are clear and very consistent gender differences.

Hall (1978) conducted a meta-analysis of 75 studies on judgments of states (predominantly affective states) from nonverbal cues, across all age groups. In that literature, and to this day, the great majority of studies are on affective judgments, most often basic emotions expressed on the face. Hall's (1983) nonoverlapping meta-analysis of 50 results was similarly weighted heavily toward emotion cues, though that review also included somewhat more diversity (e.g., the judgment of personality traits). For this chapter, a new summary was conducted, again on different studies (all published since the 1983 review). This new review does not exhaust the huge literature that has accumulated between then and now, but it was gathered in a hopefully unbiased fashion, principally by retrieving studies from several meta-analytic databases that were not focused on gender. This review, unlike the previous ones, broke up the literature according to content domains.

Table 15.1 provides a summary of these reviews. The metric of effect size is Cohen's d (Cohen, 1988), which expresses a two-group comparison in terms of standard deviation units. Thus, d = .50 means that the two groups' means are half a standard deviation apart. A positive effect size means females scored higher than males, and a negative effect size means the reverse. For the new review, the table includes only studies involving inference about affect or emotion, as other contents will be considered later in this chapter, and it also excludes studies of children. In all of these reviews, many different tests were used, some well established, others new and promising, and some created for a specific study.

[1] "Gender" in this chapter refers to the self-presented categories of male and female, such as participants check on a demographic questionnaire.

Table 15.1 *Gender differences in inferring affective states from cues*

Summary	Known effect sizes (Cohen's *d*)	All studies[a] (Cohen's *d*)	Females better[b] (% of studies)	Females significantly better[c] (% of studies)
Hall (1978)	.40 (46)	.24 (75)	84	31, 96
Hall (1983)	.52 (18)	.18 (50)	81	20, 91
New review	.45 (37)	.16 (51)	92	65, 100

Note: Effect size is positive when females excel over males, negative when males excel over females. Number of studies is in parentheses.

[a] When effect size was not known, it was entered as .00.

[b] Direction of difference, regardless of statistical significance, for known effect sizes.

[c] First entry is for known effect sizes; second entry is for studies achieving statistical significance.

The table shows incontrovertibly, over three nonoverlapping reviews, female superiority in judging affect cues. When unknown effects are included with the imputed value of $d = .00$, the effect still favors females. Furthermore, the proportions of studies showing directional advantage for females are dramatic (third column), as are the proportions of studies finding statistical significance in favor of women (fourth column).

Adding weight to this evidence are reviews reported by others. Entered in the present review as one effect size of $d = .18$ was a meta-analysis of 42 studies of adults' accuracy on the Reading the Mind in the Eyes Test (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001) done by Kirkland, Peterson, Baker, Miller, and Pulos (2013). In addition, Merten's (2005) compilation of emotion recognition data from 13 countries was entered as one effect size of $d = .22$.

Also included to only a minimal extent in Table 15.1 are the extensive findings on gender differences using the Profile of Nonverbal Sensitivity (PONS; Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979), a multichannel test of accuracy in decoding nonverbal affective cues conveyed by face, body, and content-masked voice. The new review excluded PONS studies to be consistent with the earlier meta-analyses that had included only a handful of representative studies that used the PONS test. In a separate review of only PONS test studies, the mean *d* in 133 samples of participants who took the full PONS test was .41, with 80% showing female directional advantage (Rosenthal et al., 1979). These figures are almost identical to those in Table 15.1.

The most recent meta-analysis is that of Thompson and Voyer (2014), who focused exclusively on accuracy in judging discrete emotions. Their meta-analysis is not perfectly comparable to those described above, because they analyzed accuracy and response latency together; though they said there was no overall significant difference, the effect sizes and the *p*-value for the comparison were not given. They concluded that, overall, females excel over males, with an overall *d* of .27. The authors acknowledged that this overall effect might be attenuated due to the fact that they based their review mainly (and deliberately) on individual discrete emotions, rather than total test scores.

Having settled the general question of whether there is evidence for a gender difference in inferring the meanings of affective cues, we now turn to several moderators that have been discussed in the literature.

Age

The new review conducted for this chapter mainly on young adult samples (chiefly college). However, other reviews enable statements to be made about age. McClure's (2000) meta-analysis of facial emotion judgment accuracy in children and adolescents found superior performance in girls ($d = .18$), with some substantial effect sizes occurring even in preschool samples. Hall's (1978) meta-analysis on many tests found no evidence of an interaction of age with participant gender across the lifespan, nor did Rosenthal et al.'s (1979) analysis of children, adolescents, and adults who took the PONS test. Williams et al. (2009) found no moderation by participants' age when gender differences were examined in participants between 6 and 91 years of age, nor did Ruffman, Murray, Halberstadt, and Taumeopeau (2010) when comparing young adults (mean age = 21 years) to older adults (mean age = 71 years). However, Thompson and Voyer's (2014) meta-analysis of recognizing discrete emotions found evidence for curvilinearity, with smaller effects among children and for adults over 30 than for adolescents and younger adults.

Culture

The female advantage in this skill does not appear to be culturally bound. Merten (2005) found significant female advantage in emotion recognition in the majority of 13 countries and directional advantage in all but one; Rosenthal et al. (1979) found comparable female advantage on the PONS test when comparing 17 United States college groups to 17 college groups from several other countries; and

Kirkland et al. (2013) in their meta-analysis of the Reading the Mind in the Eyes Test similarly found no difference in the extent of female superiority when comparing samples from the United Kingdom with samples from other countries. Other authors who compared data from different countries have also found similar gender differences. Furthermore, many of the individual studies in the new review were conducted outside the United States, and they found gender differences in the female direction.

Gender of targets as a possible moderator

A question often asked in this literature is whether the perceiver gender effect depends on the gender of the people whose cues are judged. Hall's (1978) between-studies meta-analysis found no evidence for such an interaction, and furthermore Hall (1978) cited several studies that found no perceiver gender X target gender interaction within their own databases. An occasional more recent study has reported an interaction of perceiver and target gender. Rotter and Rotter (1988) found that women were less accurate than men in two studies when decoding male angry facial expressions, even though the perceiver gender main effects still favored female perceivers. In their meta-analysis of recognition of discrete emotions, Thompson and Voyer (2014) found the biggest gender difference when the targets were male, though this was based on only a handful of studies.

Cue modalities

Both of the published meta-analyses by Hall (1978, 1983) found trends for visual cues (which are predominantly facial expressions) to show somewhat larger gender effects than vocal cues, though in Hall (1978) multimodal tests showed the largest differences. The present new review, based on known results, found a different pattern: effects were strongest for the voice (mean d = .77, though based on only 3 studies), next for multimodal studies (mean d = .52, 10 studies), and least for face studies (mean d = .39, 23 studies) and the one body study (d = .39). The new review is, of course, not a comprehensive compilation of the literature, so these results must be considered in that light. Thompson and Voyer's (2014) meta-analysis found that tests involving more modalities had the largest effects.

Specific emotions

Thompson and Voyer's (2014) meta-analysis calculated the gender difference for several discrete emotions. As mentioned earlier, this meta-analysis combined both accuracy and response latency; also, it treated unknown results as having an effect size of .00. In addition, the comparisons of specific emotions did not distinguish children from adults. Therefore, there are some ambiguities in interpretation. However, the pattern was clear in showing that the biggest female advantage was for judging anger and sadness, and in general for judging negative over positive emotions, though females excelled for all emotions.

Response latency

Some studies have measured how quickly men and women register their answers on affective judgment tasks. Quicker responding could reflect a more confident and automatized response, though its relation to accuracy is not clear. In the preponderance of these studies, women respond more quickly than men, often significantly faster (e.g., Hampson, van Anders, & Mullin, 2006; Vassallo, Cooper, & Douglas, 2009).

Other content domains

Obviously, one can make accurate inferences about many states other than affective ones, and one can judge many different enduring qualities (e.g., personality traits and social attributes) as well. Researchers have not given sufficient attention to these non-affective judgments. This neglect has important theoretical implications because judging emotions and other affective states are domains in which women are socialized to have special interest and expertise (Cross & Madson, 1997), whereas other domains may not be associated to the same extent, or at all, with gender roles and expectations.

Lie detection. Aamodt and Custer (2006) conducted a meta-analysis to test for individual differences in accuracy in discriminating truth from lies. The synthesis of 53 studies showed no difference in accuracy between men and women (d = .03). When they separated their sample into two groups based on profession they found that men who work in law enforcement detected lies more accurately than women in law enforcement (d = −.10) and among those not in law enforcement there was a small difference favoring women (d = .06), but neither of these findings was statistically significant.

Thoughts and feelings. In a database of 15 studies that used the empathic accuracy paradigm for measuring accuracy (see Chapter 3), wherein perceivers guess the thoughts and feelings being experienced by target persons during a spontaneous interaction, Ickes, Gesn, and Graham (2000) found a gender difference favoring women when participants were also asked to estimate their own accuracy (d = .56), but not when they were not so instructed (d = .04), which the authors took to mean the gender difference depends on whether the task is made gender relevant. This motivational account will be discussed in a later section of this chapter. Klein and Hodges (2001) found a similar pattern using different experimental manipulations. However, the empathic accuracy paradigm can show a gender difference even when participants are not evidently primed to perform in a gender-stereotypic way (e.g., Hall, Ship et al., 2014; Thomas & Fletcher, 2003). Future research could elucidate whether the empathic accuracy paradigm is particularly susceptible to motivational effects (see Hall, Blanch et al., 2009).

Pain. Ruben and Hall (2013) measured accuracy in detecting how much pain target people said they were experiencing during acute experimentally induced ischemic pain (caused by an inflated blood pressure cuff that cuts off blood flow to the lower arm). In distinction to most of the findings in this chapter, men were significantly more accurate than women in two separate studies, though both used the same videoclips as stimuli.

Status and dominance. This domain reveals another exception to the usual gender difference. In the few available studies, women have not been more accurate than men in inferring how assertive people are or in judging which person has higher status in a dyadic setting (Schmid Mast & Hall, 2004; Schmid Mast, Hall, Murphy, & Colvin, 2003).

Intelligence. Murphy, Hall, and Colvin (2003) found that women were more accurate than men at judging intelligence (as measured with standard cognitive tests) in stimulus persons shown in one-min video clips. This effect was not moderated by the gender of the target persons.

Personality traits. There is a large literature investigating accuracy in judging personality, but studies of gender differences in accuracy for judging personality make up only a small portion of this literature. These studies largely show that women are more accurate at judging personality traits than men across many different traits (Hall, Goh, Schmid Mast, & Hagedorn, in press; Letzring, 2008 (Study 2); Letzring, 2010; Vogt & Colvin, 2003). Though for the most part these findings are consistent in their small but convincing effect sizes, the individual studies within this literature vary quite a bit in the traits that they investigated and their actual significant findings.

When looking trait by trait, there are studies that suggest that women are better judges of extraversion (e.g., Ambady, Hallahan, & Rosenthal, 1995), whereas others find no gender difference (e.g., Yeagley, Morling, & Nelson, 2007). Similarly, there is a study (Lippa & Dietz, 2000) showing that women were marginally better at judging neuroticism, whereas others found no gender difference (e.g., Carney, Colvin, & Hall, 2007). The methodology that these studies used varied in stimulus type and how they measured the different traits. This literature is too small to conduct a meta-analysis, but nonsignificant findings mentioned by the authors still tended to favor women (with the exception of neuroticism in Ambady et al., 1995).

With respect to personality judgments, Chan, Rogers, Parisotto, and Biesanz (2011) found that women were more accurate than men in judging what people are like in general (sometimes called normative or stereotype accuracy), but they were not better than men at knowing how each target person's personality was different from the average person (known as distinctive accuracy).

Flirtation and romantic interest. Accuracy in judging romantic or sexual interest is a growing literature that is heavily populated by research on bias in these judgments. Researchers (e.g., Abbey, 1982) have consistently shown that men tend to perceive more sexual interest in others than women do. Simply put, men interpret nonverbal cues (of friendliness) as signals of sexual interest more often than women do.

Many studies investigate gender differences in this bias without having a truth criterion with which to compute accuracy, but there are a handful of studies that have examined accuracy in making these judgments. One finding points to men having a greater insensitivity to nonverbal cues with equal numbers of misses (saying cues indicate friendliness when they are flirtatious) and false alarms (saying cues indicate flirtation when they are friendly) (Farris, Treat, Viken, & McFall, 2008). Place, Todd, Penke, and Asendorpf (2009) found that both males and females were more accurate in decoding the interest cues of men than women, indicating that women might be less expressive in communicating these cues, which perhaps accounts for some of men's insensitivity to women's cues.

Judging social attributes. The Interpersonal Perception Task (IPT; Costanzo & Archer, 1989) shows generally minimal gender differences as reviewed by Hall and Schmid Mast (2008). On the IPT, participants make judgments of lie versus truth, kinship, intimacy, status, and competitive outcomes. If lie detection is gender neutral, recognizing kinship and intimacy are female stereotypic, and recognizing status and competitive advantage are male stereotypic; it is not surprising that the total score on the IPT shows a minimal gender difference. Furthermore, the lack of difference for detecting lies and for judging competitive advantage (which is similar to judging dominance or power) is consistent with those literatures as reviewed above.

Judging sexual orientation has received considerable interest, but these studies largely do not report on gender differences for the skill. However, Ambady, Hallahan, and Conner (1999) conducted a meta-analysis of six studies (including four that were reported on in the same publication as the meta-analysis) and found a gender difference favoring female decoders. Ambady et al. (1999) also found an interaction between perceiver gender and perceiver sexual orientation in which homosexual women were more accurate than heterosexual women, and homosexual men and heterosexual men did not differ. Their article also reported a gender difference among target people, with all perceivers achieving more accuracy when judging women than men.

Finally, Driscoll, Kelly, and Henderson (1998) found that women were more accurate than men at detecting the degree to which male targets reported having a tendency to sexually harass.

Accuracy of recalling appearance and behavior

Everyday life involves many first encounters in which appearance cues contribute to the impressions or attributions made. The literature is replete with evidence that perceivers use appearance cues – physical features of the face and body and clothes and artifacts – to judge others' affective states, personality traits, and social characteristics (e.g., Borkenau & Liebler, 1995; Ekman, Sorenson, & Friesen, 1969; Feinberg, Mataro, & Burroughs, 1992). Because noticing a person's appearance cues would likely precede any specific judgments (Funder, 1995), the question naturally arises as to whether women more thoroughly process appearance cues initially, or use these cues better in their inferences about others, or both.

Only the first part of this question has been investigated to date. Specifically, when men and women see (or read about) someone for the first time, do women have better memory for what that person looks like (i.e., better appearance accuracy)? A female advantage over males in appearance accuracy would suggest a possible gender difference in the initial processing of appearance cues.

Eleven social psychological studies (we have excluded eyewitness studies and studies in which there either was no mention of the appearance items tested or only a few appearance items were sampled by the researchers) have explored whether a gender difference exists in appearance accuracy (Hall & Schmid Mast, 2008; Horgan, McGrath, & Long, 2009; Horgan, Schmid Mast, Hall, & Carter, 2004; Schmid Mast &

Hall, 2006). Findings indicate that women outperform men under a variety of experimental setups and instructions as well as when different stimulus materials and answer formats are used (median d = .43). For instance, men's and women's memory for two people in their environment has been examined under circumstances in which they are told to pay attention to only one of the people (i.e., target) because their memory for that target's appearance would be tested later (Horgan et al., 2009). Women showed better memory for the target and nontarget (i.e., the other person), evidence that women's appearance accuracy is superior to men's both under directed- and incidental-learning conditions.

Women also have shown better appearance recall accuracy than men when a target's appearance was described to them in a vignette, when they were shown targets in color slides or videotape, and when the targets were an actual interaction partner of theirs, suggesting generality to the finding in different stimulus situations (Hall & Schmid Mast, 2008; Schmid Mast & Hall, 2006). In terms of answer formats, Hall and Schmid Mast asked men and women to write down what they recalled about a target's appearance, Horgan et al. (2004) used multiple-choice questionnaires, and Schmid Mast and Hall (2006) used a nonverbal response format that required men and women to recognize a previously seen slide of a target among four distractor slides of that same target with slightly varying appearance cues. Irrespective of the answer format, women more accurately remembered the appearance of targets than men.

Finally, Hall, Murphy, and Schmid Mast (2006) conducted several studies in which perceivers were asked to remember the nonverbal cues emitted by target persons seen on videotape. Across five studies, women showed more accurate recall than men (d = .26), contradicting some earlier research (see Hall et al., 2006).

Women's advantage over men in memory for people's appearance and behavior might be a byproduct of women being more aware of their surroundings or having superior episodic memory relative to men (Herlitz & Loven, 2013; McGivern et al., 1998). If so, the nature of the to-be-recalled information in the environmental context should not matter. Women would be expected to recall the environmental setting (e.g., the people and objects in it) as well as what transpired (e.g., what people discussed) in that setting better than men. However, the evidence suggests that women's advantage in memory for appearances and behavior may be better explained by their greater interpersonal orientation (Cross & Madson, 1997).

Compared to men, women are more perceptually drawn to people than to objects (Jobson & Watson, 1984). When men's and women's recognition memory for previously viewed faces and cars was tested, women's advantage over men was limited to the faces (McKelvie, Standing, St. Jean, & Law, 1993). When men's and women's memory for the people and objects in their surroundings was tested, women had better memory for only the people; men remembered the objects as well as women (Horgan et al., 2009). Finally, in a study dealing with memory for others' verbal statements, women recalled shared relationship (i.e., others' family members) information better than did men; men's memory was as good as women's for non-relationship information, such as the person's favorite food or exercise routine (Horgan, Stein, Southworth, & Swarbrick, 2012).

Accurate knowledge about nonverbal cues

Although not directly measuring perception of specific others' behavior, another line of research is also relevant. Rosip and Hall (2004) developed a paper-and-pencil test of factual knowledge about the meanings and functions of nonverbal behavior, which was scored for accuracy against established findings in the literature (Test of Accurate Nonverbal Cue Knowledge (TONCK)). Across four studies, women scored higher on the TONCK than men did ($d = .36$).

The question of magnitude

The research reviewed in this chapter clearly indicates that there are gender differences in several kinds of interpersonal perception accuracy. Though these are credible differences, we should still ask how big they are, for this would help us to understand the importance of the differences in everyday life. In absolute magnitude, the differences are small. For example, a gender difference of $d = .40$ explains only 4% of variation. Informative though this absolute standard is, a more nuanced understanding of the magnitude of effects is obtained by a comparative approach.

Therefore, we ask how big the differences are compared to other gender effects in social psychology, and compared to other correlates of interpersonal accuracy. Evidence for the first question was provided by Richard, Bond, and Stokes-Zoota (2003) in their large review of meta-analyses in social-personality psychology. Across 83 meta-analyses on gender differences, the average effect was $d = .24$, smaller than many of the accuracy gender differences.

Hall (2006) addressed the second question – comparison to other correlates of interpersonal accuracy. For a wide variety of social-personality correlates of interpersonal judgment accuracy other than gender, the average absolute point biserial r was .18 (d = .36, 112 studies), smaller than many of the gender differences in interpersonal judgment accuracy. In later meta-analyses of correlates of interpersonal judgment accuracy (other than gender), effects were of similarly modest magnitude (psychosocial correlates, including personality: Hall, Andrzejewski, & Yopchick, 2009; intelligence: Murphy & Hall, 2011). Thus, gender is correlated with interpersonal accuracy as well as, or better than, many other variables are. Even though gender differences in interpersonal accuracy are small in absolute magnitude, they are not so small by several comparison standards. According to Cohen (1988), effects of this size are typically visible to the naked eye, which may explain why women are stereotyped to be more accurate perceivers of others (Briton & Hall, 1995), and why men's and women's self-evaluations of interpersonal perception accuracy show the same difference (Zuckerman & Larrance, 1979).

Theoretical issues

Three themes predominate in discussions of the sources of gender differences in interpersonal accuracy, namely nature versus nurture, motivation versus knowledge, and dominance/power.

Nature versus nurture

Andersen (2006) presented biologically based arguments for the origin and development of these differences. Women may, for example, have evolved to be more sensitive to nonverbal cues than men because of advantages in terms of survival of offspring.

Evolutionary psychology provides the foundation for much of the research on gender differences in flirtation and courtship. Successful reproduction represents the end goal of flirtation and courtship from this perspective. Of importance, how women and men successfully reproduce differs, with women being more cautious and selective than men because women have a greater initial parental investment in reproduction (i.e., pregnancy, nursing, infant care, etc.) and thus are more at risk (e.g., of being raped, or raising a child without the biological father's help). Evolutionary psychologists argue that women adapted to this greater risk by becoming more aware of and defensively prepared in their surroundings relative to men (McGivern et al., 1998; McKibbin et al., 2009; Meyers-Levy & Maheswaran, 1998). If so, women's enhanced memory for others' appearance and behavior might merely be a byproduct of their natural tendency to scan their surroundings, which often includes other people, more thoroughly than men, though there exists empirical evidence to counter this argument, as reviewed in an earlier section.

Although it is difficult to adjudicate biological versus social explanations, the fact that the largest and most consistent female advantage in judging cues exists for emotions suggests, as mentioned earlier, that socialization into gender-stereotypic patterns of interests and skills might be a potent factor in explaining some gender differences in accuracy.

Motivation versus knowledge

A second theoretical discussion is about the relative contributions of motivation versus knowledge. Ickes et al. (2000) proposed that women's superiority in judging interpersonal cues is closely tied to motivational factors at the time of the skill assessment (see also Chapter 3). According to this account, both women and men recognize that this skill is female stereotypic, and this recognition motivates them to behave in a gender-congruent way, with the end result that women try harder and/or men try less hard when performing such tasks. This account puts emphasis on gender differences in short-term motivation rather than on gender differences in knowledge pertinent to the judging of cues, which might in fact be equivalent between males and females.

There is research showing that increasing motivation can increase accuracy of judging or recalling interpersonal cues, though there is also research showing the opposite (Hall, Blanch et al., 2009; Smith, Ickes, Hall, & Hodges, 2011). Possibly, too, it is easier to deflate accuracy through reducing motivation to be accurate (by reducing attention, for example) than it is to increase accuracy by adding extra motivation (Hall, Blanch et al., 2009; Horgan & Smith, 2006).

Thus, whereas the evidence regarding motivation to be accurate is mixed, there is evidence that greater knowledge of meanings and functions of nonverbal cues (assessed independently from a cue-judgment task) is positively correlated with accuracy in judging nonverbal cues (Davitz et al., 1964; Rosip & Hall, 2004); and, as reported above, women do excel on knowledge of nonverbal cues and their correlates (Rosip & Hall, 2004). However, in Rosip and Hall (2004), accurate knowledge did not statistically explain women's superiority in judging nonverbal cues. Therefore, it is still not clear whether expert knowledge underlies the gender difference in performance on interpersonal accuracy tasks.

In an important respect, however, motivation undoubtedly contributes to the observed performance differences. Whereas we have thus far discussed motivation as a proximal influence on accuracy, motivation can also be a long-term influence. Over a lifetime of being motivated to respond to gender-role expectations for higher or lower emotional and social perceptivity (Cross & Madson, 1997), both women's and men's accuracy could be affected via their degree of knowledge acquisition. Thus, a distinction must be made between long-term motivation to become skillful versus one's motivational state at the time skill is assessed: the long-term effect may be potent, while the short-term effect may be unpredictable (see also Chapter 11).

Dominance/power

A lasting theoretical debate was inspired by Henley (1977), who suggested that nonverbal gender differences, including those for interpersonal accuracy, have a common origin in women's subordinate place in society. According to this view, subordinate (weak, powerless, low dominant) individuals must develop strong interpersonal perception skills as an adaptive tactic; since women are the weak ones in society, it follows that this could account for their heightened interpersonal perception skills. This sexual politics view may have validity in some circumstances, but its viability has been weakened by research into how power, status, and dominance are actually related to interpersonal accuracy (see reviews in Hall, Schmid Mast, & Latu, 2015; and Chapter 13).

Also contradicting the dominance/power theory is Merten's (2005) study of emotion recognition accuracy in more than 40,000 people in 13 countries. That author correlated the gender differences in accuracy with country-specific gender empowerment measures. Across countries, the more gender equality there was, the *bigger* was women's advantage over men in accuracy of recognizing emotions, with a much bigger relation evident between gender empowerment and accuracy for women than for men. Consistent with this, less-“subordinate” women (in terms of gender-role attitudes and domestic chores) were also better decoders of nonverbal cues in research by Hall, Halberstadt, and O'Brien (1997).

Summary and implications of gender differences in accuracy

The causes of gender differences in interpersonal accuracy may be evolutionary, motivational, knowledge based, social structural, and due to lifetime gender socialization. One common theme may be that women are more accurate in domains that are stereotypically female. This is likely why they are better at detecting emotion and personality, and at remembering the appearance and nonverbal behavior of other people, but not at accurately detecting some other qualities that one could argue are less stereotyped as domains of female expertise, such as assertiveness and status, deception, and physical pain.

If one goal of studying interpersonal accuracy is to improve people's accuracy and consequently interpersonal relations, gaining a better understanding of gender differences is important. Considering that interpersonal accuracy is positively correlated with a wide range of intrapersonal and interpersonal variables, including in workplace settings (Hall, Andrzejewski et al., 2009), differences between the genders in interpersonal accuracy may have practical implications for the welfare of individuals and organizations.

References

Aamodt, M. G., & Custer, H. (2006). Who can best catch a liar? *Forensic Examiner*, 15, 6–11.

Abbey, A. (1982). Sex differences in attributions for friendly behavior: Do males misperceive females' friendliness? *Journal of Personality and Social Psychology*, 42, 830–838.

Ambady, N., Hallahan, M., & Conner, B. (1999). Accuracy of judgments of sexual orientation from thin slices of behavior. *Journal of Personality and Social Psychology*, 77, 538.

Ambady, N., Hallahan, M., & Rosenthal, R. (1995). On judging and being judged accurately in zero-acquaintance situations. *Journal of Personality and Social Psychology*, 69, 518–529.

Andersen, P. A. (2006). The evolution of biological sex differences in communication. In K. Dindia & D. J. Canary (Eds.), *Sex differences and similarities in communication* (2nd ed., pp. 117–135). Mahwah, NJ: Lawrence Erlbaum Associates.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: A study with normal adults, and adults with Asperger syndrome or high functioning autism. *Journal of Child Psychology and Psychiatry*, 42, 241–251.

Borkenau, P., & Liebler, A. (1995). Observable attributes as manifestations and cues of personality and intelligence. *Journal of Personality*, 63, 1–25.

Briton, N. J., & Hall, J. A. (1995). Beliefs about female and male nonverbal communication. *Sex Roles*, 32, 79–90.

Buzby, D. E. (1924). The interpretation of facial expression. *American Journal of Psychology*, 35, 602–604.

Carney, D. R., Colvin, C. R., & Hall, J. A. (2007). A thin slice perspective on the accuracy of first impressions. *Journal of Research in Personality*, 41, 1054–1072.

Chan, M., Rogers, K. H., Parisotto, K. L., & Biesanz, J. C. (2011). Forming first impressions: The role of gender and normative accuracy in personality perception. *Journal of Research in Personality*, 45, 117–120.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Costanzo, M., & Archer, D. (1989). Interpreting the expressive behavior of others: The Interpersonal Perception Task. *Journal of Nonverbal Behavior*, 13, 225–245.

Cross, S. E., & Madson, L. (1997). Models of the self: Self-construals and gender. *Psychological Bulletin*, 122, 5–37.

Davitz, J. R., Beldoch, M., Blau, S., Dimitrovsky, L., Levitt, E., Kempner Levy, P., Mattis, S., & Turner, J. Le B. (1964). Personality, perceptual, and cognitive correlates of emotional sensitivity. In J. R. Davitz (Ed.), *The communication of emotional meaning* (pp. 57–68). New York: McGraw-Hill.

Driscoll, D. M., Kelly, J. R., & Henderson, W. L. (1998). Can perceivers identify likelihood to sexually harass? *Sex Roles*, 38, 557–588.

Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science*, 164, 86–88.

Farris, C., Treat, T. A., Viken, R. J., & McFall, R. M. (2008). Sexual coercion and the misperception of sexual intent. *Clinical Psychology Review*, 28, 48–66.

Feinberg, R. A., Mataro, L., & Burroughs, W. J. (1992). Clothing and social identity. *Clothing and Textiles Research Journal*, 11, 18–23.

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, 102, 652–670.

Hall, J. A. (1978). Gender effects in decoding nonverbal cues. *Psychological Bulletin*, 85, 845–857.

Hall, J. A. (1983). *Nonverbal sex differences: Communication accuracy and expressive style*. Baltimore: Johns Hopkins University Press.

Hall, J. A. (2006). How big are nonverbal sex differences? The case of smiling and sensitivity to nonverbal cues. In K. Dindia and D. J. Canary (Eds.), *Sex differences and similarities in communication* (2nd ed., pp. 59–81). Mahwah, NJ: Lawrence Erlbaum Associates.

Hall, J. A., Andrzejewski, S. A., & Yopchick, J. E. (2009). Psychosocial correlates of interpersonal sensitivity: A meta-analysis. *Journal of Nonverbal Behavior*, 33, 149–180.

Hall, J. A., Blanch, D. C., Horgan, T. G., Murphy, N. A., Rosip, J. C., & Schmid Mast, M. (2009). Motivation and interpersonal sensitivity: Does it matter how hard you try? *Motivation and Emotion*, 33, 291–302.

Hall, J. A., Goh, J. X., & Schmid Mast, M., & Hagedorn, C. (in press). Individual differences in accuracy of assessing personality from text. *Journal of Personality*. Hall, J. A., Halberstadt, A. G., & O’Brien, C. E. (1997). “Subordination” and nonverbal sensitivity: A study and synthesis of findings based on trait measures. Sex Roles, 37, 295–317.

Hall, J. A., Murphy, N. A., & Schmid Mast, M. (2006). Recall of nonverbal cues: Exploring a new definition of interpersonal sensitivity. *Journal of Nonverbal Behavior*, 30, 141–155.

Hall, J. A., & Schmid Mast, M. (2008). Are women always more interpersonally sensitive than men? Impact of content domain and motivation. *Personality and Social Psychology Bulletin*, 34, 144–155.

Hall, J. A., Schmid Mast, M., & Latu, I. (2015). The vertical dimension of social relations and accurate interpersonal perception: A meta-analysis. *Journal of Nonverbal Behavior*, 39, 131–163.

Hall, J. A., Ship, A. N., Ruben, M. A., Curtin, E. M., Roter, D. L., Clever, S. L., . . . & Pounds, K. (2014). The Test of Accurate Perception of Patients' Affect (TAPPA): An ecologically valid tool for assessing interpersonal perception accuracy in clinicians. *Patient Education and Counseling*, 94, 218–223.

Hampson, E., van Anders, S. M., & Mullin, L. I. (2006). A female advantage in the recognition of emotional facial expressions: Test of an evolutionary hypothesis. *Evolution and Human Behavior*, 27, 401–416.

Henley, N. M. (1977). *Body politics: Power, sex, and nonverbal communication*. Englewood Cliffs, NJ: Prentice-Hall.

Herlitz, A., & Loven, J. (2013). Sex differences and the own-gender bias in face recognition: A meta-analytic review. *Visual Cognition*, 21, 1306–1336.

Horgan, T. G., McGrath, M. P., & Long, J. A. (2009). The relevance of people versus objects in explaining women's advantage over men in appearance accuracy. *Sex Roles*, 60, 890–899.

Horgan, T. G., Schmid Mast, M., Hall, J. A., & Carter, J. D. (2004). Gender differences in memory for the appearance of others. *Personality and Social Psychology Bulletin*, 30, 185–196.

Horgan, T. G., & Smith, J. L. (2006). Interpersonal reasons for interpersonal perceptions: Gender-congruent purpose goals and nonverbal judgment accuracy. *Journal of Nonverbal Behavior*, 30, 127–140.

Horgan, T.G., Stein, J. M., Southworth, J., & Swarbrick, M. (2012). Gender differences in memory for what others say about themselves and their family members. *Journal of Individual Differences*, 33, 169–174.

Ickes, W., Gesn, P. R., & Graham, T. (2000). Gender differences in empathic accuracy: Differential ability or differential motivation? *Personal Relationships*, 7, 95–109.

Jobson, S., & Watson, J. S. (1984). Sex and age differences in choice behavior: The object-person dimension. *Perception*, 13, 719–724.

Kirkland, R. A., Peterson, E., Baker, C. A., Miller, S., & Pulos, S. (2013). Meta-analysis reveals adult female superiority in "Reading the Mind in the Eyes Test." *North American Journal of Psychology*, 15, 121–146.

Klein, K. J. K., & Hodges, S. D. (2001). Gender differences, motivation, and empathic accuracy: When it pays to understand. *Personality and Social Psychology Bulletin*, 27, 720–730.

Letzring, T. D. (2008). The good judge of personality: Characteristics, behaviors, and observer accuracy. *Journal of Research in Personality*, 42, 914–932.

Letzring, T. D. (2010). The effects of judge-target gender and ethnicity similarity on the accuracy of personality judgments. *Social Psychology*, 41, 42–51.

Lippa, R. A., & Dietz, J. K. (2000). The relation of gender, personality, and intelligence to judges' accuracy in judging strangers' personality from brief video segments. *Journal of Nonverbal Behavior*, 24, 25–43.

McClure, E. B. (2000). A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychological Bulletin*, 126, 424–453.

McGivern, R. F., Mutter, K. L., Anderson, J., Wideman, G., Bodnar, M., & Huston, P. J. (1998). Gender differences in incidental learning and visual recognition memory: Support for a sex difference in unconscious environmental awareness. *Personality and Individual Differences*, 25, 223–232.

McKelvie, S. J., Standing, L., St. Jean, D., & Law, J. (1993). Gender differences in recognition memory for faces and cars: Evidence for the interest hypothesis. *Bulletin of the Psychonomic Society*, 31, 447–448.

McKibbin, W. F., Shackelford, T. K., Goetz, A. T., Bates, V. M., Starratt, V. G., & Miner, E. J. (2009). Development and Initial Psychometric Assessment of the Rape Avoidance Inventory. *Personality and Individual Differences*, 39, 336–340.

Merten, J. (2005). Culture, gender and the recognition of the basic emotions. *Psychologia*, 48, 306–316.

Meyers-Levy, J., & Maheswaran, D. (1998). Mixed messages. How men and women differ in their responses to marketing messages. *Capital Ideas*, 1(3), 7–8.

Murphy, N. A., & Hall, J. A. (2011). Intelligence and nonverbal sensitivity: A meta-analysis. *Intelligence*, 39, 54–63.

Murphy, N. A., Hall, J. A., & Colvin, C. R. (2003). Accurate intelligence assessments in social interaction: Mediators and gender effects. *Journal of Personality*, 71, 465–493.

Place, S. S., Todd, P. M., Penke, L., & Asendorpf, J. B. (2009). The ability to judge the romantic interest of others. *Psychological Science*, 20, 22–26.

Richard, F. D., Bond, C. F., Jr., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology*, 7, 331–363.

Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., & Archer, D. (1979). *Sensitivity to nonverbal communication: The PONS test*. Baltimore: Johns Hopkins University Press.

Rosip, J. C., & Hall, J. A. (2004). Knowledge of nonverbal cues, gender, and nonverbal decoding accuracy. *Journal of Nonverbal Behavior*, 28, 267–286.

Rotter, N. G., & Rotter, G. S. (1988). Sex differences in the encoding and decoding of negative facial expressions. *Journal of Nonverbal Behavior*, 12, 139–148.

Ruben, M. A., & Hall, J. A. (2013). "I know your pain": Proximal and distal predictors of pain detection accuracy. *Personality and Social Psychology Bulletin*, 39, 1346–1358.

Ruffman, T., Murray, J., Halberstadt, J., & Taumeopeau, M. (2010). Verbosity and emotion recognition in older adults. *Psychology and Aging*, 25, 492–497.

Schmid Mast, M., & Hall, J. A. (2004). Who is the boss and who is not? Accuracy of judging status. *Journal of Nonverbal Behavior*, 28, 145–165.

Schmid Mast, M., & Hall, J. A. (2006). Women's advantage at remembering others' appearance: A systematic look at the why and when of a gender difference. *Personality and Social Psychology Bulletin*, 32, 353–364.

Schmid Mast, M., Hall, J. A., Murphy, N. A., & Colvin, C. R. (2003). Judging assertiveness. *Facta Universitatis*, 2, 731–744.

Smith, J., Ickes, W., Hall, J. A., & Hodges, S. (Eds.). (2011). *Managing interpersonal sensitivity: Knowing when and when not to understand others*. New York: Nova Science Publishers.

Thomas, G., & Fletcher, G. J. O. (2003). Mind-reading accuracy in intimate relationships: Assessing the roles of the relationship, the target, and the judge. *Journal of Personality and Social Psychology*, 85, 1079–1094.

Thompson, A. E., & Voyer, D. (2014). Sex differences in the ability to recognize non-verbal displays of emotion: A meta-analysis. *Cognition and Emotion*, 28, 1164–1195.

Vassallo, S., Cooper, S. L., & Douglas, J. M. (2009). Visual scanning in the recognition of facial affect: Is there an observer sex difference? *Journal of Vision*, 9, 1–10.

Vogt, D. S., & Colvin, C. R. (2003). Interpersonal orientation and the accuracy of personality judgements. *Journal of Personality*, 71, 267–295.

Williams, L. M., Mathersul, D., Palmer, D. M., Gur, R. C., Gur, R. E., & Gordon, E. (2009). Explicit identification and implicit recognition of facial emotions: 1. Age effects in males and females across 10 decades. *Journal of Clinical and Experimental Neuropsychology*, 31, 257–277.

Yeagley, E., Morling, B., & Nelson, M. (2007). Nonverbal zero-acquaintance accuracy of self-esteem, social dominance orientation, and satisfaction with life. *Journal of Research in Personality*, 41, 1099–1106.

Zuckerman, M., & Larrance, D. T. (1979). Individual differences in perceived encoding and decoding abilities. In R. Rosenthal (Ed.), *Skill in nonverbal communication: Individual differences* (pp. 171–203). Cambridge, MA: Oelgeschlager, Gunn & Hain.

16 Interpersonal accuracy in relation to culture and ethnicity

Hillary Anger Elfenbein and Elizabeth A. Luckman

Abstract

Classic studies by Ekman and Izard provided early evidence for the cross-cultural universality of emotion recognition, through a set of studies that were later examined from the perspective of the cultural differences they also reveal. The body of evidence as a whole supports a middle ground, suggesting that both emotional expression and its perception show basic similarities across cultures and yet meaningful differences as well. We discuss both spontaneous and motivated processes in both emotional expression and recognition. Further, this chapter attempts to review this material in terms of Brunswik's lens model, which emphasizes the creation of observable cues and their interpretation by others. We also discuss cultural differences that can arise at multiple stages of the emotion process beyond emotional expression and recognition. Namely, individuals across groups can respond differently to nonverbal cues of emotion, which involves differences in the subjective interpretation of events via cognitive appraisal, differences in internal experience, and differences in emotion regulation. These, in turn, can influence accuracy in judging emotion cues across cultures.

Introductory psychology textbooks tell the tale of Paul Ekman (1972) and Carroll Izard (1971) as they traveled the world, showing a set of black-and-white photographs depicting American facial expressions to people from many cultures. Their goal was to determine whether those expressions would be recognized accurately across the globe. Since then, the hypothesis that emotion recognition is universal has been supported, challenged, reconsidered, and incorporated into increasingly integrated theoretical perspectives. Decades since the original work, there is now a large body of evidence to understand how people from distinct cultures and ethnicities express and recognize emotional states more vs. less

The two authors contributed equally to this chapter, and appear in an arbitrary order.

accurately. This chapter attempts to provide a succinct review of this research.

This chapter starts by considering broadly the findings related to cross-cultural universality and differences in emotion expression and recognition. In order to delve further into these findings, we make use of Brunswik's (1955) lens model to examine the process by which emotional states are expressed and perceived by others. Accuracy across cultures is influenced by the flow of emotion information cues. Using this model as our foundation, first, we discuss emotional expression, which is the signal that a target sends via nonverbal cues. Emotion expression includes both spontaneous emotion cues as well as cues that are more consciously regulated. Second, we discuss emotion recognition accuracy, again making a distinction between spontaneously receiving information as well as conscious motivation to interpret emotional cues in a particular way. The story unfolds while discussing both core theoretical concepts underlying the social perception process and evidence for the effects of accuracy of emotion expression and recognition in the context of differing cultures.

Cross-cultural universality versus cross-cultural differences in emotion

Since the time of Ekman and Izard's early groundbreaking work, many researchers have conducted studies in which emotional expression stimuli have been judged within and across cultural groups. Elfenbein and Ambady (2002) conducted a large-scale meta-analysis of this body of work, including 182 independent samples in 87 articles. These studies that had been conducted across the decades varied in methods, channel of communication, emotional categories, and degree of contact between the groups. In these studies, there was substantial accuracy in recognizing emotional expressions across cultural boundaries. Indeed, across the 182 samples, only one failed to reach accuracy levels greater than that expected by chance guessing, in which members of the isolated Bahinemo tribe reported that all the faces they saw of Americans appeared angry to them (Sorensen, 1975). As much as these data revealed evidence for universality, most studies also provided evidence for group differences. In particular, there was typically an in-group advantage, in that participants in the culture from which stimuli originated typically outperformed the other cultural groups that were tested (Elfenbein, 2013). Note that they found an interesting asymmetry when it came to cultural groups that lived within the same national boundary: groups in the numerical majority were at a substantially greater risk of misunderstanding their minority group neighbors than the reverse. This may result from power differences that could make minority group members more motivated or could relate to the sheer numerical difference that allows greater exposure to majority vs. minority group members.

These findings are less controversial than their interpretation. In attempting to parse what it means for emotion to be "universal," Russell (1991) argued that minimal universality would mean that emotion cues do not become entirely unrecognizable across cultural borders, whereas strict universality would mean that there are no cultural differences. The data that have been accumulated for review support a middle ground between these two extremes.

The finding of in-group advantage has been explained in terms of dialect theory (Elfenbein, 2013; Marsh, Elfenbein, & Ambady, 2003). Tomkins and McCarter (1964) wrote that cultural differences in emotional expression are like "dialects" of the "more universal grammar of emotion" (p. 127), just as linguistic dialects can differ subtly in their accents, grammar, and vocabulary—such as American vs. British English. As in verbal language, it can be more challenging to understand someone speaking a different dialect. Although the dialects of a language are still mutually intelligible, some of the meaning can get lost along the way. While arguing that different cultures have slightly different nonverbal cues used when expressing an emotion, research on dialect theory has demonstrated that there can be culture-specific elements in expressive style and that familiarity with these culture specific elements leads to greater accuracy (Elfenbein, Beaupré, Lévesque, & Hess, 2007; Elfenbein, Mandal, Ambady, Harizuka, & Kumar, 2004). The dialect theory follows closely from Brunswik's (1956) lens model, in that cultures can vary from each other in their schemas for both displaying and utilizing cues. Accuracy is maximized when these two schemas match each other—that is, when the perceiver's implicit theories for interpretation match the target's implicit theories for display.

Cross-cultural recognition accuracy can vary across emotional states. In their meta-analysis, Elfenbein and Ambady (2002) found that cross-cultural accuracy was greatest for happiness, surprise, sadness, and anger, and lowest for contempt and disgust. In-group advantage was greatest for disgust and fear, and smallest for happiness and judgment of positive–negative valence. Contempt, which is interesting to consider due to the controversy regarding its place as a potential basic emotion (Russell, 1991; Tracy & Randles, 2011), tends to be one of the most poorly recognized emotions across cultures (Elfenbein & Ambady, 2002). Recently, there has been increasing attention beyond the so-called basic emotions—which encompass anger, disgust, fear, happiness, sadness, surprise, and sometimes contempt. The self-conscious emotions of pride and shame are also accurately recognized across cultures (Tracy & Matsumoto, 2008). Cross-cultural recognition accuracy can also vary across channel of nonverbal communication. Happiness appears to be better recognized across cultures through the face than voice, and happiness showed large in-group advantage via the voice but relatively less through the face (Elfenbein & Ambady, 2002). One might speculate that angry people need to be able to convey verbal information such as demands, simultaneously while expressing their emotional states through nonverbal cues, whereas happy people have relatively less urgency to speak.

In the following section, we take a step back to examine the process of emotional information transmission more broadly. It begins by discussing a model that describes how emotion cues form an interpersonal process between individuals, in which they are expressed by one individual and perceived by the other. It continues by examining each step of the process more fully.

The lens model: emitting and perceiving cues

Central to research on emotion recognition—and to research on social judgment more generally—has been Brunswik's (1956) classic lens model. The key insight of the lens model is that people can perceive the world only indirectly. There are cues in the environment that are probabilistically related to properties of the world, and perceivers make use of these cues probabilistically. For example, the height of a building may not be measurable readily from the street, but there are useful cues available, even if these cues are imperfect. Individuals can observe the size of the shadow on the ground or count the number of stories. Perceivers use these observable cues in an attempt to understand the properties of their world. They can vary in their accuracy due to differences in what cues are available to them and how they interpret these cues. Some cues are more diagnostic than others and some are more easily detected than others. These two factors—that is, the validity of cues and the use of cues—together combine to determine the accuracy of social judgments. Nonverbal communication provides an example where cues may be more open to varying interpretations, which increases the likelihood of perception errors. In sum, the lens model as applied to communication via nonverbal cues can help us understand how a target emits social cues into the environment, and how a perceiver attempts to interpret these cues for their underlying meaning.

The accuracy of social perception is a function of both processes—that is, the presence of diagnostic cues as well as their effective utilization.

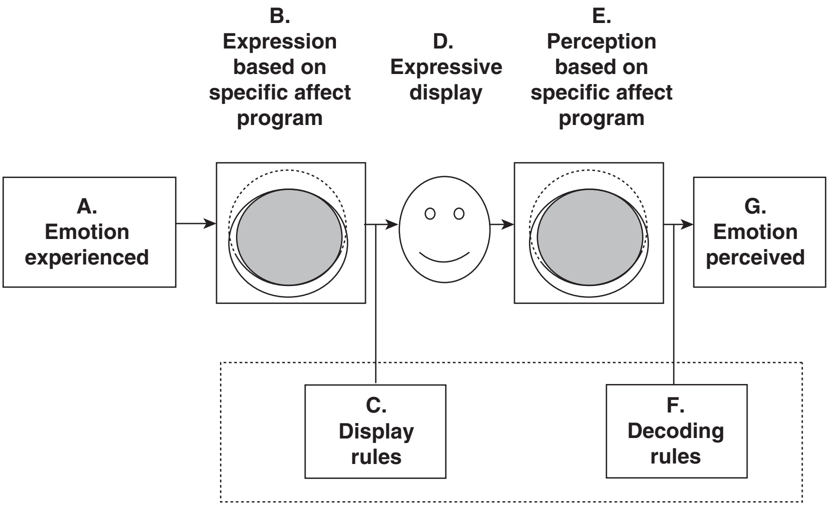

Figure 16.1 Representation of a lens model (Brunswik, 1956) of emotional expression and perception across cultures. © 2003 Hillary Anger Elfenbein.

Figure 16.1 illustrates a lens model approach to studying nonverbal communication of emotion across cultures. The left side of the model focuses on expression, which is also called encoding or the emission of cues. Encoders experience or wish to display a particular emotional state. They convey this state using a variety of cues, such as facial muscles, acoustical properties in the voice, or body movement. The right side of the model focuses on perception, which is also called decoding or the interpretation of cues. Decoders attempt to understand the speaker's emotional state through implicit analysis of these cues. As with other forms of communication, both expressing and interpreting nonverbal cues come with the risk of inaccurate judgment and misinterpretation of those signals. We will discuss in turn each of these in sides of the mirror-image process.

Cultural and ethnic differences can create barriers to accuracy in perceiving emotion. Although a great deal of the signal makes its way through—across diverse groups and even across species—some of the signal can get lost along the way. Research findings across a century and a half support the idea that the appearance of emotion expression has at least basic universality and has evolved biologically, notably through work comparing human and nonhuman emotional expressions (Darwin, 1965; Itakura, 1994; Linnankoski, Laakso, Aulanko, & Leinonen, 1994). Some of the most persuasive evidence comes from the classic round-the-world studies mentioned in the opening paragraph of this chapter (Ekman, 1972; Izard, 1971). There is also evidence that certain emotions may be more universally recognized because they serve evolutionary adaptive functions (Martens, Tracy, & Shariff, 2012; Verduyn, Van Mechelen, Tuerlinckx, & Scherer, 2013), leading to higher levels of cross-cultural accuracy. Another sizable but more recent body of research supports the role of cultural differences in emotion expression and recognition (Elfenbein, 2014; Jack, 2013; Lutz & White, 1986; Wierzbicka, 1994).

The model in Figure 16.1 emphasizes the role of both cultural universals and differences in all aspects of the emotion transmission process. These factors can exist alongside each other rather than in opposition. Each component depicted in the model includes a spontaneous process as well as an opportunity for deliberate regulation that can vary in its form across cultures. Cultural differences may affect accuracy in either the expression or recognition of emotion cues through spontaneous or motivated processes. We begin by discussing the process of emotional expression, and the accuracy implications for both spontaneous and motivated emotion expression. Then we will turn to the process of emotional recognition, again examining the accuracy implications for both spontaneous and motivated expression.

Emotional experience and expression: creating the signal

Central to the lens model process is emotional expression, namely emitting informational cues. To provide greater context, before discussing the emotional expression process, we discuss the emotional experience itself.

Emotional experience. Potential challenges to accuracy in emotion cues begin with the emotion-eliciting event. People vary in the types of emotionally evocative stimuli they find in their environment, and also in how they interpret them, both of which can be influenced by cultural norms. Longstanding process models of emotion emphasize that emotion is *about* something—starting with a stimulus in our environment, individuals engage in subjective interpretation in order to determine how to feel (Frijda & Sundararajan, 2007). Appraisal theory argues that humans generate a personal and subjective interpretation of the events around them, rather than an objective, factual analysis (Frijda, 1986; Lazarus, 1991; Mesquita & Frijda, 1992; Scherer, 1988). Even when there are objective attributes of an event, each person perceives those attributes differently. Basic emotion theorists in psychology argue that humans are hard-wired to code events rapidly and automatically in terms of the meaning for ourselves, using a cognitive appraisal process that consists of an ordered sequence of checklists (Ekman, 1992; Frijda, 1986; Frijda & Sundararajan, 2007; Lazarus, 1991; Scherer, 1988, 1995). Examples of checklists include how novel the event is or whether the event is perceived to be fair (see Frijda & Sundararajan, 2007; Smith & Ellsworth, 1985). Although the use of these checklists is universal, individuals use them subjectively. If the subjective interpretation may differ, this presents a challenge to the accuracy of experiencing and subsequently expressing an emotion.

The specific emotions that people feel are based not only on systematic differences in life experiences (Heelas, 1984), but also on complex judgments about which reasonable people can disagree. These judgments can vary across individuals and across cultures. Although our answers to the appraisal questions are subjective and idiosyncratic, theory and evidence suggest there is a universal formula that maps these answers to categorical emotional states (Scherer & Wallbott, 1994). As such, many antecedent events lead to similar emotions across cultures. There is evidence for similarity in the types of situations that elicit, for example, jealousy (Buunk & Hupka, 1987), sadness, anger, fear, and happiness (Scherer, Matsumoto, Wallbott, & Kudoh, 1988; Scherer, Summerfield, & Wallbott, 1983). In studies comparing Japanese and North American culture, researchers found that Japanese people were more likely to experience socially engaging emotions such as friendly feelings of guilt, while North American people were more likely to experience socially disengaging emotions like pride or anger (Kitayama, Mesquita, & Karasawa, 2006). In addition, research has shown that people across cultures tend to make similar ratings of which antecedent events tend to elicit which emotions (Brandt & Boucher, 1985). For example, guilt requires a negative event that a person believes they caused, whereas fear requires a future negative event that someone believes is out of their control. How a person perceives the fairness of a situation, or how a person interprets their level of control can be subject to interpretation—which leaves room for cultural differences. For example, people from independent vs. interdependent cultures tend to show greater use of the fundamental attribution error, which involves attributing events to be under the control of the individuals involved vs. determined by situational forces outside of individuals' control (Morris & Peng, 1994). Greater attributions of control might lead to greater experience of the emotions anger, pride, and guilt—all of which require the belief that a particular person is responsible for causing an event. As such, the room for personal judgments about the world around us opens the door for culturally defined norms to influence what emotional states people tend to feel and in which situations (Mesquita & Frijda, 1992). As much as individuals share the same basic formula, they vary in how they apply it. People can differ dramatically in how they interpret events—for example, whether they "look on the bright side," how they attribute blame, how much efficacy they feel to control their life's experiences, and what they believe about social standards.

In this chapter, we emphasize the role of internal emotional experience for its subsequent role in emotional expression. People can differ in how they experience emotion because they have access to different information, and also because they vary in their schemas for interpreting that information. These schemas provide an opportunity for substantial cultural differences to emerge, due to culturally shared schemas and shared meanings developed through each group's set of norms and values (Abu-Lughod, 1999; Mesquita & Frijda, 1992). For example, in a large-scale study of 37 different countries, negative emotions appeared to last significantly longer in people when the eliciting events were incongruent with the individual's goals and self-values (Verduyn et al., 2013). Even so, anger in general tends to last for relatively long durations across many cultures, which may result from the importance of anger for triggering the awareness of threat (Marinetti, Mesquita, Yik, Cragwall, & Gallagher, 2012).

This discussion of the appraisal processes highlights the room for diversity across individuals and cultures in the emotional states that they experience. This, in turn, influences the "downstream" process of expressing those emotions, which we discuss next.

Emotional expression. With this consideration of how emotional experience emerges, we now turn attention to its expression. There is substantial room for both universality and cross-cultural specificity in this key process. A helpful framework for organizing the various influences on emotional expression is Bühler's Organon model (1990), which outlines three distinct functions (Scherer, 1988). According to this model, emotional expressions function as (a) a symptom of the state of the speaker, thereby expressing emotions, intentions, and attitudes; (b) a signal to the perceiver or to the observer, thereby serving as an appeal to produce a reaction; and (c) a symbol that represents an object or event. The first of these is also called a "push" function—with emotional expression pushing itself out—while the second and third can be considered as "pull" functions—with emotional expression attempting to pull in the other party to interact. These different functions are not mutually exclusive and can even reinforce each other over time. Notably, simple reflexes that produce reliable signals can evolve to become used deliberately (Russell, Bachorowski, & Fernández-Dols, 2003). The push function within the Organon model—namely that expressions are symptoms of internal states—has received the most attention. It is at least implicitly the focus of much of the research on the communication of emotion via nonverbal cues. This function is most closely related to that which is traditionally categorized as "expression," in that the message results from the authentic internal state of the target in a spontaneous manner. However, the Organon model also emphasizes the importance of the pull functions. There has been increasing scrutiny of the idea that nonverbal cues are direct readouts that express internal feelings (Parkinson, 2005). By contrast, emotional expressions are used as signals to produce a reaction in others in a more motivated manner (Fridlund, 1994; Owren & Rendall, 2001). In pull processes, the message is intended deliberately for the perceiver's consumption. The actor hopes that his or her audience interprets the emotional cues in a particular way, in order to receive his or her deliberate message as it was intended. We consider the implications for accuracy of emotion expression related to both the push and pull functions, by engaging in separate discussions of spontaneous and motivated processes, respectively.

Spontaneous emotion expression. Darwin (1965) is considered the intellectual parent of the modern study of emotional expression, with his work on the similarity of expression across cultures and species. Importantly, he argued that certain emotional cues may play a role in natural selection. Having an emotional repertoire can allow individuals to avoid threatening or dangerous situations and to enhance cooperation. As an example, Susskind et al. (2008) argued that there are sensory benefits to the physiological responses that humans engage in when expressing fear and anger. Fear provides a larger field of vision, faster eye movements, and an increase in nasal volume and air intake. There can be an evolutionary benefit to this physiological reaction, for helping individuals escape the situation that caused the fear. The opposite case is disgust, which leads to a closing in sensory perception, and can help to prevent sensory intake of whatever led to the disgust. These biological functions of emotional expression are representative of the push or spontaneous function, are a function of human biology and, thus, are more universal in nature (Matsumoto & Hwang, 2011).

There is an interpersonal benefit to emotional expression having basic universality. Relationships emerge in the context of a shared culture, which subsequently can affect the way in which emotions are encoded and expressed by the target (De Leersnyder, Boiger, & Mesquita, 2013). Consistent with the first function of the Organon model is the notion that nonverbal cues are emitted as a spontaneous result of our internal states, for which humans and animals evolved over time the ability to read. As seen in Figure 16.1, after an emotion is experienced, it is expressed using what Ekman (1972) described as a "specific action program." In his influential neuro-cultural theory (Ekman et al., 1987), he hypothesized a one-to-one mapping between the experience of emotional categories and the specific configuration or configurations of facial muscles used to display those emotions (Ekman, Friesen, & Hager, 2002). This mapping appears to be most useful when considered as a heuristic rather than an exact formula. Researchers have rarely found the appearance of the precise total configurations hypothesized, but they do find components of these configurations. In a notable study, Carroll and Russell (1997) examined the muscle movements in Hollywood film portrayals that won awards for fine acting. They found the professional actors rarely showed facial configurations that mapped fully onto the predicted patterns, but that many expressions included activity in at least some of the predicted muscles.

The same conclusion has been made in other studies of acted portrayals (Gosselin, Kirouac, & Doré, 1995) as well as spontaneous emotional expressions (Fernández-Dols, 1997). Taken together, evidence appears to be consistent with the notion of components theory (Frijda, 1986; Scherer, 1984), which posits an association between specific muscle movements and the checklists of cognitive appraisal that were described above. According to this theory, multiple facial elements can be redundant in conveying emotions (Carroll & Russell, 1997), and so exact entire facial configurations are not necessary for accuracy. Ekman's (1972) theory about specific action programs forms a valuable description of the emotional expression process—at least when it is adapted to loosen the assumption of a complete one-to-one mapping. The processes described above are representative of the "push" function in Bühler's (1990) model (Scherer, 1988) and are more spontaneous in nature. Next, we describe emotion expression resulting from more motivated processes.

Motivated emotion expression. There are multiple potential influences of culture on the motivated or conscious transmission of emotional cues, which align with the "pull" functions of Bühler's (1990) model (Scherer, 1988). Central to the discussion of this topic in the literature has been Klineberg (1938) and Ekman's (1972) concept of display rules. Display rules are deliberately obscuring emotion regulation techniques, in which individuals may regulate their emotional displays to conform to social norms. Ekman (1972) defined display rules as conscious management techniques to deintensify, intensify, neutralize, and mask particular emotional displays. He argued that members of each culture would express their emotions in exactly the same way if some groups were not constantly monitoring themselves and adjusting their displays to fit social norms. To emphasize the role of display rules as deliberate, Ekman and colleagues (1987) argued that individuals' faces read out their emotional cues at all times, *unless they choose to consciously control it* [italics added]. Anecdotal examples of individuals choosing deliberately to regulate their emotion expression might include bluffing in a poker game or hiding fear during a scary movie.

As discussed above, stimuli and cognitive appraisal are key to determining emotional experience and expression, and can be affected by the sociocultural environment (Mesquita & Frijda, 1992). Cultures have shared norms, values, and expectations, which are imbued into the thinking of the members and influence what is expressed. Given that the use of display rules, in which people deliberately obscure emotion regulation techniques to conform to social norms, exist in the service of social relationships, differences in social norms across cultures influence groups' particular display rules. Gross (1998) developed a process model of emotion regulation that identifies a number of steps that individuals can undertake to regulate their emotions. At the chronologically earliest stage, they can decide how to allocate their attention. Subsequently, individuals can attempt to change their emotional expression through reappraisal—i.e., reevaluating the situation to yield a potentially new interpretation—or suppression—i.e., attempting to deny their internal experience. In general, reappraisal as a strategy can lead to lower physiological responding, whereas suppression can lead to greater physiological responses due to the inhibitory processes engaged in emotional suppression (Gross, 1998). Cultural differences in norms and expectations could lead to cross-cultural differences in emotion regulation that, in turn, influence the appearance and intensity of emotional expression.

People tend to express what is important to them on a personal, normative, and cultural basis. As such, there are important differences in the base rates of what emotions people tend to express. This is important to incorporate into our understanding of cultural differences and universals in emotional expression and recognition because "practice makes perfect." Cultural groups that more openly show certain emotions may be better able to recognize them as well. In a recent meta-analysis of research on emotion expression and culture, van Hemert and colleagues (2007) identified a variety of ecological, sociopolitical, and aggregated psychological determinants of levels of emotion expression. They found higher levels of emotional expressivity in countries with looser norms, higher levels of democracy, higher individualism, and more service industry workers. By contrast, countries with tighter social norms—i.e., norms that are enforced more strictly—demonstrated lower general levels of expressiveness. They found that countries with higher levels of religiosity actually demonstrated higher levels of positive emotion, which interestingly they had hypothesized in the opposite direction. Boiger and colleagues (2013) demonstrated that elements deemed important to the national culture had an effect on the type of emotion people from that culture were likely to express. The authors compared American culture, identified as more individualistic, with Belgian culture, identified as having stronger egalitarian goals. They found support for their hypothesis that Americans would be more likely to demonstrate anger, as it is more aligned with an individualistic culture; while the egalitarian culture of Belgians would be more likely to display shame. These emotions, they argued, were contextually relevant to the distinctions in the cultures. Likewise, Mesquita (2001) found that people in more interdependent cultures were more likely to express emotions that were connected to their social worth, were more likely to represent reality, and recognized the role of relationships. People in independent cultures were less likely to focus on their social worth, more likely to represent an individual and subjective perspective of reality, and less likely to recognize the role of relationships. In studies comparing Japanese and North American culture, researchers found that Japanese people were more likely to experience socially engaging emotions like friendly feelings of guilt while North American people were more likely to experience socially disengaging emotions like pride or anger (Kitayama et al., 2006). Cultural differences can also affect an individual's emotional reactivity. One study identified that cultural factors determine the focus on the aspect of the self, whether individual or relational, and that this determined the intensity of emotional reactivity (Chentsova-Dutton & Tsai, 2010). Based on familiarity, types of cultural differences in the frequency of expressing particular emotions feed into the likelihood of recognizing them.

Emotion recognition: perceiving the signal

Returning to Brunswik's lens model (1956), we now examine the right side of the model, namely emotion recognition. The lens model emphasizes that accurate perception is a matter of detecting cues in the environment, and there needs to be a match between the style of display produced and the style expected by that of the perceiver (Jack, Caldara, & Schyns, 2012). Perceivers receive nonverbal cues and interpret them based on prior knowledge, idiosyncratic habits and preferences, and social norms. As such, the recognition of emotion cues can not only be automatic and spontaneous, but it can also be motivated and regulated.

Spontaneous emotion recognition. As discussed above, there is a substantial body of research on accuracy in recognizing emotional cues across cultures, starting with the work of Ekman (1972) and Izard (1971).

In their research designs, participants made multiple-choice judgments of images of people expressing emotion, and achieved far better performance than the recognition rates that would be expected by chance guessing alone, e.g., 16.7% for a response among six choices. Ekman and Izard interpreted this finding in favor of the universality of accuracy in emotion recognition—which is a conclusion that was initially controversial, came to be accepted, and in recent years has become understood as incomplete.

Dialect theory has attempted to provide a middle ground, by arguing that emotions can be interpreted accurately across cultures, and yet there is an in-group advantage that allows individuals to interpret emotions more accurately from their own group members. Evidence for dialect theory comes from multiple labs, and the body of findings has been increasing over time (Dailey et al., 2010; Kang & Lau, 2013; Kleinsmith, De Silva, & Bianchi-Berthouze, 2006; Thompson & Balkwill, 2006; Wickline, Bailey, & Nowicki, 2009). There has even been recent research with members of relatively isolated populations, such as a judgment study of nonlinguistic vocalizations that compared Americans with Namibian villagers (Sauter, Eisner, Ekman, & Scott, 2010). Other studies use far different methods: In-group advantage has been generated in simulations of machine learning (Laukka, Neiberg, & Elfenbein, 2014). In a two-step process that matches the expression and perception stages of Brunswik's (1956) lens model, first cross-cultural differences were detected in the acoustic expression patterns of speech that were diagnostic to distinguish one emotional state from another. In the second step, machine algorithms that were trained to recognize these expressions were more accurate when they were tested on stimuli from the same cultural origin used to train them. Accuracy suffered when the algorithms trained with expressions from one culture and yet had to recognize expressions from another. This strongly suggests the role of familiarity and learning in cross-cultural accuracy in emotion recognition. Along these lines with human participants, cultural learning appears to reduce the cross-cultural gap, such that students abroad learn over time how to recognize the expressions from their host culture (Elfenbein & Ambady, 2003). New members to a culture will learn over time the values and norms that are most important or most ubiquitous to that culture. This acculturation process has been shown to enhance the emotion recognition accuracy of new members (Prado et al., 2013). Interestingly, individuals are not only more accurate when judging in-group expressions, but they also tend to be more confident about those judgments (Beaupré & Hess, 2006).

Some research on dialect theory has also demonstrated practical implications, notably in applied psychiatry research. A number of studies over the years had demonstrated that there was greater emotional impairment for African American versus Caucasian schizophrenics, which was a source of concern that researchers attempted to explain. In attempting to explain the discrepancy, Pinkham and colleagues (2008) noted that all the stimuli in these previous studies were based on Caucasian facial expressions. When they tested both ethnic groups with stimulus material that originated from both ethnic groups, they found that this observation no longer held.

As in the discussion above about the influence of base rates on emotional expression, the base rates of particular emotions in the social environment can influence accuracy in emotion perception. Some groups have greater familiarity with some emotions than others. In terms of recognizing expressions, groups vary in their opportunities for practicing and learning over time. Particularly in the case of ambiguity, this type of Bayesian processing provides a heuristic to judge a person as experiencing a state that seems to them statistically more probable. In an extreme example, Umiltà, Wood, Loffredo, Ravera, and Gallese (2013) tested a sample of survivors of Civil war in Sierra Leone and found that, relative to other emotions, participants were particularly likely to judge sad displays as anger, and yet also judge other emotions as sadness. This effect was pronounced for participants who had been child soldiers. Over time, individuals can learn to avoid the recognition of emotions that are not productive for their particular environments. As such, individuals can vary across cultures in which emotional states they judge more accurately.

Motivated emotion recognition. Up until this point, emotion recognition has been described somewhat passively, as something that happens to the perceiver when in the presence of emotion cues. However, the perceiver is an active participant in the communication process. Perceivers can be more vs. less interested in perceiving a signal and can have their own opinions about what they wish the signal to be.

Out-group bias exists when individuals are less motivated to understand the emotions of people from visibly foreign cultural groups. This can result from indifference or even lack of caring (Hugenberg, Miller, & Claypool, 2007). Evidence for out-group bias comes from studies that use stimuli in which the nonverbal cues of emotion are exactly identical across cultural groups, and yet participants still achieve higher accuracy when judging individuals from their own group. In such cases, nonverbal dialects cannot explain away higher levels of in-group accuracy. The out-group bias effect has been found with real cultural groups (van der Schalk et al., 2011), minimal groups (Young & Hugenberg, 2010), and even false feedback about group membership (Thibault, Bourgeois, & Hess, 2006).

Decoding rules can also influence emotion recognition across cultures. Matsumoto (1989) extended the concept of display rules to coin the term "decoding rules," which are norms for deliberately deceptive regulation in perceiving others' emotions. He argued that Americans are simply more effective at recognizing emotions because Americans do not suppress their true understanding of emotional displays out of concern for group harmony. In this sense, like any other type of information, people can perceive what they want to perceive. As such, decoding rules can be a "flip side" of display rules. In an act of dyadic emotion regulation, people can assist each other with conforming to norms. Individuals who display inappropriate emotions can be assisted by other people who interpret them the way ideally they should have behaved. This notion fits within theories of ideal affect (Tsai, 2007), which describes the way that people want ideally to feel—and how they want other people to feel. Individuals interpret situations actively, through cognitive appraisals that can help to conform to the internal states they desire in others. Cultural groups can also differ in the extent to which people tend to hold particular emotions in favor vs. disfavor. An example comes from a recent study comparing American and German cultures, which found cultural differences in the desire to avoid negative affect (Koopmann-Holm & Tsai, 2014). In particular, Americans were more likely to focus on the positive, while Germans were more likely to focus on the negative. There can be cultural variation in the extent to which individuals make use of contextual cues in emotion recognition. Masuda and colleagues (2008) found that people in the Japanese culture were more likely than the United States to incorporate into their judgment of a target's emotions the emotional displays of other people surrounding the target. Participants from a Western culture were more likely to ignore the emotion information of the surrounding group, and to interpret the emotion expression of the central stimulus person in terms of their individual affective response. Research increasingly points out the important and often underappreciated importance of context in forming judgments. In particular, Chinese and American subjects utilized cultural context differently when identifying emotion expression (Stanley, Zhang, Fung, & Isaacowitz, 2013). These effects can create roadblocks in judging other people's emotions even when perceivers try to be as accurate as possible. This illustrates how individuals can be motivated by the social context regarding group vs. individual norms (Masuda et al., 2008).

Interestingly, individuals can also differ across cultures in their assumptions for the reason why an emotional expression was produced in the first place. Again, we can consider the functions of emotional expression as delineated by Bühler's (1990) Organon model (Scherer, 1988), and consider the perceiver as an active partner in attempting to judge which of these functions ("pull") is operating. Cultures with looser social normative constraints—such as Western groups—may be more likely to assume that the push function is operating and infer that the emotion they judge truly represents the target's internal state. By contrast, individuals from cultures with greater need to constrain behavior might be more likely to anticipate that the target is self-regulating (Scherer & Brosch, 2009; Yuki, Maddux, & Masuda, 2007). Related to this is how individuals judge the authenticity of others' expressions. One often-studied judgment of authenticity is that of the Duchenne marker, which is a wrinkling around the corner of the eyes that purportedly distinguishes real from false smiles (Ekman, Davidson, & Friesen, 1990). In a study testing this physical expression of emotion among Gabonese and Mainland Chinese living in Canada, Thibault, Levesque, Gosselin, & Hess (2012) found that the Gabonese did not use the Duchenne marker at all, whereas Mainland Chinese only showed sensitivity to the marker when judging faces of French-Canadians. As such, our judgments of others' emotions are shaped by culture-specific schemas for interpreting the link between intentions and displays. In the process of judging displays, people implicitly judge not only the emotional category but also the extent to which they believe those displays resulted from push vs. pull processes.

Likewise, there can be cultural variability in which particular cues are the focus of attention, even which part of the face is the most salient. In cultures where emotional norms are driven toward more subdued emotion expression, such as Japan, perceivers focus more attention on the target's eyes, which are surrounded by facial muscles that are relatively harder to control (Jack, Blais, Scheepers, Schyns, & Caldara, 2009; Yuki et al., 2007). By contrast, in cultures where emotional norms are more open to expression, such as the United States, perceivers focus more attention on the mouth, because it is the more openly expressive part of the face that can be readily manipulated for display. In this way, United States vs. Japanese participants in these studies paid attention to the cues that were most likely to represent pull vs. push processes.

Conclusion

The perception of emotional cues across cultures has been an active and often dramatic area of research for decades, and continues to evolve. Examining the body of evidence as a whole, the heated debate needs to be replaced with calmer voices because only one conclusion can fit the data: Both emotional expression and its perception show basic similarities across cultures and yet meaningful differences as well. Cultural group membership has influences throughout the emotion process, starting with environmental stimuli, and extending to an expressor's subjective appraisal of the stimuli, internal experience, regulation, expression, and then the other party's emotion perception. Being part of the same cultural group can provide access to shared experiences and internal states, norms for expressing oneself, styles for producing expressive cues conditional on wanting to express, expectations for the appearance of cues, and schemas for interpreting them. It is important to emphasize that although there is a gap, it appears that people can overcome it. The boundaries between cultures are becoming more porous as interaction between people across distinct cultures is increasingly common and necessary. The ability to read nonverbal cues is an important element of social interaction, and we argue that continued research in this area has the potential to inform and enhance communication in the global environment.

References

Abu-Lughod, L. (1999). *Veiled sentiments: Honor and poetry in a Bedouin society*. Berkeley: University of California Press.

Beaupré, M. G., & Hess, U. (2006). An ingroup advantage for confidence in emotion recognition judgments: The moderating effect of familiarity with the expressions of outgroup members. *Personality and Social Psychology Bulletin*, 32, 16–26.

Boiger, M., Deyne, S. D., & Mesquita, B. (2013). Emotions in "the world": Cultural practices, products, and meanings of anger and shame in two individualist cultures. *Frontiers in Psychology*, 4. Available at http://dx.doi.org/10.3389/fpsyg.2013.00867.

Brandt, M. E., & Boucher, J. D. (1985). Judgments of emotions from the antecedent situations in three cultures. In R. Lagunes & Y. H. Poortinga (Eds.), *From a different perspective: Studies of behavior across cultures* (pp. 348–362). Lisse, Netherlands: Swets & Zeitlinger.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, 62, 193–217.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley: University of California Press.

Bühler, K. (1990). *Theory of language: The representational function of language*. Amsterdam: John Benjamins Publishing.

Buunk, B., & Hupka, R. B. (1987). Cross-cultural differences in the elicitation of sexual jealousy. *The Journal of Sex Research*, 23, 12–22.

Carroll, J. M., & Russell, J. A. (1997). Facial expressions in Hollywood's portrayal of emotion. *Journal of Personality and Social Psychology*, 72, 164–176.

Chentsova-Dutton, Y. E., & Tsai, J. L. (2010). Self-focused attention and emotional reactivity: The role of culture. *Journal of Personality and Social Psychology*, 98, 507–519.

Dailey, M. N., Joyce, C., Lyons, M. J., Kamachi, M., Ishi, H., Gyoba, J., & Cottrell, G. W. (2010). Evidence and a computational explanation of cultural differences in facial expression recognition. *Emotion*, 10, 874–893.

Darwin, C. (1965). *The expression of the emotions in man and animals*. Chicago: University of Chicago Press.

De Leersnyder, J., Boiger, M., & Mesquita, B. (2013). *Cultural regulation of emotion: individual, relational, and structural sources. Frontiers in Psychology*, 4.

Ekman, P. (1972). Universals and cultural differences in facial expressions of emotion. In J. Cole (Ed.), *Nebraska Symposium on Motivation, 1971* (Vol. 19, pp. 207–282). Lincoln, NE: University of Nebraska Press.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6, 169–200.

Ekman, P., Davidson, R. J., & Friesen, W. V. (1990). The Duchenne smile: Emotional expression and brain physiology: II. *Journal of Personality and Social Psychology*, 58, 342–353.

Ekman, P., Friesen, W. V., & Hager, J. C. (2002). *Facial Action Coding System investigator's guide*. Salt Lake City, UT: Research Nexus.

Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., . . . Tzavaras, A. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology*, 53, 712–717.

Elfenbein, H. A. (2013). Nonverbal dialects and accents in facial expressions of emotion. *Emotion Review*, 5, 90–96.

Elfenbein, H. A. (2014). The many faces of emotional contagion: An affective process theory of affective linkage. *Organizational Psychology Review*, 4, 326–362.

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, 128, 203–235.

Elfenbein, H. A., & Ambady, N. (2003). When familiarity breeds accuracy: Cultural exposure and facial emotion recognition. *Journal of Personality and Social Psychology*, 85, 276–290.

Elfenbein, H. A., Beaupré, M., Lévesque, M., & Hess, U. (2007). Toward a dialect theory: Cultural differences in the expression and recognition of posed facial expressions. *Emotion*, 7, 131–146.

Elfenbein, H. A., Mandal, M., Ambady, N., Harizuka, S., & Kumar, S. (2004). Hemifacial differences in the in-group advantage in emotion recognition. *Cognition and Emotion*, 18, 613–629.

Fernández-Dols, J. M. (1997). Spontaneous facial behavior during intense emotional episodes: Artistic truth and optical truth. In J. A. Russell (Ed.), *The psychology of facial expression* (pp. 255–274). Cambridge, UK: Cambridge University Press.

Fridlund, A. J. (1994). *Human facial expression: An evolutionary view*. San Diego: Academic Press.

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

Frijda, N. H., & Sundararajan, L. (2007). Emotion refinement: A theory inspired by Chinese poetics. *Perspectives on Psychological Science*, 2, 227–241.

Gosselin, P., Kirouac, G., & Doré, F. Y. (1995). Components and recognition of facial expression in the communication of emotion by actors. *Journal of Personality and Social Psychology*, 68, 83–96.

Gross, J. J. (1998). Antecedent- and response-focused emotion regulation: Divergent consequences for experience, expression, and physiology. *Journal of Personality and Social Psychology*, 74, 224–237.

Heelas, P. (1984). Emotions across cultures: Objectivity and cultural divergence. *Royal Institute of Philosophy Supplements*, 17, 21–42.

Hugenberg, K., Miller, J., & Claypool, H. M. (2007). Categorization and individuation in the cross-race recognition deficit: Toward a solution to an insidious problem. *Journal of Experimental Social Psychology*, 43, 334–340.

Itakura, S. (1994). Differentiated responses to different human conditions by chimpanzees. *Perceptual and Motor Skills*, 79, 1288–1290.

Izard, C. E. (1971). *The face of emotion*. East Norwalk, CT: Appleton-Century-Crofts.

Jack, R. E. (2013). Culture and facial expressions of emotion. *Visual Cognition*, 21, 1248–1286.

Jack, R. E., Blais, C., Scheepers, C., Schyns, P. G., & Caldara, R. (2009). Cultural confusions show that facial expressions are not universal. *Current Biology*, 19, 1543–1548.

Jack, R. E., Caldara, R., & Schyns, P. G. (2012). Internal representations reveal cultural diversity in expectations of facial expressions of emotion. *Journal of Experimental Psychology: General*, 141, 19–25.

Kang, S. -M., & Lau, A. S. (2013). Revisiting the out-group advantage in emotion recognition in a multicultural society: Further evidence for the in-group advantage. *Emotion*, 13, 203–215.

Kitayama, S., Mesquita, B., & Karasawa, M. (2006). Cultural affordances and emotional experience: Socially engaging and disengaging emotions in Japan and the United States. *Journal of Personality and Social Psychology*, 91, 890–903.

Kleinsmith, A., De Silva, P. R., & Bianchi-Berthouze, N. (2006). Cross-cultural differences in recognizing affect from body posture. *Interacting with Computers*, 18, 1371–1389.

Klineberg, O. (1938). Emotional expression in Chinese literature. *The Journal of Abnormal and Social Psychology*, 33, 517–520.

Koopmann-Holm, B., & Tsai, J. L. (2014). Focusing on the negative: Cultural differences in expressions of sympathy. *Journal of Personality and Social Psychology*, 107, 1092–1115.

Laukka, P., Neiberg, D., & Elfenbein, H. A. (2014). Evidence for cultural dialects in vocal emotion expression: Acoustic classification within and across five nations. *Emotion*, 14, 445–449.

Lazarus, R. S. (1991). Cognition and motivation in emotion. *American Psychologist*, 46, 352–367.

Linnankoski, I., Laakso, M., Aulanko, R., & Leinonen, L. (1994). Recognition of emotions in macaque vocalizations by children and adults. *Language & Communication*, 14, 183–192.

Lutz, C., & White, G. M. (1986). The anthropology of emotions. *Annual Review of Anthropology*, 15, 405–436.

Marinetti, C., Mesquita, B., Yik, M., Cragwall, C., & Gallagher, A. H. (2012). Threat advantage: Perception of angry and happy dynamic faces across cultures. *Cognition and Emotion*, 26, 1326–1334.

Marsh, A. A., Elfenbein, H. A., & Ambady, N. (2003). Nonverbal "accents": Cultural differences in facial expressions of emotion. *Psychological Science*, 14, 373–376.

Martens, J. P., Tracy, J. L., & Shariff, A. F. (2012). Status signals: Adaptive benefits of displaying and observing the nonverbal expressions of pride and shame. *Cognition and Emotion*, 26, 390–406.

Masuda, T., Ellsworth, P. C., Mesquita, B., Leu, J., Tanida, S., & Van de Veerdonk, E. (2008). Placing the face in context: Cultural differences in the perception of facial emotion. *Journal of Personality and Social Psychology*, 94, 365–381.

Matsumoto, D. (1989). Cultural influences on the perception of emotion. *Journal of Cross-Cultural Psychology*, 20, 92–105.

Matsumoto, D., & Hwang, H. S. (2011). Culture and emotion: The integration of biological and cultural contributions. *Journal of Cross-Cultural Psychology*, 43, 91–118.

Mesquita, B. (2001). Emotions in collectivist and individualist contexts. *Journal of Personality and Social Psychology*, 80, 68–74.

Mesquita, B., & Frijda, N. H. (1992). Cultural variations in emotions: A review. *Psychological Bulletin*, 112, 179–204.

Morris, M. W., & Peng, K. (1994). Culture and cause: American and Chinese attributions for social and physical events. *Journal of Personality and Social Psychology*, 67, 949–971.

Owren, M. J., & Rendall, D. (2001). Sound on the rebound: Bringing form and function back to the forefront in understanding nonhuman primate vocal signaling. *Evolutionary Anthropology: Issues, News, and Reviews*, 10, 58–71.

Parkinson, B. (2005). Do facial movements express emotions or communicate motives? *Personality and Social Psychology Review*, 9, 278–311.

Pinkham, A. E., Sasson, N. J., Calkins, M. E., Richard, J., Hughett, P., Gur, R. E., & Gur, R. C. (2008). The other-race effect in face processing among African American and Caucasian individuals with schizophrenia. *American Journal of Psychiatry*, 165, 639–645.

Prado, C., Mellor, D., Byrne, L. K., Wilson, C., Xu, X., & Liu, H. (2013). Facial emotion recognition: a cross-cultural comparison of Chinese, Chinese living in Australia, and Anglo-Australians. *Motivation and Emotion*, 38, 420–428.

Russell, J. A. (1991). Negative results on a reported facial expression of contempt. *Motivation and Emotion*, 15, 281–291.

Russell, J. A., Bachorowski, J.-A., & Fernández-Dols, J.-M. (2003). Facial and vocal expressions of emotion. *Annual Review of Psychology*, 54, 329–349.

Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences*, 107, 2408–2412.

Scherer, K. R. (1984). On the nature and function of emotions: A component process approach. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 293–317). New York: Psychology Press.

Scherer, K. R. (1988). Criteria for emotion-antecedent appraisal: A review. In V. Hamilton, G. H. Bower, & N. H. Frijda (Eds.), *Cognitive perspectives on emotion and motivation* (pp. 89–126). The Netherlands: Springer.

Scherer, K. R. (1995). In defense of a nomothetic approach to studying emotion-antecedent appraisal. *Psychological Inquiry*, 6, 241–248.

Scherer, K. R., & Brosch, T. (2009). Culture-specific appraisal biases contribute to emotion dispositions. *European Journal of Personality*, 23, 265–288.

Scherer, K. R., Matsumoto, D., Wallbott, H., & Kudoh, T. (1988). Emotional experience in cultural context: a comparison between Europe, Japan, and the US. In K. R. Scherer (Ed.), *Facets of emotion: recent research* (pp. 5–30). Hillsdale, NJ: Lawrence Erlbaum Associates.

Scherer, K. R., Summerfield, A. B., & Wallbott, H. G. (1983). Cross-national research on antecedents and components of emotion: A progress report. *Social Science Information/sur Les Sciences Sociales*, 22, 355–385.

Scherer, K. R., & Wallbott, H. G. (1994). Evidence for universality and cultural variation of differential emotion response patterning. *Journal of Personality and Social Psychology*, 66, 310–328.

Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, 48, 813–838.

Sorensen, E. R. (1975). Culture and the expression of emotion. In T. R. Williams (Ed.), *Psychological Anthropology* (pp. 361–371). Chicago, IL: Aldine.

Stanley, J. T., Zhang, X., Fung, H. H., & Isaacowitz, D. M. (2013). Cultural differences in gaze and emotion recognition: Americans contrast more than Chinese. *Emotion*, 13, 36–46.

Susskind, J. M., Lee, D. H., Cusi, A., Feiman, R., Grabski, W., & Anderson, A. K. (2008). Expressing fear enhances sensory acquisition. *Nature Neuroscience*, 11, 843–850.

Thibault, P., Bourgeois, P., & Hess, U. (2006). The effect of group-identification on emotion recognition: The case of cats and basketball players. *Journal of Experimental Social Psychology*, 42, 676–683.

Thibault, P., Levesque, M., Gosselin, P., & Hess, U. (2012). The Duchenne marker is not a universal signal of smile authenticity—But it can be learned! *Social Psychology*, 43, 215–221.

Thompson, W. F., & Balkwill, L. L. (2006). Decoding speech prosody in five languages. *Semiotica*, 158, 407–424.

Tomkins, S. S., & Mc Carter, R. (1964). What and where are the primary affects? some evidence for a theory. *Perceptual and Motor Skills*, 18, 119–158.

Tracy, J. L., & Matsumoto, D. (2008). The spontaneous expression of pride and shame: Evidence for biologically innate nonverbal displays. *Proceedings of the National Academy of Sciences*, 105, 11655–11660.

Tracy, J. L., & Randles, D. (2011). Four models of basic emotions: A review of Ekman and Cordaro, Izard, Levenson, and Panksepp and Watt. *Emotion Review*, 3, 397–405.

Tsai, J. L. (2007). Ideal affect: Cultural causes and behavioral consequences. *Perspectives on Psychological Science*, 2, 242–259.

Umiltà, M. A., Wood, R., Loffredo, F., Ravera, R., & Gallese, V. (2013). Impact of civil war on emotion recognition: the denial of sadness in Sierra Leone. *Frontiers in Psychology*, 4, 1–10.

Van der Schalk, J., Fischer, A., Doosje, B., Wigboldus, D., Hawk, S., Rotteveel, M., & Hess, U. (2011). Convergent and divergent responses to emotional displays of ingroup and outgroup. *Emotion*, 11, 286–298.

Van Hemert, D. A., Poortinga, Y. H., & van de Vijver, F. J. R. (2007). Emotion and culture: A meta-analysis. *Cognition and Emotion*, 21, 913–943.

Verduyn, P., Van Mechelen, I., Tuerlinckx, F., & Scherer, K. (2013). The relation between appraised mismatch and the duration of negative emotions: Evidence for universality. *European Journal of Personality*, 27, 481–494.

Wickline, V. B., Bailey, W., & Nowicki, S. (2009). Cultural in-group advantage: Emotion recognition in African American and European American faces and voices. *Journal of Genetic Psychology*, 170, 5–30.

Wierzbicka, A. (1994). Emotion, language, and cultural scripts. In S. Kitayama & H. R. Markus (Eds.), *Emotion and culture: Empirical studies of mutual influence* (pp. 133–196). Washington, DC: American Psychological Association.

Young, S. G., & Hugenberg, K. (2010). Mere social categorization modulates identification of facial expressions of emotion. *Journal of Personality and Social Psychology*, 99, 964–977.

Yuki, M., Maddux, W. W., & Masuda, T. (2007). Are the windows to the soul the same in the East and West? Cultural differences in using the eyes and mouth as cues to recognize emotions in Japan and the United States. *Journal of Experimental Social Psychology*, 43, 303–311.

17 Interpersonal accuracy

Real and perceived links to prosocial behavior

Sara D. Hodges and Adrienne A. P. Wise

Abstract
This chapter examines the relationship between interpersonal accuracy and prosociality. These two concepts seem intuitively intertwined, perhaps in part because both of them have been linked to the very broad and diffuse idea of "empathy." We first highlight studies that support the relationship between interpersonal accuracy and prosociality. Notably, the predominant measure of interpersonal accuracy in studies supporting the link is emotion recognition – and in particular, recognition of fear. We next explore why the connection between these two constructs may be over-perceived. One possible reason is that both interpersonal accuracy and prosociality can be viewed as multidimensional constructs that have been operationalized in multiple ways. Links found between certain pairs of measures may not generalize to all measures. We conclude with future directions that may help further delineate the mechanisms that connect these two associated, yet distinct, constructs.

Everyone knows that people who are more interpersonally accurate also behave more prosocially – right? Think about that kind guidance counselor in high school who somehow sensed that you didn't want to study engineering in college and was so helpful in finding you art schools to apply to. Or the hospice nurse who recognized that Great Aunt Mary wanted to die at home and helped the family to make her comfortable during her last days. Or that nice man on the plane who noticed that two other passengers were falling in love and gave up his seat so they could sit together.

We can all come up with examples of beloved and vaunted people in our lives who have this great combination: they have an uncanny ability to know and understand others – that is, interpersonal accuracy – *and* they act in prosocial ways on those others' behalf. However, just because two components together make a great combo is not evidence that the two

Acknowledgment: We thank Colton B. Christian for his helpful comments on this chapter.

components have to co-occur: Chocolate and mint are a beloved and vaunted combination too, but they occur in isolation of each other at least as often as they do together (think about chocolate milk, chocolate chip cookies, and Nutella – or toothpaste, chewing gum, and mojitos). However, our own personal experience as researchers who study interpersonal accuracy suggests that in a lot of people's heads, interpersonal accuracy and prosociality are assumed to go together. We see this perceived correlation in laypeople, who, upon hearing about our accuracy research, send us newspaper articles about heroic helpers ("sounds like your research!") and even assume that we *personally* are prosocial – because we study interpersonal accuracy. We see it in other researchers, in their papers that we review (and some that make it into press), where scales developed to measure interpersonal accuracy are used as evidence of kindness and compassion toward others.

Like chocolate and mint, interpersonal accuracy and prosociality complement each other as a lovely combination, but in addition to doing duets, they may also perform solo. They may even sometimes work in inverse (e.g., when interpersonal accuracy is used to "personalize" harm toward others – see Hodges & Biswas-Diener, 2007; Hodges & Myers, 2007). We will demonstrate in this chapter that interpersonal accuracy can be a predictor of prosociality – but reliably under only some conditions. We will also present a case for why the co-occurrence of these two constructs may be overestimated.

Defining terms – interpersonal sensitivity

Hall (2011a) distinguishes between interpersonal sensitivity that reflects someone accurately noticing another person's state or trait, and interpersonal sensitivity that reflects a "wise and tactful response to what one has perceived" (p. 319). Hall makes the case that the two sorts of sensitivities should be considered separately. For the purposes of this chapter (and of this entire volume), we will stick largely with the first sort of interpersonal sensitivity that involves accurate perception. It is easier to delineate generally because it requires a criterion for accuracy and it is also easier than the second sort of sensitivity to distinguish from prosocial behavior. For example, if you compliment someone at your gym on her flat stomach while you observe her doing sit-ups, it could well mean that you have made a "wise and tactful response" to what you perceived. However, the compliment could also be seen as constituting prosocial behavior.

The accuracy side of interpersonal sensitivity has been measured in a variety of ways. One common method is to measure accuracy at identifying facial emotional expressions (a method often found in studies of prosocial behavior). Such measures often use a standard set of stimuli, such as the Diagnostic Analysis of Nonverbal Accuracy 2 – Adult Faces (DANVA2-AF; Nowicki & Duke, 1994) or one of the versions of the Reading the Mind in Eyes test (e.g., Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001; although this tests the ability to read emotion in the eyes only, not complete faces). Another variation is to show emotional faces to perceivers and ask them how *much* of a particular emotion the target person feels (e.g., Haugen, Welsh, & McNulty, 2008; Papp, Kouros, & Cummings, 2010).

Other interpersonal accuracy measures have been studied much less in conjunction with prosociality. One method that is dynamic over the course of a social experience was introduced by Levenson and Ruef (1992) and has been used by Zaki and colleagues (e.g., Zaki, Bolger, & Ochsner, 2008). A target person is filmed, then watches the film and provides ratings over time of how positive or negative she feels, using a dial to make continuous ratings. These ratings serve as the criterion for accuracy. Perceivers observe the target over the same period and also adjust the dial rating over time depending on how positively or negatively they think the target feels.

A more open-ended (but also dynamic) measure of accuracy is to record targets, then ask them to watch the recording and retrospectively report the time points whenever they had a thought or feeling, along with the content of that thought or feeling. In this method, developed by Ickes and colleagues (e.g., Ickes, Stinson, Bissonnette, & Garcia, 1990), perceivers are then asked to guess what the target was thinking or feeling at the same time points. Both of these measures have been called empathic accuracy (see Chapter 3).

Finally, there is variety of other methods that measure perceivers' accuracy at assessing more enduring characteristics of a target, such as the target's personality (e.g., Funder, 1995), appearance (e.g., Horgan, Schmid Mast, Hall, & Carter, 2004) or more specific characteristics such as sexual orientation or teaching ability (Ambady, Bernieri, & Richeson, 2000) (see Chapters 5, 6, and 15).

Although all these different measures of accuracy seem conceptually related, there is limited empirical evidence that they represent a unified factor (see Chapter 18). Researchers tend to gravitate toward using one method (or maybe two similar ones) in their studies. Hall and Bernieri's "state of the science" book on *Interpersonal Sensitivity* (published in 2001) provides about two dozen separate chapters that each focus on a different methodology (including separate chapters for two different measures of the ability to identify emotional expressions). Toward the end of the book, Zebrowitz (2001) warns in her chapter that a true understanding of interpersonal sensitivity will be impeded as long as researchers “continue to probe its separate components” (p. 334), but the decade and a half since has produced little to show in terms of heeding her words. Progress may be impeded not just by the practice of “specializing” in only one measure of accuracy, but also by the practice of either explicitly or implicitly suggesting that one’s findings would generalize if one were to use other measures.

Defining terms – prosociality

Defining prosociality presents problems too, as the term can cover a wide gamut of human behavior, and the delineations around it have fuzzy edges. In terms of this chapter, acts that help others (such as providing needed assistance or resources) clearly constitute prosocial behavior. Kind behaviors (compassionate reactions, compliments, other expressions of positive affect) may also be prosocial, but these behaviors may also be performed for entirely self-serving reasons. Prosocial *traits*, such as empathy (as often defined), agreeableness, and warmth, have been studied more in connection with interpersonal accuracy than prosocial *behaviors*. Prosocial trait measures are often self-reported, and they, like prosocial *intentions* (which are also generally self-reported), may be susceptible to inaccuracies and inflation due to simple social desirability response biases (people want to appear prosocial). If everyone fudged reports of their own prosociality upward the same amount, correlations with other constructs such as interpersonal accuracy would be unaffected. However, inaccurate inflation may not be the same at all levels or for all measures: Marsh, Kozak, and Ambady (2007) caution against the use of self-report empathy scales as a measure of prosocial responding, observing “that people who are particularly nonempathic are unable to recognize their lack of empathy” (p. 247). There may also be gaps between self-reports and behavior caused by the fact that acting on prosocial intentions often comes at a personal cost (e.g., see Batson & Thompson, 2001).

An additional complicating factor echoes what we saw for interpersonal sensitivity: Is prosociality a single dimension or itself a multidimensional construct? (see Dunfield & Kuhlmeier, 2013 for a review). Expressing a theme that may sound familiar (and that may become downright repetitive in our discussion of empathy later in this chapter), researchers who study prosociality tend to collect just one measure of prosocial behavior that then is presented as functioning as a proxy for prosociality more generally. However, at least among younger humans, when Dunfield and

Kuhlmeier (2013) measured different forms of prosociality in the same research participants, they found that toddlers' tendencies to engage in different forms of prosociality were uncorrelated. Of course, different forms of prosocial behavior may be more interconnected in adults (Hubbard, Srivastava, Degras, Harbaugh, & Mayr, 2015), perhaps as abstract concepts such as morality and social identity develop. In this chapter, we will be biased toward focusing on prosocial behavior rather than traits when we can, because some of the traits associated with prosociality are much broader and encompass other constructs than mere prosociality (e.g., agreeableness in the Big Five model of personality).

A final complicating factor is that, as noted above, "empathy" may be considered a prosocial trait, but "empathy" may also be considered its own fuzzy, broad construct. (To make matters worse, the term "empathy" also frequently refers to a construct that encompasses interpersonal accuracy as well, a problem we will cover later in this chapter.) Our intuitions are that in common usage, "prosociality" more emphasizes being kind and doing kind things for others, whereas "empathy" emphasizes feeling for them and understanding them. However, the distinction between the two is very loose, and there is a great deal of overlap in the studies reviewed in this chapter that purport to be measuring one or the other. Rather than adding another set of definitions to the fray, we will generally classify studies as measuring what the researchers who conducted them report they are measuring.

Evidence for the link between interpersonal accuracy and prosociality

As we foreshadowed in our introduction, there is a small collection of studies supporting an association between accuracy and prosociality. Many investigations use the identification of facial emotional expressions as the measure of interpersonal accuracy, although, interestingly, many of the studies supporting the link specifically find results with accuracy at identifying *fear* expressions, a detail we will contemplate later in this chapter. Below, we highlight these and other studies supporting the link – not with an exhaustive review, but instead by featuring a nonrandom sample designed to illuminate, as well as probe, the link.

Traits

Looking first at studies of prosocial traits, Hall, Andrzejewski, and Yopchick's (2009) meta-analysis turned up a number of correlates with interpersonal sensitivity. Although many of these correlates were measured using self-report scales, some of these scales were probably not particularly transparent (i.e., obvious to respondents what traits were being measured) and were part of larger instruments designed to provide a profile based on a number of dimensions, rather than being devoted to exclusively measuring a single obviously desirable trait (such as affiliation). Among the traits Hall et al. examined, their category of "warmth prosociality" correlates was not significantly related to interpersonal sensitivity, but empathy, affiliation, conscientiousness, and tolerance all were. In addition, interpersonal sensitivity was positively related to the Communality and Socialization scales of the California Personality Inventory. Interpersonal sensitivity was inversely related to a couple of negative traits that might be proxies for the absence or opposite of prosociality, specifically neuroticism and self-rated miscellaneous negative traits, but notably not aggression.

A review of research on emotional intelligence (EI) by Mayer, Salovey, and Caruso (2004) reported that the "perceiving emotions" component of EI (the component most related to interpersonal accuracy) had a small but significantly positive correlation with agreeableness (as did Hall et al. 2009's meta-analysis). It also had very small but significant correlations with neuroticism (negatively correlated), extraversion, and openness (both positively correlated), all of which could be considered markers of prosociality if one were to use the most inclusive funnel to gather adjectives related to a sort of "über" positive trait. EI more generally – that is, including all components and not just accurately perceiving emotions – was also found to negatively predict *anti*social activities such as bullying, destructive behavior, and deviance. In a separate study of EI in the workplace using a small sample of employees, Lopes, Grewal, Kadis, Gall, and Salovey (2006) found that those who scored higher on the EI subscale of perceiving emotions (as well as EI generally) were rated more positively by their coworkers on dimensions that could potentially reflect prosociality (e.g., creating a positive work environment; see also Hall et al., 2009, for a review of the relationship between interpersonal accuracy and other workplace outcomes that could imply prosociality).

Besel and Yuille (2010) found that higher scores on Davis's Empathic Concern subscale of the Interpersonal Reactivity Index (IRI; a self-report empathy scale; Davis, 1983) predicted participants' ability to correctly identify facial emotions with brief exposure (approximately 50 ms) to the faces. In contrast, higher scores on the Empathy Quotient (EQ; Lawrence, Shaw, Baker, Baron-Cohen, & David, 2004) predicted accuracy at longer exposures (approximately 2000 ms). (The EQ and IRI empathic concern scales themselves were correlated .44.) However, as mentioned earlier, there appears to be a unique importance of the ability to specifically recognize *fear* expressions in predicting prosociality. When Besel and Yuille (2010) examined the EQ's relationship to accuracy at identifying specific emotions, fear was the emotion whose identification was most related to the EQ (and, in the interest of full reporting, this relationship was found only when faces were viewed at the slower presentation time of 2000 ms). We will consider explanations and speculations about why recognition of fear may be related to prosocial traits later in this chapter.

Prosocial behavior

Turning now to prosocial *behaviors*, some evidence for a connection between fear identification and prosociality is again apparent. Marsh et al. (2007) found convincing evidence for the link between fear recognition and prosocial behavior in a series of three studies. In the first study, participants heard the case of a young woman in need (based on the "Katie Banks" story used by Batson and colleagues in much of their work – e.g., Batson, Early, & Salvarani, 1997) and were then asked to donate on her behalf. Participants were also asked to identify facial expressions using the DANVA. Participants who were more accurate at identifying fear expressions donated more. (The researchers also, interestingly, found that people who were better at identifying happy expressions gave *less* – something the researchers speculate may be related to mood congruency: Participants who were not moved by Katie's sad story may have been less sad – and thus happier and perhaps better able to identify happy faces.)

In their second study, Marsh et al. (2007) used a different, novel measure of prosocial responding. Participants were asked to make attractiveness ratings of target people, and were either told that those people would receive the feedback (prosocial condition, where making higher ratings was the more prosocial response), or that the ratings were solely for the purpose of norming stimuli (control condition). As expected, participants who could more accurately identify fearful facial expressions (again using a measure that asked participants to identify facial expressions) also gave higher attractiveness ratings to the target people, but only in the prosocial condition. Ability to identify sad expressions correctly also predicted higher attractiveness ratings in the prosocial condition, which is consistent with the idea that ability to identify others' distress expressions predicts prosocial responding. A final study quite similar in methodology to the second study put all participants in the prosocial condition – that is, led them to believe that targets would receive the feedback, making more attractive ratings more prosocial, or at least more kind. In this third study, ability to identify fear expressions was the only variable that predicted higher ratings of target attractiveness from participants.

In a later study, Marsh et al. (2014) compared fear recognition abilities in a group of people who had committed the extremely prosocial act of donating a kidney to a nonrelative to fear recognition in a control group. There was a group-by-emotion-recognition interaction effect: Donor group members did relatively better than controls in identifying fear expressions, whereas the controls did relatively better at identifying anger expressions. (Simple effects tests were not significant. The entire sample size for the study was 39, including 19 kidney donors.) The donor group also appeared to show greater reactivity to the fear expressions – they exhibited greater brain response in the right half of their amygdala (a part of the brain connected with processing fear stimuli) and activity in this region was correlated with accuracy at recognizing fearful expressions.

A study by Johnson (2012) provides a few other tidbits hinting at (and slightly complicating) the special relationship between fear recognition and prosocial behavior in a study whose main intent was to examine the correlates of "literary transportation" (i.e., the state of being fully engaged and emotionally caught up in a story). As part of the study, participants first read an evocative story designed to evoke compassion. Johnson also not only looked at accuracy at identifying happy, fearful, and neutral facial expressions but also looked at perceptual bias to over-perceive these emotions. Specifically, he examined whether participants were biased to perceive faces as expressing one of the two emotions (happiness and fear) when the faces were actually neutral. He found[1] that perceptual fear bias – a tendency to see facial stimuli as expressing fear even when they are not – predicted that participants were more likely to help an experimenter who "accidentally" dropped a handful of pens. (Interestingly, the perceptual bias to view faces as expressing fear was present for faces presented for 2000 ms, but not for faces presented for only 50 ms, which Johnson interpreted as indicating that the bias requires "some deliberation" – p. 154.) Notably, prosocial behavior was not associated with a bias to over-perceive emotions generally: Perceptual bias in perceiving *happy* faces did not predict helping, so it does not appear that merely having a low threshold for perceiving emotion is associated with prosocial behavior. Furthermore, the perceptual bias in

[1] In Study 2 of the paper, Johnson ran a minimum of 11 multiple regression equations and did not statistically adjust the significance level for testing multiple hypotheses; thus, the results should be treated with some caution.

perceiving fear was itself related to Davis's empathic concern trait subscale of the IRI and Batson's (e.g., Batson et al., 1997) empathic concern adjectives (designed to tap sympathy for specific targets). Thus, participants who reported more concern for others generally, and more concern for the characters in the story that they had just read, tended to overperceive fear in a separate set of faces. In contrast, – and somewhat at odds with Besel and Yuille's (2010) results discussed earlier, actual *accuracy* at identifying fear expressions (not bias in seeing fear expressions) was *negatively* predicted by Davis's IRI empathic concern subscale. That is, participants who scored higher in self-reported trait empathic concern in Johnson's Study 2 tended to be worse at identifying which faces were fearful and which were not.

What's so special about fear?

Why is recognizing fear a special predictor of prosociality? Answers to this question are still somewhat speculative, but one theme that comes up repeatedly in explanation attempts is that the *inability* to recognize fear is a robust marker of constructs that might be considered the opposite, or at least the absence, of prosociality – specifically psychopathy (Marsh & Blair, 2008; Marsh et al., 2011).[2] One explanation of this relationship hinges on "the notion that aversive cues [such as fearful faces] do not appropriately modulate the behavior of individuals whose behavior is marked by antisociality and a lack of empathy" (Marsh et al., 2007, p. 247). The expression of fear can be seen as a signal of submission (Marsh et al., 2007) and its "wide eyes and high brows" in humans (p. 15039, Marsh et al., 2014) may even be seen as resembling features of a baby's face, priming behaviors aimed at those who are vulnerable and need care (see Marsh, Adams, & Kleck, 2005; Marsh & Ambady, 2007). However, individuals high in psychopathy seem immune to and unmoved by these cues – whereas people who are particularly sensitive at recognizing them tend to demonstrate higher levels of prosociality.

In addition to the special link between processing fear expressions and prosociality, more global deficits in decoding emotional expressions may also predict the absence of or reduction in prosocial behavior. Knafo, Steinberg, and Goldner (2011) measured the link between interpersonal accuracy and prosociality in very young children (3–6-year olds), with a somewhat different measure of accuracy than

[2] Although deficits in fear recognition are most pronounced, other specific deficits, such as in identifying disgust (see Sato, Uono, Matsuura, & Toichi, 2009), responding to pain (Cheng, Hung, & Decety, 2012), and general deficits (e.g., Dawel, O'Kearney, McKone, & Palermo, 2012) have also been found among those scoring high in psychopathy.

others discussed so far. To measure accuracy, Knafo et al. (2011) used an emotion identification task that asked children to read stories and then verbally state the emotion the protagonist felt and identify which of three faces best represented how the protagonist felt. Then, to measure prosociality, the experimenter feigned three behaviors that were designed to elicit "self-initiated" prosocial behaviors from the child (e.g., the experimenter feigned an injury in order to see if the child would help her up or comfort her). In addition, the experimenter directly asked the child for help in three other situations (e.g., looking for a misplaced item). Children who did poorly on the emotion identification task (i.e., scored in the bottom 20% for their age) also scored low specifically on the self-initiated prosocial behaviors.

There is an intriguing postscript to the results of the Knafo et al. (2011) study, given that it follows on the heels of our earlier discussion of the unique link between fear identification and prosocial behavior. Knafo et al. (2011) intended for their emotion identification measure to be made up of five items, each corresponding to a different emotion (disgust, happiness, sadness, anger, and fear); however, when examining the inter-item reliability of this measure, the fear item – and only the fear item – did not load on the same factor as the others and was discarded. Without knowing additional details about the results, it is hard to know what to make of this twist – and indeed, we don't know whether fear was correlated with prosociality in their study.

The positive correlation between prosocial behavior and the ability to correctly recognize fear expressions (or indeed, in Johnson's 2012 paper, to be *biased* to see fear, such that neutral facial expressions are perceived as fear) is both fascinating and frustrating – the latter because it raises a lot of unanswered questions. For example, what is the causal direction of this relationship? Maybe a lower degree of prosocial orientation leads people also to not care about or pay attention to fear expressions. Or alternatively, maybe a greater tendency to recognize fear expressions predicts greater prosociality, because people with better fear recognition know when help or compassion is needed. And, of course, both directions of causality might be operating.

Contemplating an answer to the question about causal direction leads to speculations about another question: Is the relationship between prosociality and ability to recognize fear linear? If prosociality's link to fear recognition is mediated by attention, it would be reasonable to expect that incremental increases in prosociality might lead to incremental improvements in fear recognition. However, if deficits in fear recognition deprive people of the key cues that trigger prosocial behavior, then we might imagine more of an "all or nothing" relationship driven by categorical variables: People who recognize fear act prosocially; people who don't recognize fear do not act prosocially.

Looking in the right places to find the link between accuracy and prosociality

The previous section hints at the fact that links between interpersonal accuracy and prosociality may be easier to find when looking at the lower end of both distributions. For example, in the Knafo et al. (2011) study described above, the researchers reported that they "did not find a significant linear correlation between [their measure of emotion identification] and prosocial behavior" (p. 197), raising the possibility that among children (and perhaps adolescents and adults), the most important information in predicting prosociality may be whether someone scores above or below some minimal level of accuracy. It may be a mistake to assume that the same neat linear relationship between accuracy and prosocial behavior occurs along the entire range of values of the variables used to measure these two constructs. In line with our discussion of "all-or-nothing" effects with fear recognition, we can speculate that finding a positive correlation between accuracy and prosociality might really be better described as higher mean prosociality in the group that scores above some threshold for accuracy than in the group that is below that threshold, and even studies that have found linear relationships between accuracy and prosociality might also have found curvilinear effects, if they had looked for them.

Dunfield and Kulhmeier (2013) astutely observed that for people to make a decision to act prosocially or not, they must first identify that there is a need for prosocial behavior and what the prosocial behavior would be. People missing these first two skills are severely restricted in terms of their likelihood of behaving prosocially. Someone who cannot distinguish between clear facial expressions of basic emotions (e.g., between a broad smile and a deep grimace) likely will also be unable to establish what would be helpful or responsive to the person displaying these expressions, as these two expressions would presumably dictate very different courses of action.

To further complicate matters, various psychological and mental disorders are associated with deficits in various forms of interpersonal accuracy – including autism (e.g., Clark, Winkielman, & McIntosh, 2008; Demurie, DeCorel, & Roeyers, 2011); schizophrenia (Harvey, Zaki, Lee, Ochsner, & Green, 2013) and some forms of dementia (e.g., Fernandez-Duque & Black, 2005; Fernandez-Duque, Hodges, Baird, & Black, 2010; see also Chapter 9). People with these disorders also may frequently display behaviors that reflect a distinct absence of prosociality (e.g., inappropriate emotional outbursts, uncooperative behavior, and lack of compassion), but the blunt effects of very serious mental disorders can make it hard to pinpoint precisely which deficits are producing which outcomes. For example, people with severe autism often lack language skills (Tager-Flusberg, 2005), which greatly constrains their ability to decode interpersonal communications accurately *or* to signal prosocial intentions. In summary, in examining the relationship between interpersonal accuracy and prosociality, researchers should at least consider the possibility that statistically significant linear relationships may be driven by mathematical relationships that would be better described using non-linear models.

Moderating factors

Finally, one last line of research contributes important information for delineating the link between interpersonal accuracy and prosociality. Côté et al. (2011), noting inconsistent results in past studies of accuracy and prosociality, hypothesized that the link was moderated by power. Due to the heightened goal focus that accompanies power, they believed high-power people would show a greater association between accuracy and prosociality and were able to demonstrate support for the idea using a variety of operationalizations of all three constructs.

Côté et al.'s first study utilized greater respiratory sinus arrhythmia (RSA) as a physiological marker of prosociality. In brief, RSA is variation in the heart rate that is related to respiration – RSA is the degree to which the heart beats faster during inspiration of breath, and slower on expiration of breath (Berntson, Cacioppo, & Quigley, 1993). RSA is an index of parasympathetic nervous system control and is associated with emotion regulation (Musser et al., 2011). It has been linked to a prosocial orientation (see Côté et al., 2011, for a review), but to a number of other psychological constructs as well.

Côté et al. (2011) found that RSA predicted accuracy at guessing how much another person was feeling certain emotions (not just fear) among participants who self-reported a higher sense of power. In their second study, Côté et al. manipulated power and also manipulated prosociality by arousing compassion via evocative media (versus control media). In this study, the outcome measure was a variation on Ickes' (e.g., Ickes et al., 1990) empathic accuracy paradigm. Participants were asked to guess which emotions were being felt by a person whom they thought they would be interacting with later in the study. The target's own emotion ratings – the criterion for accuracy – had been collected previously after recording an interview with the target. Participants who were in the combined high-power and high-compassion condition showed the greatest accuracy at inferring the target's emotions. Finally, in a third study, the researchers took advantage of naturally occurring power differences associated with workers who held different jobs in the workplace. They also collected agreeableness scores as a measure of prosocial orientation in these workers and measured workers' ability to identify facial expressions. Once again, Côté et al. (2011) found a stronger relationship between accuracy and agreeableness among the more powerful.

Summarizing the link

Thus, to summarize, the search for studies that support the accuracy–prosociality link does not come back empty-handed, nor is the yield overwhelming. It seems fair to say that there *is* a link, but to qualify it. First, studies finding the link exclusively rely on emotion recognition as the measure of accuracy. Accuracy at identifying emotional expressions is a key form of interpersonal accuracy and a well-studied one too, but it is unclear at this point whether the link with prosociality would be found if other forms of interpersonal sensitivity were used – such as trait judgments or attentiveness to appearance. When it comes to the link between the ability to infer another person's thoughts and prosociality, what little evidence there is actually hints at no relationship[3] (something to be discussed in more detail later).

Second, the robustness of the link appears to be further moderated by use of fear identification as the measure of accuracy and may not reliably appear when accuracy at identifying other emotions is tested. Third, particularly when recognition of emotions other than fear is the measure of accuracy, it is unclear whether or not there is a smooth linear relationship between the two constructs across the full range of both distributions. Finally, Côté et al.'s (2011) work examining the moderating role of power on the link between accuracy and prosociality identifies an important moderator of the link, and it also provides a more general nudge that researchers should consider and test the possibility of other moderators of the link. We think that power and other possible moderating variables may be overlooked and the strength of the perceived link between accuracy and prosocial behavior can be overestimated, due to a confirmatory bias to overgeneralize a relationship that is seen just under specific circumstances. We turn now to other reasons the link may be overestimated.

[3] Côté et al.'s work describes an "empathic accuracy" paradigm that is related to Ickes' thought-feeling inference paradigm, but perceivers are provided with a list of emotions to rate the target on, rather than generating their own inferences.

Why might the link between accuracy and prosociality be overestimated?

Part of the challenge in finding robust support for the idea that interpersonal accuracy and prosocial behavior are linked is that only a limited number of studies have measured both constructs simultaneously. This may be a function of these two topics emerging from two separate research traditions historically, with the study of interpersonal accuracy emerging from the personality psychology tradition (e.g., Gage & Cronbach, 1955) and the study of prosocial behavior growing out of social psychology and sociology. However, perhaps more insidious than the dearth of studies including both constructs is the fact that when these constructs are studied separately, many of the results are labeled studies of "empathy." Associating the term empathy with both interpersonal accuracy and prosociality may have contributed to over-perceiving the extent to which these two constructs are correlated.

Many papers on empathy open with a sentence along the lines of, "There are multiple definitions of empathy." There is now even some consensus that the multiple definitions are not necessarily due to theoretical disagreements about empathy, but instead reflect the fact that empathy is a multidimensional construct (Davis, 1983; Hodges & Myers, 2007; Zaki & Ochsner, 2012), with several of those dimensions being relevant to the relationship between interpersonal accuracy and prosocial behavior. Specifically, studies of what has been called *cognitive empathy* (e.g., the ability to take others' perspective and solve theory of mind problems) often reflect concerns with accuracy, even if they sometimes rely on self-reported attention to accuracy, rather than making comparisons with any actual criterion of accuracy. *Empathic concern* is an emotional dimension of empathy that is defined as tender heartedness and feelings of compassion and is by itself considered a prosocial response. It has also been persuasively linked to prosocial behaviors such as volunteering to help someone in need in an extensive corpus of work by Batson and colleagues (e.g., Batson et al., 1997). In contrast, *personal distress* is another emotional dimension of empathy that is more self-than other-oriented, but it reflects a perceiver's own distress at encountering someone in need. It may impede prosocial behavior by motivating people to try to escape a distressing situation rather than sticking around to help (Batson et al., 1997; Lamm, Batson, & Decety, 2007).

Empathic accuracy is not the same thing as empathy

To complicate matters further, some measures of interpersonal accuracy go by the name of "empathic" accuracy (including Ickes', and Levenson and Ruef's, paradigms) and seeing the "empathy" right there in their name leads some people to assume that these measures must somehow reflect prosociality. However, unlike emotion recognition, which has been linked to prosociality under some circumstances (ones we have described earlier in this chapter), there is scant to no evidence that performance on empathic accuracy tasks predicts prosociality. In fact, there is even a little concrete evidence that these two constructs are *not* related: Klein and Hodges (2001) found that reported sympathy (a.k.a., empathic concern) for a target person was not significantly correlated with accuracy at inferring that target person's thoughts (i.e., using Ickes' measure of empathic accuracy). At least in Ickes' paradigm, accuracy depends very much on what the target says (e.g. Gesn & Ickes, 1999; Lewis, Hodges, Laurent, Srivastava, & Biancarosa, 2012). It is possible that perceiving and focusing on emotional displays (e.g., facial expressions) are key to finding a link between accuracy and prosociality.

Two studies that were not designed to measure the relationship between accuracy and prosociality nonetheless do measure empathic accuracy along with other empathic constructs and their results provide indirect evidence about the absence of a relationship between interpersonal accuracy and prosocial responding. Hodges, Kiel, Kramer, Veach, and Villanueva (2010) collected a variety of measures that have fallen under empathy's giant umbrella, including an accuracy measure (empathic accuracy using Ickes' paradigm), and two measures of empathic responding that are more related to prosociality: participants' self-reported empathic concern for a target, and participants' self-reported understanding of the target. Hodges et al.'s main goal for the study was to examine how perceivers' personal experience affected scores on various dimensions of empathy expressed toward targets, who were all first-time mothers discussing their adjustment to the birth of their child. Participants who themselves had experience relevant to becoming a new mother self-reported greater empathic concern and perceived understanding of the new mother targets (i.e., they scored higher on both constructs related to prosociality). In contrast, on accuracy measures, participants with relevant personal experience were more accurate at guessing how other new mothers would respond to a questionnaire about adjusting to new motherhood, but this advantage was largely due to their ability to guess how the stereotypic new mother would answer. When guessing the idiosyncratic thoughts of new mothers using the Ickes' empathic accuracy paradigm, experience had no effect. Thus, personal experience predicted empathy constructs related to self-reported prosociality, but was less successful at predicting empathy constructs related to accuracy. Hodges et al.'s (2010) study highlights the importance of specifying constructs precisely, rather than stating that "Variable X" increases empathy – which may be taken to mean X increases accuracy, prosociality, or both.

In a study comparing elderly controls to patients with frontotemporal dementia (FTD) and patients with Alzheimer's disease (AD), Fernandez-Duque et al. (2010) again demonstrated that a variable – in this case, being diagnosed with different forms of dementia or not – did not show the same pattern with prosociality as it did with accuracy. When it came to accurately guessing the thoughts of an ambivalent target, both kinds of patients showed accuracy deficits relative to elderly control participants. (These accuracy measures were derived from Ickes' empathic accuracy paradigm but were substantially simplified to accommodate the patient samples.) However, when it came to a prosociality measure – collected only for patients (and not for the elderly controls) by asking caregivers to evaluate the patients using items on Davis's IRI – the Alzheimer's patients scored higher on perspective-taking tendencies and empathic concern than the FTD patients. Thus, both disorders seemed equally damaging to interpersonal accuracy, while one (FTD) had a worse impact on prosociality.

Finally, correlations between self-report empathy scales and Ickes' measure of empathic accuracy are not consistently found, which is perhaps not surprising given that few perceiver characteristics have been found to reliably predict empathic accuracy as measured with that paradigm (see Hodges, Lewis, & Ickes, 2015), whereas contextual variables often do predict accuracy. We should probably also not be that surprised that although reading emotional expressions shows some relationship to prosociality, reading thoughts does not. Emotional expressions – at least those accompanying the basic emotions – have been hypothesized to have evolved specifically for signaling purposes (Knutson, 1996). These expressions are distinctive and few in number and the ability to read them appears to be hardwired (Matsumoto & Willingham, 2009) – and thus they were ideal signals for triggering adaptive responses, like helping someone in need (particularly adaptive if the person in need were kin). However, the vast majority of possible thoughts entertained in someone's head are not accompanied by shared signals and are in fact largely private. Instead of "reading" thoughts like visible facial expressions, accuracy at inferring them requires construction and integration of general schemas (Gesn & Ickes, 1999; Hodges et al., 2010; Lewis et al., 2012). Thus, for example, when inferring the thoughts of a new mother, people rely on what they know about the experiences of new mothers in general. Accuracy is also helped by specific background knowledge of the target person (Stinson & Ickes, 1992).

In summary, a variety of studies suggest that variables that can be used to predict empathy measures with a prosocial flavor cannot be used in a parallel fashion to predict empathic accuracy (specifically empathic accuracy measured with the Ickes paradigm and close variants). Furthermore, empathic accuracy (measured with the Ickes paradigm) itself is not consistently correlated with prosocially flavored empathy measures (e.g., self-reports of empathic concern). When hearing that "X predicts empathy," an assumption that this means X affects both empathic accuracy and empathic concern may not only inflate perceptions that accuracy and prosociality are correlated, but would also more often than not be wrong.

All kinds of empathy on the brain

At the neural level, Zaki and Ochsner (2012; see also Zaki, Weber, & Ochsner, 2012) identified two distinct and nonoverlapping neural systems driving two kinds of empathy. They further explored how the two systems intersect with the neural underpinnings of prosociality. The first of the two empathy systems, experience sharing, is defined as "vicariously sharing targets' internal states" (p. 675), including their emotions. Thus, experience sharing may involve feeling sad when another person feels sad. Experience sharing therefore could lead to accuracy via the indirect path of a perceiver correctly perceiving a target's experience and then resonating to it, but such resonance does not necessarily imply a conscious appraisal of the other person's state. However, feeling sad for a person who has gotten drunk and who is cheerful now but will regret it in the morning would *not* fit with Zaki and Ochsner's experience sharing, because it lacks the resonance component of feeling what the other person is feeling. The other empathy system in Zaki and Ochsner's model is mentalizing, or "explicitly considering (and perhaps understanding) targets' states" (p. 675). It would seem that achieving accuracy thus likely draws on mentalizing, but the act of mentalizing by no means guarantees accuracy in perceiving others (see Bombari, Schmid Mast, Brosch, & Sander, 2013, for a discussion of the relationship of these two systems to interpersonal accuracy).

Zaki and Ochsner see both empathy systems as potentially leading to prosociality in the form of concern for a target person – because a perceiver resonates to a target's pain and/or because a perceiver draws mental conclusions about a target's sorry state. Thus, in their model, both empathic pathways may lead to prosociality. Furthermore, Zaki, Weber, Bolger, and Ochsner (2009) found evidence that brain regions in both pathways have greater activation in research participants who are more accurate at identifying the valence and extremity of a target person's emotions. Thus, in their model, both empathic pathways may also be linked to interpersonal accuracy. However, whether interpersonal accuracy and prosociality are directly linked at the neural level remains to be seen. As an analogy, reading more books and writing more letters may improve one's vocabulary, and reading more books and writing more may make one a more interesting conversation companion, but there is not necessarily a direct connection between one's vocabulary and one's conversation skills.

Other speculations

We speculate (without empirical evidence) that there are other possible – and to be blunt, fairly cynical – reasons why the real correlation between interpersonal accuracy and prosocial behavior may be smaller than people's intuitions about it. First, although one's interpersonal accuracy could help one to recognize that prosocial behavior is needed, that recognition may also be accompanied by the knowledge that the prosocial behavior that is called for is costly – in terms of giving up or sharing resources or one's time, or in terms of feeling negative emotions such as distress or frustration. If we think about graphing the linear relationship between accuracy and prosociality, then individuals who are interpersonally sensitive and who also intend to – but do not actually – behave prosocially may be spuriously populating the high accuracy/high prosociality quadrant. There, they would thus be providing support for a "real" positive correlation between interpersonal accuracy and prosocial *intentions* (which, granted, would be an interesting connection – perhaps indicative of a person who cares a lot about what other people think of him or her), but they might also artificially inflate the perceived link between interpersonal accuracy and *actual* prosocial behavior.

There may be other perceiver characteristics that are inflating the "high/high" quadrant. For example, being interpersonally sensitive is not only perceived to be part of the female gender role (Hodges, Laurent, & Lewis, 2011) but on many measures of interpersonal accuracy, women outscore men (Hall & Schmid Mast, 2008; see also Chapter 15). Similarly, caring and compassion are big components of the female gender role (Helgeson, 1994; prosociality for men may take on a more agentic and less communal flavor – Eagly, 2009) and women are overrepresented in certain service and nurturing roles, such as nursing and teaching (National Science Foundation, 2013). We are not claiming that the positive correlations between accuracy and prosociality are all due to gender – many of the studies that have found this relationship did so while controlling for gender (e.g., Besel & Yuille, 2010; Marsh et al., 2007, Study 3). However, the fact that both constructs might be associated with stereotypes of women may be illusorily boosting people's perceived correlations. Similarly, there may also be a halo effect for people we like or perceive as nice – as "good" people, they are overperceived as having all positive qualities: attractiveness, intelligence, strength – as well as interpersonal sensitivity and prosociality.

Second, people may misrepresent their understanding of others as an excuse for not behaving prosocially. A person who recognizes a homeless man's pleading expression and dull eyes as signs of hunger but claims not to can get away with not sharing her sandwich. However, if other people believe that she really did not know the man was hungry, then her "falsified data" inflates the correlation between accuracy and prosociality, by adding another data point in the low accuracy/low prosociality quadrant. Third, showy and salient displays of prosocial behavior – particularly ones that appear to be costly to the person performing them – may obscure or drown out the fact that they are not in fact driven by accurate perceptions of what the target person needs. A man who gives his date his jacket to wear after halftime at a football game may get credit for both perceiving her discomfort and giving up resources, even though if he had truly been interpersonally accurate, he would have known that she was uncomfortable because of the bawdy behavior of other fans in their section. However, we might *fail* to add his data point to the low accuracy/high prosociality quadrant. Similarly, imagine someone who used her ability to accurately read people for evil purposes – for example, choosing to do the thing that she knows will hurt a person most (what we have called "Machiavellian cognitive empathy" – Hodges & Myers, 2007). Very unkind interpersonal behaviors may distract us from noticing that these behaviors actually required a high level of accuracy, thus resulting in underrepresentation in the high accuracy/low-prosociality quadrant.

Future directions and concluding thoughts

We think there are a number of interesting directions for further exploring and delineating the relationship between accuracy and prosocial behavior. So far, most approaches have at least implicitly pursued the relationship between the two constructs by looking to find them both in the *perceiver* – a sensitive person who is also prosocial. However, another approach that might prove fruitful would be to look at characteristics of targets – is there a kind of person who is both easier to read *and* more likely to evoke prosocial responses? For example, there is a sprinkling of studies using different measures of accuracy that suggest that we are better at reading in-group targets (Elfenbein & Ambady, 2003; Young & Hugenberg, 2010) and targets we are closer to (Stinson & Ickes, 1992), plus a whole literature on how we treat in-group members better, often in ways that reflect prosocial response (Brewer, 2007).

Other target characteristics may similarly function as "third variables" affecting both accuracy and prosociality. For example, targets who have fate control over us (that is, high-power people) may merit both a closer read (see Chapter 13) *and* attempts to behave more prosocially (keeping in mind that prosocial responses may be self-serving as well). Thus, we not only want to know what our boss is considering doing in the big company reorganization, but we also want to ingratiate ourselves by volunteering to help implement those plans. However, to some extent this speculation suggests that greater motivation can improve accuracy. This may be true for Ickes' empathic accuracy paradigm (although even then the evidence is somewhat indirect; Hodges et al., 2015; Klein & Hodges, 2001), but less so for other forms of accuracy, especially emotion identification (Hall, 2011b; Hall, Blanch, et al., 2009; see also Chapter 2). Given past research suggesting that powerful people send clearer signals (Hall, Rosip, Smith LeBeau, Horgan, & Carter, 2006; Snodgrass, Hecht, & Ploutz-Snyder, 1998), a related possibility is that powerful people are not only easier to read, but also clearer about what kind of prosocial assistance they need.

We think the concept of "motivated inaccuracy" (e.g., see Ickes, Simpson, & Oriña, 2005) may be interesting to study in connection with prosocial responding. Motivated inaccuracy is the idea that people may not accurately read others when what they would learn is threatening. It has been mostly studied using the Ickes' empathic accuracy paradigm, largely with romantic couples. For example, dating couples asked to infer how much their partner finds a rival attractive may be inaccurate in order to avoid seeing a threat to the relationship (Simpson, Ickes, & Blackstone, 1995). Motivation to be inaccurate may overlap with prosocial motivations (e.g., not wanting to embarrass the target by identifying his true feelings) as well as self-serving ones, and thus under certain circumstances, we may find that accuracy is reliably *negatively* correlated with prosociality – or at least with prosocial intentions.

In closing, the authors of this chapter like chocolate–mint as much as the next gal or guy – we are happy when we run across it, just as we are happy when we run across people who both are interpersonally sensitive and behave prosocially. That said, we hope this chapter has demonstrated that although there are times when accuracy and prosociality work in tandem, those instances are constrained. We also think there are instances when these constructs operate independently of each other, but their correlation may be overperceived, perhaps due to illusory correlations, confirmation biases, and overgeneralization. Both interpersonal sensitivity and prosociality are sometimes lumped together under the concept of "empathy," which may contribute to over-perception of their co-occurrence. Interpersonal sensitivity and prosociality can also both be parsed into multiple sub-constructs and when describing the relationship between these two larger constructs, it is important to be precise about what actual measures are involved. At the same time, adopting this level of precision may lead to new insights about the mechanisms that connect accuracy and prosocial behavior (for example, the interesting theories being developed about why accurate perception of fear expressions is specifically connected with prosocial responses) in future research.

References

Ambady, N., Bernieri, F. J., & Richeson, J. A. (2000). Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology, Vol. 32* (pp. 201–271). San Diego, CA: Academic Press.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the mind in the eyes" Test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry*, 42, 241–251.

Batson, C. D., Early, S., & Salvarani, G. (1997). Perspective taking: Imagining how another feels versus imagining how you would feel. *Personality and Social Psychology Bulletin*, 23, 751–758.

Batson, C. D., & Thompson, E. R. (2001). Why don't moral people act morally? Motivational considerations. *Current Directions in Psychological Science*, 10, 54–57.

Berntson, G. G., Cacioppo, J. T., & Quigley, K. S. (1993). Respiratory sinus arrhythmia: Autonomic origins, physiological mechanisms, and psychophysiological implications. *Psychophysiology*, 30, 183–196.

Besel, L. D., & Yuille, J. C. (2010). Individual differences in empathy: The role of facial expression recognition. *Personality and Individual Differences*, 49, 107–112.

Bombari, D., Schmid Mast, M., Brosch, T., & Sander, D. (2013). How interpersonal power affects empathic accuracy: Differential roles of mentalizing versus mirroring? *Frontiers in Human Neuroscience*, 7, Article ID 375. doi:10.3389/fnhum.2013.00375.

Brewer, M. B. (2007). The social psychology of intergroup relations: Social categorization, ingroup bias, and outgroup prejudice. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles* (2nd ed., pp. 695–715). New York: Guilford Press.

Cheng, Y., Hung, A.-Y., & Decety, J. (2012). Dissociation between affective sharing and emotion understanding in juvenile psychopaths. *Development and Psychopathology*, 24, 623–636.

Clark, T. F., Winkielman, P., & McIntosh, D. N. (2008). Autism and the extraction of emotion from briefly presented facial expressions: Stumbling at the first step of empathy. *Emotion*, 8, 803–809.

Côté, S., Kraus, M. W., Cheng, B. H., Oveis, C., van der Löwe, I., Lian, H., & Keltner, D. (2011). Social power facilitates the effect of prosocial orientation on empathic accuracy. *Journal of Personality and Social Psychology*, 101, 217–232.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multi-dimensional approach. *Journal of Personality and Social Psychology*, 44, 113–126.

Dawel, A., O'Kearney, R., McKone, E., & Palermo, R. (2012). Not just fear and sadness: Meta-analytic evidence of pervasive emotion recognition deficits for facial and vocal expressions in psychopathy. *Neuroscience and Biobehavioral Reviews*, 36, 2288–2304.

Demurie, E., De Corel, M., & Roeyers, H. (2011). Empathic accuracy in adolescents with autism spectrum disorders and adolescents with attention-deficit/hyperactivity disorder. *Research in Autism Spectrum Disorders*, 5, 126–134.

Dunfield, K. A., & Kuhlmeier, V. A. (2013). Classifying prosocial behavior: Children's responses to instrumental need, emotional distress, and material desire. *Child Development*, 84, 1766–1776.

Eagly, A. H. (2009). The his and hers of prosocial behavior: An examination of the social psychology of gender. *American Psychologist*, 64, 644–658.

Elfenbein, H. A., & Ambady, N. (2003). When familiarity breeds accuracy: Cultural exposure and facial emotion recognition. *Journal of Personality and Social Psychology*, 85, 276–290.

Fernandez-Duque, D., & Black, S. E. (2005). Impaired recognition of negative facial emotions in patients with frontotemporal dementia. *Neuropsychologia*, 43, 1673–1687.

Fernandez-Duque, D., Hodges, S. D., Baird, J. A., & Black, S. E. (2010). Empathy in frontotemporal dementia and Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology*, 32, 289–298.

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, 102, 652–670.

Gage, N. L., & Cronbach, L. (1955). Conceptual and methodological problems in interpersonal perception. *Psychological Review*, 62, 411–422.

Gesn, P. R., & Ickes, W. (1999). The development of meaning contexts for empathic accuracy: Channel and sequence effects. *Journal of Personality and Social Psychology*, 77, 746–761.

Hall, J. A. (2011a). Clinicians' accuracy in perceiving patients: Its relevance for clinical practice and a narrative review of methods and correlates. *Patient Education and Counseling*, 84, 319–324.

Hall, J. A. (2011b). Manipulated motivation and interpersonal accuracy. In J. Smith, W. Ickes, J. A. Hall, & S. D. Hodges (Eds.). *Managing interpersonal* *sensitivity: Knowing when – and when not – to understand others* (pp. 1–20). Hauppage, NY: Nova Science Publishers.

Hall, J. A., Andrzejewski, S. A., & Yopchick, J. E. (2009). Psychosocial correlates of interpersonal sensitivity: A meta-analysis. *Journal of Nonverbal Behavior*, 33, 149–180.

Hall, J. A., & Bernieri, F. J. (Eds). (2001). *Interpersonal sensitivity: Theory and measurement*. Mahwah, NJ: L. Erlbaum Associates.

Hall, J. A., Blanch, D. C., Horgan, T. G., Murphy, N. A., Rosip, J. C., & Schmid Mast, M. (2009). Motivation and interpersonal sensitivity: Does it matter how hard you try? *Motivation and Emotion*, 33, 291–302.

Hall, J. A., Rosip, J. C., Smith LeBeau, L., Horgan, T. G., & Carter, J. D. (2006). Attributing the sources of accuracy in unequal-power dyadic communication: Who is better and why? *Journal of Experimental Social Psychology*, 42, 18–27.

Hall, J. A., & Schmid Mast, M. (2008). Are women always more interpersonally sensitive than men? Impact of content domain and motivation. *Personality and Social Psychology Bulletin*, 34, 144–155.

Harvey, P. O., Zaki, J., Lee, J., Ochsner, K., & Green, M. F. (2013). Neural substrates of empathic accuracy in people with schizophrenia. *Schizophrenia Bulletin*, 39, 617–628.

Haugen, P. T., Welsh, D. P., & McNulty, J. K. (2008). Empathic accuracy and adolescent romantic relationships. *Journal of Adolescence*, 31, 709–727.

Helgeson, V. S. (1994). Relation of agency and communion to well-being: Evidence and potential explanations. *Psychological Bulletin*, 116, 412–428.

Hodges, S. D., & Biswas-Diener, R. (2007). Balancing the empathy expense account: Strategies for regulating empathic response. In T. F. D. Farrow & P. W. R. Woodruff (Eds.), *Empathy in mental illness* (pp. 389–407). Cambridge: Cambridge University Press.

Hodges, S. D., Kiel, K. J., Kramer, A. D. I., Veach, D., & Villanueva, B. R. (2010). Giving birth to empathy: The effects of similar experience on empathic accuracy, empathic concern, and perceived empathy. *Personality and Social Psychology Bulletin*, 36, 398–409.

Hodges, S. D., Laurent, S. M., & Lewis, K. L. (2011). Specially motivated, feminine, or just female: Do women have an empathic accuracy advantage? In J. L. Smith, W. Ickes, J. A. Hall, & S. D. Hodges (Eds.), *Managing interpersonal sensitivity: Knowing when – and when not – to understand others* (pp. 59–73). Hauppage, NY: Nova Science Publishers.

Hodges, S. D., Lewis, K. L., & Ickes, W. (2015). The matter of other minds: Empathic accuracy and the factors that influence it. In P. Shaver, M. Mikulincer (Eds.), J. A. Simpson, & J. Dovidio (Assoc. Eds.), *APA handbook of personality and social psychology: Vol 3. Interpersonal relations* (pp. 319–348). Washington, DC: American Psychological Association.

Hodges, S. D., & Myers, M. W. (2007). Empathy. In R. F. Baumeister and K. D. Vohs (Eds.), *Encyclopedia of social psychology* (pp. 296–298). Thousand Oaks, CA: Sage.

Horgan, T. G., Schmid Mast, M., Hall, J. A., & Carter, J. D. (2004). Gender differences in memory for the appearance of others. *Personality and Social Psychology Bulletin*, 30, 185–196.

Hubbard, J., Srivastava, S., Degras, D., Harbaugh, W., & Mayr, U. (March, 2015). *Looking for the heart in the brain: Neuroimaging analyses of life-span differences in charitable giving.* Poster presented at the annual meeting of the Cognitive Neuroscience Society, San Francisco, CA.

Ickes, W., Simpson, J. A., & Oriña, M. (2005). Empathic accuracy and inaccuracy in close relationships. In B. F. Malle & S. D. Hodges (Eds.), *Other minds: How humans bridge the divide between self and others* (pp. 310–322). New York: Guilford.

Ickes, W., Stinson, L., Bissonnette, V., & Garcia, S. (1990). Naturalistic social cognition: Empathic accuracy in mixed-sex dyads. *Journal of Personality and Social Psychology*, 59, 730–742.

Johnson, D. R. (2012). Transportation into a story increases empathy, prosocial behavior, and perceptual bias toward fearful expressions. *Personality and Individual Differences*, 52, 150–155.

Klein, K. J. K., & Hodges, S. D. (2001). Gender differences, motivation and empathic accuracy: When it pays to understand. *Personality and Social Psychology Bulletin*, 27, 720–730.

Knafo, A., Steinberg, T., & Goldner, I. (2011). Children's low affective perspective-taking ability is associated with low self-initiated pro-sociality. *Emotion*, 11, 194–198.

Knutson, B. (1996). Facial expressions of emotion influence interpersonal trait inferences. *Journal of Nonverbal Behavior*, 20, 165–182.

Lamm, C., Batson, C. D., & Decety, J. (2007). The neural substrate of human empathy: Effects of perspective-taking and cognitive appraisal. *Journal of Cognitive Neuroscience*, 19, 42–58.

Lawrence, E. J., Shaw, P., Baker, D., Baron-Cohen, S., & David, A. S. (2004). Measuring empathy: Reliability and validity of the Empathy Quotient. *Psychological Medicine*, 34, 911–919.

Levenson, R. W., & Ruef, A. M. (1992). Empathy: A physiological substrate. *Journal of Personality and Social Psychology*, 63, 234–246.

Lewis, K. L., Hodges, S. D., Laurent, S. M., Srivastava, S., & Biancarosa, G. (2012). Reading between the minds: The use of stereotypes in empathic accuracy. *Psychological Science*, 23, 1040–1046.

Lopes, P. N., Grewal, D., Kadis, J., Gall, M., & Salovey, P. (2006). Evidence that emotional intelligence is related to job performance and affect and attitudes at work. *Psicothema*, 18, 132–138.

Marsh, A. A., Adams, R. B., Jr., & Kleck, R. E. (2005). Why do fear and anger look the way they do? Form and social function in facial expressions. *Personality and Social Psychology Bulletin*, 31, 73–86.

Marsh, A. A., & Ambady, N. (2007). The influence of the fear facial expression on prosocial responding. *Cognition and Emotion*, 21, 225–247.

Marsh, A. A., & Blair, R. J. R. (2008). Deficits in facial affect recognition among antisocial populations: A meta-analysis. *Neuroscience and Biobehavioral Reviews*, 32, 454–465.

Marsh, A. A., Finger, E. C., Schechter, J. C., Jurkowitz, I. T. N., Reid, M. E., & Blair, R. J. R. (2011). Adolescents with psychopathic traits report reductions in physiological responses to fear. *Journal of Child Psychology and Psychiatry*, 52, 834–841.

Marsh, A. A., Kozak, M. N., & Ambady, N. (2007). Accurate identification of fear facial expressions predicts prosocial behavior. *Emotion*, 7, 239–251.

Marsh, A. A., Stoycos, S. A., Brethel-Haurwitz, K. M., Robinson, P., VanMeter, J. W., & Cardinale, E. M. (2014). Neural and cognitive characteristics of extraordinary altruists. *Proceedings of the National Academy of Sciences*, 111, 15036–15041.

Matsumoto, D., & Willingham, B. (2009). Spontaneous facial expressions of emotion of congenitally and noncongenitally blind individuals. *Journal of Personality and Social Psychology*, 96, 1–10.

Mayer, J. D., Salovey, P., & Caruso, D. R. (2004). Emotional intelligence: Theory, findings, and implications. *Psychological Inquiry*, 15, 197–215.

Musser, E. D., Backs, R. W., Schmitt, C. F., Ablow, J. C., Measelle, J. R., & Nigg, J. T. (2011). Emotion regulation via the autonomic nervous system in children with attention-deficit/hyperactivity disorder (ADHD). *Journal of Abnormal Child Psychology*, 39, 841–852.

National Science Foundation, National Center for Science and Engineering Statistics. (2013). *Women, Minorities, and Persons with Disabilities in Science and Engineering: 2013*. Special Report NSF 13–304. Arlington, VA. Available at www.nsf.gov/statistics/wmpd/.

Nowicki, S., & Duke, M. P. (1994). Individual differences in the nonverbal communication of affect: The Diagnostic Analysis of Nonverbal Accuracy Scale. *Journal of Nonverbal Behavior*, 18, 9–35.

Papp, L. M., Kouros, C. D., & Cummings, E. M. (2010). Emotions in marital conflict interactions: Empathic accuracy, assumed similarity, and the moderating context of depressive symptoms. *Journal of Social and Personal Relationships*, 27, 367–387.

Sato, W., Uono, S., Matsuura, N., & Toichi, M. (2009). Misrecognition of facial expressions in delinquents. *Child and Adolescent Psychiatry and Mental Health*, 3, Article ID 27. doi:10.1186/1753-2000-3-27.

Simpson, J., Ickes, W., & Blackstone, T. (1995). When the head protects the heart: Empathic accuracy in dating relationships. *Journal of Personality and Social Psychology*, 69, 629–641.

Snodgrass, S. E., Hecht, M. A., & Ploutz-Snyder, R. (1998). Interpersonal sensitivity: Expressivity or perceptivity? *Journal of Personality and Social Psychology*, 74, 238–249.

Stinson, L., & Ickes, W. (1992). Empathic accuracy in the interactions of male friends versus male strangers. *Journal of Personality and Social Psychology*, 62, 787–797.

Tager-Flusberg, H. (2005). Language and communication disorders in autism spectrum disorders. In M. L. Bauman & T. L. Kemper (Eds.), *The neurobiology of autism* (pp. 45–58). Baltimore, MD: Johns Hopkins University Press.

Young, S. G., & Hugenberg, K. (2010). Mere social categorization modulates identification of facial expressions of emotion. *Journal of Personality and Social Psychology*, 99, 964–977.

Zaki, J., Bolger, N., & Ochsner, K. (2008). It takes two: The interpersonal nature of empathic accuracy. *Psychological Science*, 19, 399–404.

Zaki, J., & Ochsner, K. N. (2012). The neuroscience of empathy: Progress, pitfalls and promise. *Nature Neuroscience*, 15, 675–680.

Zaki, J., Weber, J., Bolger, N., & Ochsner, K. (2009). The neural bases of empathic accuracy. *Proceedings of the National Academy of Sciences*, 106, 11382–11387.

Zaki, J., Weber, J., & Ochsner, K. (2012). Task-dependent neural bases of perceiving emotionally expressive targets. *Frontiers in Human Neuroscience*, 6, Article ID 228. doi:10.3389/fnhum.2012.00228.

Zebrowitz, L. A. (2001). Groping for the elephant of interpersonal sensitivity. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement* (pp. 333–350). Mahwah, NJ: Lawrence Erlbaum Associates.

Part III

Conclusions

18 Is there a general skill in perceiving others accurately?

R. Thomas Boone and Katja Schlegel

Abstract
Over the past decades, there has been increased interest in and recognition of the importance of interpersonal accuracy (IPA) for successful social interaction. Yet, while there has been some success in creating instruments that assess individuals' accuracy in inferring others' personality, affect and emotions, thoughts and feelings, deception, and social affordances, there have been no systematic efforts to examine the degree of overlap between these measures. The relatively few attempts in which relationships have been reported have shown surprisingly low correlations. Here, we utilized meta-analytic techniques to map out how different assessments of IPA are related to each other and to evaluate the question whether there is a global IPA skill. The results revealed a modest but significant overall correlation that was moderated by domain and nonverbal channels available in each measure. Implications for a unified construct of IPA are discussed.

> "And like the blind men, who lacked an overarching theory of elephantness and could not find it through local observations, psychologists lack an overarching theory of interpersonal sensitivity and will not find it through their local observations." Zebrowitz (2001, p. 334).

Accurately judging others' feeling states, emotions, intentions, traits, and other characteristics is the ether that binds us all together as social beings. Interpersonal accuracy (IPA) is thus a skill that is important on all levels at all times in all social interactions, from the first encounter with someone to managing a relationship with a close friend or partner. This skill makes social relationships and interactions more manageable and predictable and is adaptive from an evolutionary perspective (Ambady & Skowronsky, 2008). While the evolutionary perspective provides a distal explanation (species specific social adaptations) for the importance of IPA, models of interactions in the form of social dilemmas and exchange theory offer more proximal mechanisms (the decision to trust or not trust another in an ongoing interaction). In this capacity, IPA has impact on two levels. The first process is through highlighting the critical importance of accurate first impressions to determine others' suitability as a partner (Boone & Buck, 2003; Boone & Macy, 1999). Making accurate judgments of personality, emotions, and trustworthiness leads to the successful strategy of being able to lock into a cooperative relationship early, described as the cooperator's advantage (Orbell & Dawes, 1993). The second process is through allowing the perception of others as agents with the potential for loyalty and connectedness rather than a generalized other in a series of interchangeable interactions. Frank (1988) argued that emotions and related judgments are the glue that pulls us out of cold rational choice and allows us to build relationships and successfully engage in social coordination.

Even at zero acquaintance, judgment accuracy is often above guessing level, though the degree of accuracy varies from domain to domain and from measure to measure. Within groups, there is also individual variation, with some individuals being more accurate than others. Higher levels of IPA have been tied to socially relevant outcomes such as popularity, success in school and work, occupation, and interpersonal relationships (Halberstadt, Parker, & Castro, 2013; Hall, Andrzejewski, & Yopchick, 2009; Sternglanz & DePaulo, 2004).

Speaking of IPA as a general construct suggests a global ability underlying accurate judgments in all the different domains – a person who is good at judging others' emotions should also be good at judging whether another person is telling the truth. However, little is known about whether IPA is indeed a global, unitary ability or not. The goal of this chapter is to braid together the different IPA domains and measurement instruments to evaluate if a global, individual-level IPA skill exists. To this end, first, previous research from a historical perspective will be summarized. Second, different possible dimensional structures of IPA will be outlined. Third, these possible structures will be examined empirically using meta-analysis on correlations between different IPA measures. Finally, the results of this meta-analysis will be discussed with respect to psychometric obstacles that face researchers aiming to measure IPA and the mechanisms and processes that might underlie IPA.

Historically, the different domains in which IPA has been assessed were investigated in different lines of research with little overlap. For example, emotion researchers focused on the issue of universality versus cultural specificity in emotional expressions (Ekman et al., 1987; Izard, 1994), while researchers interested in the detection of social attributes looked at which broad categories of groups could be reliably identified (Borkenau & Liebler, 1992; Rule, Garrett, & Ambady, 2010), such as sexual orientation or religion. Research on first impressions focused on which personality traits were most readily inferred and how much or little exposure was required for an accurate inference (Ambady & Rosenthal, 1993; Carney, Colvin, & Hall, 2007). Lie detection research looked for factors that could improve accuracy, such as the motivation of the sender (DePaulo, Kirkendol, Tang, & O'Brien, 1988). For the most part, the main goal of such domain-specific endeavors was to demonstrate judgmental accuracy based upon group means and to compare different groups or variations in the stimuli with respect to performance. Individual performance levels were not always of primary interest. Therefore, measures of accuracy were usually designed to reach group accuracy and consisted of stimuli that best exemplified the trait or state of interest.

Arguably, the domain that focused on individual differences in accuracy the earliest was the field of affect and emotion recognition (Boone & Buck, 2004; Hall, 2001). To date, most standard IPA tests have been developed in this domain, with one of the earliest tests being the Profile of Nonverbal Sensitivity (PONS; Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979) in which an actor portrays different affect-laden scenes. For each scene, the test-taker is asked to choose which of two options correctly describes the enacted situation. Over the years, numerous tests measuring accuracy in recognizing discrete emotions have been published, such as the Brief Affect Recognition Test (BART; Ekman & Friesen, 1974), Communication of Affect Receiving Test (CARAT; Buck, 1976), Diagnostic Analysis of Nonverbal Accuracy (DANVA; Nowicki & Duke, 1994), Japanese and Caucasian BART (JACBART; Matsumoto et al., 2000), Multimodal Emotion Recognition Test (MERT; Bänziger, Grandjean, Scherer, & Scherer, 2009), Emotion Recognition Index (ERI; Scherer & Scherer, 2011), or the Geneva Emotion Recognition Test (GERT; Schlegel, Grandjean, & Scherer, 2014). In other IPA domains, many fewer standard tests are available. A notable exception is the Interpersonal Perception Task (IPT; Costanzo & Archer, 1989) that measures judgments related to competition, deception, kinship, intimacy, and status. Most studies in IPA domains other than emotion recognition used nonstandard instruments targeted at a specific research question and with little reference to other measures. To describe the state of the field, Zebrowitz (2001) used the analogy of a group of blind men trying to determine the nature of an elephant, each examining a different part of the elephant, leading each man to infer different qualities about that singular elephant.

Earlier efforts to explore the relationship between different IPA tests have yielded little to no support for a unified construct. Buck (1984) reviewed studies comparing individuals' scores on the PONS, BART, and CARAT, and noted generally low intercorrelations. For example, the correlation between the CARAT and PONS was $r = .04$. Hall (2001) reported similar findings, with the correlations between the PONS and the IPT being $r = .20$ and $r = .22$ in two samples, the correlation between the PONS and the CARAT being $r = .16$, and the correlation between the CARAT and the IPT being $r = .10$. These results present an odd paradox: Although these tests all measure accuracy in judging affect-related states and demonstrated predictive validity with social outcomes and other correlates, they seem to be quite independent from each other. The story appears to end there.

Based on theoretical considerations, several hypotheses regarding the structure of IPA are conceivable: First, there could be a single skill underlying IPA in all content domains. Second, there could be a common skill within each content domain, but the skills might be more or less unrelated across domains. Third, there could be no discernible structure in interpersonal accuracy, with each measurement tool assessing a unique facet of the general construct. In the following, we will discuss some reasons that might support each view.

IPA as single ability

There are several reasons to assume that IPA can be considered a single global ability. First, IPA measures generally require paying attention to and processing cues that are emitted by another person through physical appearance as well as verbal and nonverbal behavior. The lens Model (Brunswik, 1956) suggests that the mechanism through which the state or trait of a sender is encoded and decoded is based on these cues. Across the different IPA domains, there has been a systematic effort to isolate the relevant cues in different nonverbal media that signal the sender's state or trait of interest in different nonverbal channels including the face, voice, and body (DeMeijer, 1989; Ekman et al., 1987; Izard, 1994). Individuals with higher IPA might be more attentive to all of these cues, developing greater sensitivity to variation and nuance. Other researchers have proposed that there exists a superordinate set of cues relating to spatiotemporal forms, in other words geometric and temporal patterns that underlie cue features in all the nonverbal modalities whether visual or auditory (Scherer & Oshinsky, 1977). Recent research has shown that sensitivity to these spatiotemporal cues is related to emotion receiving ability across several modalities (Castro & Boone, 2015). Although not empirically extended to other IPA domains, such sensitivity to spatiotemporal cues may allow the receiver better utilization of a wide range of nonverbal cues and offer a global, or at least shared, underlying mechanism.

Further support for the view of a global IPA skill is that measures across domains tend to correlate in the same way with other individual difference variables. For instance, women tend to be more accurate than men in most nonverbal assessments, especially in emotion perception tasks (Hall & Matsumoto, 2004; Hall & Schmid Mast, 2008; Rosip & Hall, 2004). Furthermore, in emotion recognition, typically age effects are found. While increasing throughout childhood and adolescence, emotion recognition skills reach their height in young adulthood and decline in middle age and older adulthood (for reviews, see Chapter 10, and Halberstadt et al., 2013). Another process that might shape a common skill underlying all IPA domains could be early attachment processes that involve the development of communication and relationship between infant/toddler and caregiver, which several theorists place at the center of human sociability (Buck, 2014).

IPA as a set of distinct but correlated skills

Although all IPA judgments are based on nonverbal cues and/or physical appearance, IPA tasks differ considerably in the type of cue processing. Several distinctions on the level of cue processing can be made. First, interpersonal judgments can refer to others' states (such as emotions or deception) and others' traits (such as personality, intelligence, social status, or age; Hall, Bernieri, & Carney, 2005). Second, judgments can refer to attentional accuracy (noticing and recalling others' behaviors) or to inferential accuracy (making judgments about others' behaviors). Third, judgments can be about continuous evaluations (How extraverted is a person?), categorical choices (Which out of five emotions did the person express?), or dichotomous decisions (Did this person tell the truth or not?). These judgments are further complicated by differences in how the stimuli are presented (e.g., static photo vs. dynamic video, posed vs. spontaneous expressions). Thus, such judgments might require different cognitive decision making processes, and accuracy between the different domains might be not highly related or even unrelated. In addition, the different types of judgments might also be influenced by various individual characteristics, traits, and self-perceptions of the judge. For example, it would be plausible to assume that a strong belief in justice might motivate someone to be an accurate judge of deception but not necessarily a good judge of emotion expressions. Similarly, it might be that a person who rates him- or herself as high on emotional intelligence would be better at recognizing others' emotions, but not at evaluating others' extraversion. Despite these differences between IPA domains, this perspective assumes that within each domain, there should be a common underlying skill because of the similar type of judgment and decision being made. In this case, global IPA would be a multifaceted construct with looser connections across domains, but tighter connections within domains.

IPA as a set of unrelated and instrument-specific skills

This perspective is supported by the few empirical studies that found low correlations between different IPA tests and assumes that the main reason is that IPA measures differ a lot in their content and tasks, even within one domain. Emotion recognition tests use pictures of facial expressions, voice recordings, images of body postures, video clips with or without sound, or combinations of these channels, in which someone expresses an emotional state. In addition, diagnostic cues for different emotions tend to appear selectively in specific channels (App, McIntosh, Reed, & Hertenstein, 2011) rather than globally across channels.

Personality judgments are typically based on still pictures of faces with neutral expressions, on voice recordings, or on video clips of the target person doing a variety of things (introducing him- or herself, negotiating, or being interviewed for a job), often with the linguistic information maintained in the clips. Deception detection requires catching the liar regarding at least two different components: An affective component such as anxiety or guilt, or a cognitive component where the would-be liar must manage the complexity of lying that lines up with all the details of the truth (DePaulo, LeMay, & Epstein, 1991). Arguably, within each IPA domain, judgments represent a combination of tasks relying on different types of information and might therefore involve quite distinct processes. Some tests, such as the IPT, have fairly heterogeneous content, in an apparent effort to span a wider range of social content and IPA judgments. Heterogeneous content and different channels (modalities) also require knowledge about which cues to pay attention to for various types of judgments. Individuals might specialize only in some types of content or judgments, leading to lower associations between measures that differ in these respects.

Even within one channel, such as the face, different mechanisms might be involved in inferring information from someone's facial features versus his or her emotional expressions, which involve muscle movements (Ekman & Friesen, 1978). Furthermore, within one channel, stimuli differ in the spontaneity of encoded behavior and in stimulus dynamism (Hall et al., 2005). Spontaneity can range from spontaneous (e.g., unrehearsed behavior recorded during a social interaction) to deliberate and posed (e.g., facial expressions posed by actors).

Each way in which stimuli in IPA tasks can differ might require distinct skills from the judge that draw on different types of cognitive processing, such as more superficial and automatic processes versus more conscious and resource-consuming processes. These skills might be quite independent, thus leading to no specific pattern between the different IPA measures even within a content domain. For example, judging a prototypic "basic" emotion such as disgust might require much less cognitive effort than judging a subtle or more complex emotion such as embarrassment. Nevertheless, each of the various tasks and ways to measure IPA might still represent a justifiable and valid assessment approach that captures a different facet of how people make accurate interpersonal judgments.

To summarize, there are a number of arguments for each of the three possible perspectives on the structure of IPA. The following section presents some preliminary results of an extensive meta-analysis aimed at evaluating the proposed perspectives undertaken by the present authors along with Judith A. Hall.

A meta-analysis

The present meta-analysis examined the correlations between IPA measures across the following domains of IPA: Emotions, thoughts and feelings, affective states more broadly, personality, deception, and social affordances, the latter referring to judgments on social categories and group affiliations such as political ideology, or being gay or straight. In particular, the meta-analysis aimed to address the following questions: What is the average correlation between IPA measures? Do measures within one domain correlate more highly than measures across domains? And do measures using the same cue channels correlate more highly than measures using different cue channels?

Description of dataset. The present meta-analysis included published and unpublished studies that reported correlations between two or more measures in which judges assessed states, traits, or personal attributes of target persons based on their recorded verbal and/or nonverbal behavior. These measures had to be scored for accuracy according to an external criterion. The search for studies involved an extensive search of published literature as well as unpublished dissertations, contacting researchers in the IPA field, and announcements on the list serves of respective professional organizations. To be included in the present analysis, studies had to (1) be reported in English, (2) have a sample with a mean age of at least 14 years, and (3) have a sample size of at least ten. For the current analysis, we excluded any correlations in which two measures represented subparts of the same test, defined as measures with overlapping cues from the same target material where learning effects from answering previous items could have occurred.

Seventy-seven sources were identified that reported a total of 92 studies with independent samples. These 92 studies provided a total of 587 effect sizes, each of which was a Pearson correlation reflecting the association between two IPA measures. The number of effect sizes per study ranged from 1 to 82 with a median of 9.5 effect sizes per study.

We identified and coded the domain and cue channels of each of the two involved measures for each effect size. Specifically, each measure was sorted into one of six IPA domains: personality, emotion, situational affect (using emotional enactments to infer the social event being enacted), deception, thoughts and feelings (measures involving empathy and intention), and social affordances (using personal attributes to infer target social group). Each measure was also coded with respect to its cue channel configuration, which included the following categories: (1) face only, (2) voice only (speech that is content-masked so that no linguistic content is intelligible), (3) body only, (4) face with body, (5) face with voice, (6) face with voice and body, (7) face with voice and linguistic cues, (8) face with voice, body, and linguistic cues, and (9) eyes only.

Analysis and results. In cases where a given study sample was administered more than two IPA tests, the multiple effects resulting from correlating each test with each other test introduced non-independence into the database. Multilevel modeling (MLM), using a random effects approach to adjust for the nested data structure, yielded an unweighted averaged effect size of $r = .19$ ($SD = .15$); which significantly differed from zero, $t(71.6) = 10.07$, $p = <.001$, though it was not of large magnitude.

To explore how effect sizes vary for different domain combinations and cue channel combinations, a series of MLM analyses was conducted for each combination separately. Each MLM was based on all effect sizes for a particular combination and computed in the same way as the overall effect (see above). These results are presented in Table 18.1 for domain combinations and Table 18.2 for cue channel combinations. In addition, the *q* graph package for R (Epskamp, Cramer, Waldorp, Schmittmann, & Borsboom, 2012) was used to visually display the pattern of correlations provided in Tables 18.1 and 18.2, respectively. Specifically, the Fruchterman–Reingold algorithm (Fruchterman & Reingold, 1991) was applied to create a layout in which the distance and connection between nodes (representing the domains or cue channel configurations, respectively) reflects the closeness of their association.

Table 18.1 *Average correlations for each interpersonal domain combination*

	Personality	Emotion	Situational affect	Deception	Thoughts and feelings	Social affordances
Personality	.08** (69)					
Emotion	.07** (76)	.29 *** (108)				.
Situational affect	.08* (7)	.19*** (94)	.12** (11)			
Deception	–.03 (10)	.09 (14)	.09** (18)	.40 (2)		
Thoughts and feelings	.08 (10)	.25 *** (14)	.06 (2)			
Social affordances	.05 (11)	.06 **(46)	.08*** (57)	.05 (11)	.00 (6)	.01 (21)

Note. Total number of effect sizes was $N = 587$. The correlations were adjusted for non-independence using MLM. Effects on the diagonals refer to pairs of measures belonging to the same domain. The asterisks refer to the significance levels comparing the average correlation per combination of each interpersonal domain to zero,†$p < .10$.*$p < .05$.**$p < .01$.**$p < .001$,***$p < .001$.

Regarding the different IPA domains, Table 18.1 shows that the combinations that have been studied the most include emotion with emotion, situational affect with emotion, emotion with personality, and personality with personality. (Most studies in the emotion category were concerned with the so-called basic emotions, but other emotional states were also included.) In contrast, accuracy in judging thoughts and feelings (an accuracy construct based on the empathic accuracy paradigm; see Chapter 3) contributed few effect sizes in total. The combinations thoughts and feelings with thoughts and feelings, and thoughts and feelings with deception, have never been reported. Figure 18.1 therefore omitted the thoughts and feelings domain. Table 18.1 shows that the largest mean effect was found when emotion measures were correlated with other emotion measures ($r = .31$, $p < .001$). Other combinations that yielded effect sizes of notable magnitude were emotion correlated with situational affect ($r = .18$, $p < .001$) and emotion correlated with thoughts and feelings ($r = .23$, $p < .01$). Deception, personality, and social affordances each showed weak correlations with any other domain. Personality and deception measures were significantly correlated with other measures in the same domain, but social affordances were not. Figure 18.1 visualizes this pattern and shows that a) relationships between all domains (even when not significant) were positive with the exception of personality with deception and b) situational affect and emotion were the most closely related domains. This is not surprising because, as defined earlier, situational affect also involved the judgment

Table 18.2 *Average correlations for each nonverbal channel configuration combination*

	F	V	FV	B	FB	FVB	FVL	FVBL	E
F	.25 *** (38)								
V	.20*** (33)	.22*** (15)							
FV	.35* (3)		.25 (2)						
B	.15** (17)	.07 (6)		.08* (6)					
FB	.18***(19)	.11 (5)		.15† (4)	.10*** (55)				
FVB	.21*** (26)	.30** (7)	.25* (5)	.20 (2)	.42*** (8)	.16*** (40)			
FVL	.08 (6)								
FVBL	.07** (48)	.10*** (43)		.12*** (25)	.06 (19)	.10*** (23)		.06 ** (120)	
E	.16 (2)	.09 (1)		.38 (1)				.26** (5)	.06 (3)

Note. Total number of effect sizes was N = 587. F = Face only; V = Prosody only; FV = Face and prosody; B = Body only; FB = Face and body; FVB = Face, prosody, and body; FVL = Face, prosody, and linguistic; FVBL = Face, prosody, body, and linguistic; and E = Eyes only. The asterisks refer to the significance levels comparing the average correlation per cue channel combination to zero,†p < .10.*p < .05.**p < .01.** p < .001,***p < .001.

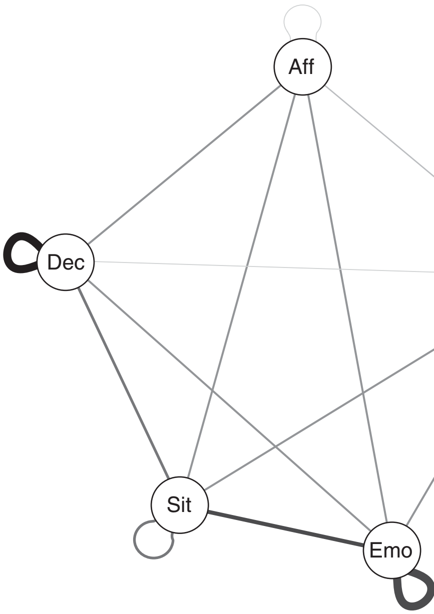

Figure 18.1 Visual representation of the correlational pattern within and between five IPA domains.

of emotional cues, though not tested in terms of discrimination between discrete emotions. For example, on the PONS test (Rosenthal et al., 1979), test-takers decide what kind of affective situation the target person was in, for example talking to a lost child versus ordering food in a restaurant, or asking forgiveness versus expressing jealous rage.

Regarding cue channel configurations, Table 18.2 shows that the most frequently examined configurations included measures using face, voice, body, and linguistic cues (FVBL) with other FVBL measures, face and body (FB) with other FB measures, and FVBL with face-only measures. Measures including the face and voice (FV), the face, body, and linguistic cues without voice (FBL), and the eyes only were not widely studied. In Figure 18.2, the FBL and FVBL categories and the FV and FVB (face, voice, and body) categories were therefore combined and eyes were omitted. Table 18.1 shows that the largest effects were found for FVB measures, with the highest effect size observed when FVB measures were correlated with FB measures ($r = .45$, $p < .05$). Generally, measures with the same or overlapping channel configurations tended to be substantially correlated with each other, with the exception of measures using all four

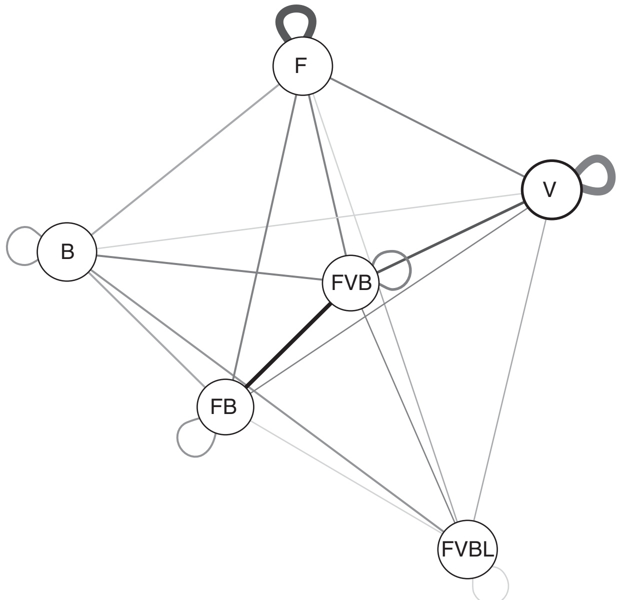

Figure 18.2 Visual representation of the correlational pattern within and between the six main channel configurations.

channels of face, voice, body, and linguistic. Figure 18.2 visualizes the correlational pattern and shows that measures including multiple channels (with the exception of linguistic cues) overall tend to have the highest associations with other measures. This is especially the case for measures covering three channels (FVB), which are therefore located in the center of the figure. In contrast, measures of single channels, in particular voice only and body only, tend to have somewhat weaker links with other measures. Tests including linguistic cues generally are more distinct from other measures. This finding is likely to be related to the fact that domains and cue channel configurations are confounded. In particular, FVBL measures mostly refer to personality and almost never to emotion.

Overall, these results suggest that IPA might not be a single skill, but rather a set of distinct but correlated skills, in line with our second theoretical proposition. Accuracy of judging situational affect and emotion seem to be based on shared mechanisms as suggested both by the higher intra-domain correlations and the closer relationship between these domains. Moreover, emotion might be the most central or global IPA domain. In contrast, personality and social affordances are rather disparate categories. Results also suggest that multichannel nonverbal measures might capture a more global IPA dimension in comparison to single channel measures and measures with linguistic content. However, this meta-analysis also demonstrated that, in line with previous studies, correlations between IPA measures are generally low.

The present findings can be interpreted both from a theoretical as well as a psychometric angle. The psychometric angle examines how IPA measures have been developed in the past, highlights some potential reasons that might have contributed to the low correlations between tests, and provides some general recommendations for future test developers in IPA. The theoretical angle tries to explain the present findings in terms of possible brain mechanisms underlying somewhat specific skills under the umbrella of global IPA.

Psychometric issues in the measurement of IPA

From a psychometric perspective, IPA is a very diverse field with a variety of measures that rely on different types of stimuli, response formats, and ways to define accuracy. Notably, many researchers use nonstandard measures that are often created only for one specific study. Such measures are usually not validated and their psychometric properties are often not reported. In the present meta-analysis, 43% of all effect sizes were based on at least one nonstandard measure. This could be one factor contributing to the low correlations between IPA measures that we found. The widespread use of nonstandard measures is related to the fact that for several IPA domains, in particular personality judgments, no standard tests to measure individual differences exist. In fact, as mentioned earlier, the only domain with a variety of standard tests is emotion recognition and the related domain we called situational affect. However, even for standardized tests, psychometric properties are often not clear. For example, the factorial structure is only known for a few tests (e.g., for the PONS, see Rosenthal et al., 1979; for the MERT and the GERT, see Schlegel, Grandjean, & Scherer, 2012). It also remains to be investigated whether IPA tests have the same measurement properties across cultures, languages, genders, or age groups. Several researchers have previously noted that a possible reason for the low intercorrelations of IPA measures could be psychometric problems (Hall et al., 2005; Kenny, 2013).

An issue of particular importance with respect to psychometric quality is the internal consistency or reliability of a test. In IPA, low reliability (<.50 in terms of Cronbach's alpha) is very common (Kenny, 2013) and, often, reliability is not reported at all. There are several possible reasons for low internal consistency. First, during test development, items might not have been evaluated and selected based on how well they distinguish or discriminate between people with different ability levels. This might result in the inclusion of items that do not measure the intended construct well, for example because they lack the necessary cues. Second, the test might be too easy to be able to discriminate between individuals in the higher ability range, which limits variance and consequently leads to lower inter-item correlations (but see Kenny, 2013). Third, the test might not measure one underlying ability (which is a central assumption in the classical reliability concept) but could consist of multiple factors that are not highly correlated and thus restrict overall internal consistency. Such factors might for example be defined by different cue channels, method factors (e.g., subsets of items with the same response options clustering together), or different kinds of content. If these dimensions are not identified and the test is mistakenly treated as unidimensional, internal consistency might be low.

While IPA as a field lacks the strong psychometric tradition of other fields like cognitive intelligence, it should also be noted that the type of stimuli used in IPA research – portrayals of other individuals who emit verbal and nonverbal cues – are generally much more naturalistic and complex than in other fields of performance-based testing like cognitive intelligence. While in intelligence tests item characteristics and difficulty can be systematically varied, portrayals of human behavior such as emotional expressions cannot be as easily manipulated and controlled. This complicates the endeavor of creating internally consistent tests.

Given that IPA is a diverse field with measures that were created for a variety of different purposes, an overall "one-size-fits-all" instruction for future test development would not be appropriate. However, the following general recommendations might help the field to coordinate and systematize how measures are constructed, and thus, to better understand why different measures are more or less correlated. Generally, test development and test validation should aim at following the steps that have been proposed by various researchers (e.g., DeVellis, 2012). In particular, test development should be based on a clear definition of the objectives of the new instrument, including decisions about what skill exactly will be measured and of which subskills it might consist, who will be tested, and who will administer the test. These questions have important implications for various decisions during the process of generating stimuli and response options. For example, what will be the criterion for assessing accuracy – e.g., a panel of experts, consensus of a large group of raters, some sort of objective coding, or the self-ratings of the target (Ekman, Friesen, & Ellsworth, 1972)? What is the appropriate response format for making the respective judgments – continuous ratings, a forced choice format, or free responses (e.g., Frank & Stennett, 2001; Russell, 1994)? Should stimuli be presented in a static or dynamic mode and should cue channels be presented in isolation or in a multimodal way? Should stimuli be posed or spontaneous (e.g., Kang, 2012)? Should stimuli be preselected for producing a desired accuracy level or for containing specific diagnostic cues, or should the stimuli not be selected for having any particular features (Hall, Andrzejewski, Murphy, Schmid Mast, & Feinstein, 2008)? While a case can be made for each choice, the choices that previous researchers made have added considerable method variance to the dataset of the present meta-analysis, which might partly explain the rather low test intercorrelations. It is daunting to envision a study in which a standardized approach to measurement was used on stimuli that were equally effective in measuring individual differences in all IPA domains, in order to reduce such method variance.

Aside from choices concerning stimuli and their presentation, researchers in IPA should be encouraged to make use of modern psychometric techniques to improve the quality of the item selection process. Almost all psychometric work in IPA was based on Classical Test Theory (CTT). However, the more recent framework of Item Response Theory (IRT) is much more flexible than CTT and offers models adapted to different item types commonly used in IPA, including binary items (i.e., two-option multiple choice) and ordinal items (e.g., ratings). Advantages of IRT over CTT include the possibility to select items with the highest measurement precision for a desired ability range and to assess measurement precision of items independently from the other items and the particular participant sample (see Embretson & Reise, 2000). IRT also provides specific reliability indices for tests with binary items for which traditional Cronbach's alpha is not appropriate (Dimitrov, 2003). While the use of IRT is common in other fields, IPA researchers only recently began adopting this framework. Schlegel and colleagues (2014) have successfully used IRT in the development and validation of the GERT. In the present meta-analysis, the GERT showed a higher correlation with other ERA tests than the average coefficient in this domain ($r = .46$ vs. $r = .31$), which suggests that IRT could contribute to better IPA measurement.

Dissecting IPA: common and specific mechanisms and processes

The aim of this section is to discuss the results of our meta-analysis by presenting theoretical and empirical evidence for both a general, broad IPA skill as well as for domain- or task-specific IPA subskills. Such evidence can be found in the fields of social cognition and social and affective neuroscience (Lieberman, 2010). From a theoretical perspective, the notion of embodiment has been very influential and can provide a ground for assuming a common basis in IPA skills across domains. Embodiment refers to the idea that all cognitive representations and operations are grounded in the body (Niedenthal, Barsalou, Winkielman, Krauth-Gruber, & Ric, 2005). Embodiment theories propose that the verbal or nonverbal behavior of another person triggers perceptual, somato-visceral, and motoric re-experiencing or simulation of this person's state in the perceiver. Such re-experiencing activates the perceiver's conceptual knowledge necessary to make a judgment about the other person.

To date, the embodiment hypothesis has been predominantly studied in emotion recognition. In particular, there is now converging evidence for the occurrence of motoric re-experiencing or mimicry during the perception and interpretation of facial expressions (Gallese, Keysers, & Rizzolatti, 2004), although it remains debated whether this motoric type of embodiment is a requirement for accurate emotion recognition (see Chapter 11; Bogart & Matsumoto, 2010; Hess & Blairy, 2001). Emotional contagion, which describes emotional re-experiencing without motoric reenactment, has been proposed as an alternative embodied mechanism in emotion recognition (Hatfield, Cacioppo, & Rapson, 1993). Importantly, embodiment is assumed to underlie most areas of understanding others' behavior, including making inferences about another person's affective state, thoughts, intentions, or traits (Niedenthal et al., 2005). For instance, when observing a facial expression in another person, the perceiver will reexperience or simulate its associated perceptual, somato-visceral, and possibly motoric elements that will activate knowledge about traits and states related to this expression. Depending on the specific task (e.g., inferring the emotional state versus judging the truthfulness of this person), the relevant knowledge is accessed and used to make a judgment. This idea implies a general and broad IPA mechanism across many, though perhaps not all, of its domains.

While embodiment could be the common mechanism underlying several domains of IPA, some theories suggest that more specific processes could be at work for different contents and types of interpersonal perception and judgments. For example, according to Associated Systems Theory (Carlston, 1994) social stimuli are simultaneously represented in separate visual, verbal/semantic, affective, and action systems. Each of these systems contains low-level structures for the perception of cues and high-level structures for abstract concepts that are used in the production of behavioral responses. These four systems could be differentially engaged in interpersonal perception depending on the specific task.

For example, typical emotion recognition tasks might primarily engage the affective system while personality judgments and the identification of thoughts and feelings might rely more heavily on the verbal/semantic system. Given that these systems are assumed to operate largely independently (although conceptual representations are linked between systems), it is likely that accuracy in one task does not necessarily predict accuracy in another task. Moreover, Carlston (1992) suggested that individuals might differ in how effectively they use each of the four systems and in the extent to which they use one or several of the systems (representational complexity). As a consequence, individuals might display specific patterns of accuracy across IPA domains and cue channels with relative individual strengths. Such individual differences in accuracy patterns across judges might result in less strong associations between IPA measures in different domains and cue channels.

Besides these theoretical propositions, the increasing number of empirical studies in social and affective neuroscience can inform the question of which mechanisms and processes underlie the different IPA domains and cue channels. Some studies have for example looked at whether tasks that require participants to label emotional expressions engage the same brain regions as tasks that require participants to rate personality traits (e.g., Heberlein & Saxe, 2005). Other relevant current research focuses on disentangling which brain circuits are implied in inferring *temporary* states in others such as intentions, beliefs, or thoughts, versus *enduring* dispositions of others such as traits or social categories (e.g., Bombari, Schmid Mast, Brosch, & Sander, 2013; van Overwalle, 2009; Zaki & Ochsner, 2011).

Although to date there is no agreement yet about the nature and distinctiveness of the mechanisms involved in making various interpersonal judgments, results suggest that the processing of emotional information might subserve many types of social cognition. For example, from his meta-analysis on neuroscientific data, van Overwalle (2009) concluded that decoding affective cues represents a general-purpose mechanism of social cognition that is often involved in forming a trait impression, in inferring someone's thoughts, as well as in detecting deception.

A current limitation of experimental and neuroscientific research on IPA is that it has largely omitted individual differences. While focusing on the processes that take place while people attempt to understand each other, researchers often neglect how successful these attempts are (Zaki & Ochsner, 2011). In other words, little is known about why some individuals are more accurate at making interpersonal judgments than others, especially in the healthy population. In recent years, some studies have started to explore this question by examining structural features of the brain in relation to IPA and social adjustment (e.g., Frühholz, Schlegel, & Grandjean, 2015). Such studies are a promising next step toward understanding the neural basis of individual differences in IPA.

Taken together, theoretical accounts and empirical results suggest that embodied cognitive processes, in particular in the processing of affective information, could characterize a broad, common IPA skill. In addition, the different IPA domains and tasks are likely to draw on more specific skills represented by special brain circuits and processes. Moreover, individual differences in the underlying processes or in personality traits affecting these processes can lead to a dissociation of accuracy levels between IPA tasks. Such processes and traits could include preferences for global versus local processing (Christman, 2001), abstract versus concrete representations, self-referent versus external sources of information (Carlston, 1992), or attention to visual versus vocal cues (Bänziger et al., 2009). Thus, while individuals can be assessed on general IPA, they are also likely to vary in terms of relative domain-specific strengths and weaknesses.

Such relative differences in IPA subskills can have real-life consequences, as has been demonstrated for channel-specific sensitivity to nonverbal cues. Specifically, higher sensitivity to less controllable or "leaking" vocal or bodily cues relative to easily controllable facial cues in others can be related to lower rapport with an interaction partner (Puccinelli & Tickle-Degnen, 2004). In the workplace, Tickle-Degnen (1998) demonstrated that relatively higher accuracy in reading covert cues might constitute an advantage or a disadvantage depending on the demands of the professional setting. However, differential predictive effects of channel- or domain-specific IPA skills to date remain understudied and should be addressed in future research (Hall et al., 2009).

Summary, implications, and conclusion

This chapter examined the question of whether IPA is one skill or several skills and whether a dimensional structure can be found among the great variety of IPA measures that we were able to study in conjunction with one another. While theoretically, different structures of the IPA domain – or even an absence of structure – could be justified, empirically, this question has never been addressed in a comprehensive way. We therefore conducted a meta-analysis of previous research that examined correlations between measures assessing individual differences in how accurately people judge others' affective states, deception, thoughts and feelings, personal attributes, or personality traits.

The results suggest that the associations between IPA measures are generally low, but they tend to be higher among measures in the same interpersonal domain. IPA thus seems to be a very broad construct that subsumes loosely related domain-specific facets which in themselves likely draw on sets of rather heterogeneous skills. However, all measures are conceptually linked because they refer to making judgments about others' characteristics based on verbal and nonverbal cues. Such a view of IPA is in line with Bollen and Lennox's (1991) perspective on *"causal indicators"* for establishing the validity of a construct. This perspective assumes that a broad construct such as IPA is captured in the most valid way when all facets that causally determine the construct are assessed. The less these facets or indicators are correlated, the more unique incremental variance they can add to explaining global IPA. This has indeed been argued in the IPA context by Bänziger, Scherer, Hall, and Rosenthal (2011). The various facets, which in our case are represented by the different IPA tests, are not interchangeable. Therefore, omitting any of the facets would omit part of the overall IPA construct and hence reduce its validity. This *"causal indicators"* perspective is opposed to the *"effect-indicators"* perspective that represents the basis of CTT. In the *"effect-indicators"* approach, indicators of a construct (i.e., test items) are not seen as potentially independent constituents of the construct, but as "effects" caused by the construct. Given that each indicator represents a repeated effect or outcome of the same construct, the indicators are supposed to be substantially correlated. The results of the present meta-analysis suggest that the *"causal-indicators"* approach reflects the IPA domain more appropriately: the variety of weakly correlated IPA measures across domains and channels captures many different facets of global IPA and collectively add to its construct validity.

The present results have several implications for future studies using IPA measures. Most importantly, IPA measures are not interchangeable. It seems that even within the more homogeneous domains like emotion recognition, instruments share only a limited amount of variance. Researchers should thus consider the various characteristics along which IPA measures can differ to make a selection that is matched to the specific research question. If researchers want to assess emotion recognition ability, they could decide to use a test that presents stimuli in different separate cue channels (e.g., the face, voice, or postures DANVA tests) in case they have reasons to specifically expect accuracy in these channels to be linked to their phenomenon of interest. In contrast, if researchers intend to measure emotion recognition in a way that is more ecologically valid and more closely linked to everyday social functioning, they might choose a test with dynamic, naturalistic, and multimodal stimuli that cover many positive and negative emotions (e.g., the GERT; Schlegel et al., 2014) instead of a test showing just static photographs of full-blown exemplars of a few prototypic emotional expressions (e.g., the JACFEE; Biehl et al., 1997). Alternatively, researchers might use several IPA measures and make different predictions for each measure. For example, if a researcher wanted to compare professional ballet dancers to non-dancers, they might hypothesize that dancers would make more accurate judgments than non-dancers based on the body channel, but not based on other channels. Another implication of the present meta-analysis is that for several IPA domains, in particular personality judgments and deception, valid standardized tests need to be developed. This would create more coherence in the field, allow for better comparisons of results across studies, and enable a more systematic investigation of the structure of IPA. Furthermore, it would be desirable to develop a test of general IPA sampling the different domains.

To conclude, let us go back to Zebrowitz's (2001) metaphor comparing IPA to an elephant, in which the different domains and subskills represent different body parts that lack clear common features of "elephantness" if studied in isolation "in the dark." We believe that with the present meta-analysis, we have gained a fuller understanding of the elephant. However, our results also suggest that a different metaphor might describe IPA more appropriately. Given their distinctiveness, the different interpersonal domains could also be compared to different types of fruit such as citrus fruits, apples, and bananas. Whereas more specific skills can be considered different varieties of these fruits, e.g., clementines, grapefruits, or blood oranges, IPA refers to their superordinate category "fruit." All fruit have something in common, but are quite distinct and have their own characteristics, with some fruit being more similar to each other than others. The fruit metaphor for IPA implies that for many research questions this superordinate category is too coarse to make concrete predictions and it might be more suitable to stick to the different fruit varieties. Nevertheless, we believe that IPA does merit being studied and presented in the literature as a global construct that is central to interpersonal communication and social functioning.

References

Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology*, 64, 431–441.

Ambady, N. E., & Skowronski, J. J. (2008). *First impressions*. New York: Guilford Publications.

App, B., McIntosh, D. N., Reed, C. L., & Hertenstein, M. J. (2011). Nonverbal channel use in communication of emotion: How may depend on why. *Emotion*, 11, 603–617.

Bänziger, T., Grandjean, D., Scherer, K. R., & Scherer, K. R. (2009). Emotion recognition from expressions in face, voice, and body. *The Multimodal Emotion Recognition Test (MERT)*. *Emotion*, 9, 691–704.

Bänziger, T., Scherer, K. R., Hall, J. A., & Rosenthal, R. (2011). Introducing the MiniPONS: A Short Multichannel Version of the Profile of Nonverbal Sensitivity (PONS). *Journal of Nonverbal Behavior*, 35, 189–204.

Biehl, M., Matsumoto, D., Ekman, P., Hearn, V., Heider, K., Kudoh, T., & Ton, V. (1997). Matsumoto and Ekman's Japanese and Caucasian Facial Expressions of Emotion (JACFEE): Reliability data and cross-national differences. *Journal of Nonverbal Behavior*, 21, 3–21.

Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, 110, 305–314.

Boone, R. T., & Buck R. (2003). Emotional expressivity and trustworthiness: The role of nonverbal behavior in the evolution of cooperation. *Journal of Nonverbal Behavior*, 27, 163–182.

Boone, R. T., & Buck R. (2004). Emotion receiving ability: A new view of measuring individual differences in the ability to accurate. In G. Geher (Ed.), *Measuring emotional intelligence: Common ground and controversy* (pp. 73–89). New York: Nova Science Publishers.

Boone, R. T., & Macy, M. W. (1999). Unlocking the doors to the prisoner's dilemma: Dependence, selectivity, and cooperation. *Social Psychology Quarterly*, 62, 32–52.

Bogart, K. R., & Matsumoto, D. (2010). Facial mimicry is not necessary to recognize emotion: Facial expression recognition by people with Moebius syndrome. *Social Neuroscience*, 5, 241–251.

Bombari, D., Schmid Mast, M., Brosch, T., & Sander, D. (2013). How interpersonal power affects empathic accuracy: Differential roles of mentalizing versus mirroring? *Frontiers in Human Neuroscience*, 7, 375.

Borkenau, P., & Liebler, A. (1992). Trait inferences: Sources of validity at zero acquaintance. *Journal of Personality and Social Psychology*, 62, 645–657.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments* (2nd ed.). Berkeley: University of California Press.

Buck, R. (1976). A test of nonverbal receiving ability: Preliminary studies. *Human Communication Research*, 2, 162–171.

Buck, R. (1984). *The communication of emotion*. New York: Guilford Press.

Buck, R. (2014). *Emotion: A biosocial synthesis*. Cambridge: Cambridge University Press.

Carlston, D. E. (1992). Impression formation and the modular mind: The Associated Systems Theory. In L. L. Martin & A. Tesser (Eds.), *The construction of social judgments* (pp. 301–341). Hillsdale, NJ: Erlbaum.

Carlston, D. E. (1994). *Associated Systems Theory: A systematic approach to cognitive representations of persons*. Hillsdale, NJ: Erlbaum.

Carney, D. R., Colvin, C. R., & Hall, J. A. (2007). A thin slice perspective on the accuracy of first impressions. *Journal of Research in Personality*, 41, 1054–1072.

Castro, V. L., & Boone, R. T. (2015). Sensitivity to spatiotemporal percepts predicts the perception of emotion. *Journal of Nonverbal Behavior*, 39, 215–240.

Christman, S. D. (2001). Individual differences in Stroop and local-global processing: A possible role of interhemispheric interaction. *Brain and Cognition*, 45, 97–118.

Costanzo, M., & Archer, D. (1989). Interpreting the expressive behavior of others: The Interpersonal Perception Task. *Journal of Nonverbal Behavior*, 13, 225–245.

DeMeijer, M. (1989). The contribution of general features of body movement to the attribution of emotions. *Journal of Nonverbal Behavior*, 13, 247–268.

DePaulo, B. M., Kirkendol, S. E., Tang, J., & O'Brien, T. P. (1988). The motivational impairment effect in the communication of deception: Replications and extensions. *Journal of Nonverbal Behavior*, 12, 177–202.

DePaulo, B. M., LeMay, C. S., & Epstein, J. A. (1991). Effects of importance of success and expectations for success on effectiveness at deceiving. *Personality and Social Psychology Bulletin*, 17, 14–24.

DeVellis, R. F. (2012). *Scale development: Theory and applications* (Vol. 26). Thousand Oaks, CA: Sage.

Dimitrov, D. M. (2003). Marginal true-score measures and reliability for binary items as a function of their IRT parameters. *Applied Psychological Measurement*, 27, 440–458.

Ekman, P., & Friesen, W. V. (1978). *Facial Action Coding System: A technique for the measurement of facial movement*. Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., & Friesen, W. V. (1974). Nonverbal behavior and psychopathology. In R. J. Friedman & M. Katz (Eds.), *The psychology of depression: Contemporary theory and research* (pp. 3–31). Washington, DC: Winston & Sons.

Ekman, P., Friesen, W. V., & Ellsworth, P. (1972). *Emotion in the human face*. New York: Pergamon.

Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., . . . & Tzavaras, A. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology*, 53, 712–717.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Hillsdale, NJ: Erlbaum.

Epskamp, S., Cramer, A. O., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). Qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software*, 48, 1–18.

Frank, M. G., & Stennett, J. (2001). The forced-choice paradigm and the perception of facial expressions of emotion. *Journal of Personality and Social Psychology*, 80, 75–85.

Franks, R. H. (1988). *Passions within reason: The strategic role of emotions*. New York: W. W. Norton.

Fruchterman, T. M., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience*, 21, 1129–1164.

Frühholz, S., Schlegel, K., & Grandjean, D. (2015). *Amygdala structure and core dimensions of the affective personality*. Manuscript submitted for publication.

Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences*, 8, 396–403.

Halberstadt, A. G., Parker, A. E., & Castro, V. L. (2013). Nonverbal communication: Developmental perspectives. In J. A. Hall & M. L. Knapp (Eds.), *Handbook of Communication Science* (Vol. 2, pp. 93–128). Berlin: Mouton de Gruyter.

Hall, J. A. (2001). The PONS Test and the psychometric approach to measuring interpersonal sensitivity. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement* (pp. 143–160). Mahwah, NJ: Lawrence Erlbaum.

Hall, J. A., Andrzejewski, S. A., Murphy, N. A., Schmid Mast, M., & Feinstein, B. (2008). Accuracy of judging others' traits and states: Comparing mean levels across tests. *Journal of Research in Personality*, 42, 1476–1489.

Hall, J. A., Andrzejewski, S. A., & Yopchick, J. E. (2009). Psychosocial correlates of interpersonal sensitivity: A meta-analysis. *Journal of Nonverbal Behavior*, 33, 149–180.

Hall, J. A., Bernieri, F. J., & Carney, D. R. (2005). Nonverbal behavior and interpersonal sensitivity. In J. A. Harrigan, R. Rosenthal, & K. R. Scherer (Eds.), *Handbook of nonverbal behavior research methods in the affective sciences* (pp. 237–281). New York: Oxford.

Hall, J. A., & Schmid Mast, M. (2008). Are women always more interpersonally sensitive than men? Impact of goals and content domain. *Personality and Social Psychology Bulletin*, 34, 144–155.

Hall, J. A., & Matsumoto, D. (2004). Gender differences in judgments of multiple emotions from facial expressions. *Emotion*, 4, 201–206.

Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1993). Emotional contagion. *Current Directions in Psychological Science*, 2, 96–99.

Heberlein, A. S., & Saxe, R. R. (2005). Dissociation between emotion and personality judgments: Convergent evidence from functional neuroimaging. *NeuroImage*, 28, 770–777.

Hess, U., & Blairy, S. (2001). Facial mimicry and emotional contagion to dynamic emotional facial expressions and their influence on decoding accuracy. *International Journal of Psychophysiology*, 40, 129–141.

Izard, C. E. (1994). Innate and universal facial expressions: Evidence from developmental and cross-cultural research. *Psychological Bulletin*, 115, 288–299.

Kang, S.-M. (2012). Individual differences in recognizing spontaneous emotional expressions: Their implications for positive interpersonal relationships. *Psychology*, 3, 1183–1188.

Kenny, D. A. (2013). Issues in the measurement of judgmental accuracy. In S. Baron-Cohen, H. Tager-Flusberg, & M. V. Lombardo (Eds.), *Understanding other minds: perspectives from developmental social neuroscience* (pp. 104–116). Oxford, UK: Oxford University Press.

Lieberman, M. D. (2010). Social cognitive neuroscience. In S. T. Fiske, D. T. Gilbert, & L. Gardner (Eds.), *Handbook of social psychology* (5th ed., pp. 143–193). Hoboken, NJ: Wiley.

Matsumoto, D., LeRoux, J., Wilson-Cohn, C., Raroque, J., Kooken, K., Ekman, P., . . . Goh, A. (2000). A new Test to measure emotion recognition ability: Matsumoto and Ekman's Japanese and Caucasian Brief Affect Recognition Test (JACBART). *Journal of Nonverbal Behavior*, 24, 179–209.

Niedenthal, P. M., Barsalou, L. W., Winkielman, P., Krauth-Gruber, S., & Ric, F. (2005). Embodiment in attitudes, social perception, and emotion. *Personality and Social Psychology Review*, 9, 184–211.

Nowicki, S., & Duke, M. P. (1994). Individual differences in the nonverbal communication of affect: The Diagnostic Analysis of Nonverbal Accuracy Scale. *Journal of Nonverbal Behavior*, 18, 9–35.

Orbell, J., & Dawes, R. (1993). Social welfare, cooperator's advantage, and the option of not playing the game. *American Sociological Review*, 58, 787–800.

Puccinelli, N. M., & Tickle-Degnen, L. (2004). Knowing too much about others: Moderators of the relationship between eavesdropping and rapport in social interaction. *Journal of Nonverbal Behavior*, 28, 223–243.

Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., & Archer, R. (1979). *Sensitivity to nonverbal communication: The PONS Test*. Baltimore: Johns Hopkins University Press.

Rosip, J. C., & Hall, J. A. (2004). Knowledge of nonverbal cues, gender, and nonverbal decoding accuracy. *Journal of Nonverbal Behavior*, 28, 267–286.

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

Russell, J. A. (1994). Is there universal recognition of emotion from facial expressions? A review of the cross-cultural studies. *Psychological Bulletin*, 115, 102–141.

Scherer, K. R., & Oshinsky, J. S. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion*, 1, 331–346.

Scherer, K. R., & Scherer, U. (2011). Assessing the Ability to Recognize Facial and Vocal Expressions of Emotion: Construction and Validation of the Emotion Recognition Index. *Journal of Nonverbal Behavior*, 35, 305–326.

Schlegel, K., Grandjean, D., & Scherer, K. R. (2012). Emotion recognition: Unidimensional ability or a set of modality- and emotion-specific skills? *Personality and Individual Differences*, 53, 16–21.

Schlegel, K., Grandjean, D., & Scherer, K. R. (2014). Introducing the Geneva Emotion Recognition Test: An example of Rasch-based test development. *Psychological Assessment*, 26, 666–672.

Sternglanz, R., & DePaulo, B. (2004). Reading nonverbal cues to emotions: The advantages and liabilities of relationship closeness. *Journal of Nonverbal Behavior*, 28, 245–266.

Tickle-Degnen, L. (1998). Working well with others: The prediction of students' clinical performance. *The American Journal of Occupational Therapy*, 52, 133–142.

Van Overwalle, F. (2009). Social cognition and the brain: a meta-analysis. *Human Brain Mapping*, 30, 829–858.

Zaki, J., & Ochsner, K. (2011). Reintegrating the study of accuracy into social cognition research. *Psychological Inquiry*, 22, 159–182.

Zebrowitz, L. A. (2001). Groping for the elephant of interpersonal sensitivity. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement* (pp. 333–350). Mahwah, NJ: Erlbaum.

19 What we know and the future of interpersonal accuracy research

Nora A. Murphy

Abstract
This chapter reviews emerging themes and commonalities woven throughout the preceding chapters. What do we know about interpersonal accuracy (IPA)? We know that people can be interpersonally accurate and that there are many measures of IPA. We know that situational settings and interpersonal context matter in IPA, and people can be trained to improve their IPA. But questions remain. How accurate is accurate? What is the role of motivation in IPA? How do we distinguish between target and perceiver effects in IPA? And where exactly do we stand in terms of IPA theory? I conclude with several suggestions about possible future directions for IPA researchers, including further research on the social outcomes and mechanisms of IPA. Ultimately, I remain enthusiastic about how far we have come in understanding IPA and excited about future IPA endeavors.

When first approached to write this chapter, I immediately recalled a chapter written by Leslie Zebrowitz for Hall and Bernieri's (2001) book on methodologies for interpersonal accuracy (IPA).[1] Apparently I was not alone in my recollection as Zebrowitz's chapter seems to be well known (and cited) among the chapters in this book. To recap, in that chapter, Zebrowitz (2001) made the analogy of six blind men describing different parts of an elephant's anatomy. While all the men were feeling the same beast, you would not know it from their widely varied descriptions. Zebrowitz contended that unless IPA researchers articulate an overarching theory to understanding IPA, we are destined to be the blind men groping the elephant; we may be able to

[1] The Hall and Bernieri (2001) book used the term "interpersonal sensitivity" and focused specifically on "decoding" accuracy (i.e., accuracy of a perceiver making a judgment about other targets). Thus, their definition of interpersonal sensitivity coincides with the definition of IPA used here. However, the terms interpersonal sensitivity and interpersonal accuracy might have different meanings depending on the author and usage (Hodges & Wise, Chapter 17). For example, sensitivity could refer to measurement that includes both judge perception and target expressivity (Snodgrass, 2001).

correctly describe discrete components but we will not fully comprehend the whole animal. Hodges and Wise (Chapter 17) also recalled Zebrowitz's forewarning in their chapter and concluded that little has resulted from her message. Boone and Schlegel (Chapter 18) also remembered Zebrowitz's chapter and believed we have made progress toward seeing the complete picture of the IPA elephant. After reading the chapters of this book, I was left feeling quite heartened about our progress in the field of IPA. It is true that an overarching IPA theory remains elusive, a point to which I return later, but the research described here within left me optimistic about future directions and possibilities available to those of us who study IPA. In this chapter, I will review some of the major themes and commonalities, as well as the areas where the research is ambiguous or conflicted, which emerged in reading through the previous chapters. Subsequently, I will indicate areas where there remain gaps or divergences in the study of IPA. The chapter concludes with possible future directions, answering the question, "If I were to decide the future of IPA research over the next 50 years, what would it look like?"

What do we know about interpersonal accuracy?

People can do it. We know people can be interpersonally accurate. People can accurately detect so many traits and states of others that to list the domains in which accuracy was achieved would be tedious and excessive. Limiting ourselves to the domains mentioned in this book, we know that people can accurately recognize all sorts of emotional expressions from angry and sad to guarded and proud (Bänziger, Chapter 2; Ekman & Friesen, 1971; Schlegel, Grandjean, & Scherer, 2014; Tracy & Robins, 2008b). People can recognize many personality traits including assertiveness, extraversion, intelligence, and masculinity/femininity, among others (Back & Nestler, Chapter 5; Funder, 2012; Lippa & Dietz, 2000; Murphy, 2007; Schmid Mast, Hall, Murphy, & Colvin, 2003). People can accurately read others' thoughts and feelings, and detect attributes such as sexual orientation, political affiliation, and status (Alaei & Rule, Chapter 6; Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001; Ickes, Chapter 3; Rule & Ambady, 2008; Samochowiec, Wanke, & Fiedler, 2010). People are accurate in their judgments of where groups stand on their attitudes, especially relative to other groups (West, Chapter 7). People are even meta-accurate in knowing how others see them in terms of reputation, status, and personality (Carlson & Barranti, Chapter 8) and the list goes on and on.

Lest we get too optimistic here, to be clear, people are not always accurate. Historically, there are many examples of erroneous person perception and identification of biases and errors that may lead to inaccurate judgments, such as the fundamental attribution error and correspondence bias (Gawronski, 2004; Ross, 1977). We know that deception detection is notoriously not too great (Bond & DePaulo, 2006; Burgoon & Dunbar, Chapter 4). Judging less expressive traits such as conscientiousness is not always accurate (e.g., Borkenau, Mauer, Riemann, Spinath, & Angleitner, 2004). Moreover, people frequently confuse anger with fear, disgust, or contempt expressions (e.g., Du & Martinez, 2011; Wehrle, Kaiser, Schmidt, & Scherer, 2000). So yes, people can be interpersonally accurate, but not always, and often not very well (see "How accurate is accurate?" section below).

There are many measures of interpersonal accuracy. We also know that there are all sorts of standard and nonstandard tests to measure IPA. Standard tests are established measures with published validity and psychometric properties whereas nonstandard tests are typically "one-off" measures developed specifically by the researcher for a particular study (Hall, Andrzejewski, Murphy, Schmid Mast, & Feinstein, 2008). As described in previous chapters (Bänziger, Chapter 2; Boone & Schlegel, Chapter 18; Ruben, Chapter 14), there are many standard IPA measures of discrete emotion recognition, such as the Japanese and Caucasian Brief Affect Recognition Test (JACBART; Matsumoto et al., 2000), the Diagnostic Analysis of Nonverbal Behavior (DANVA2; Nowicki & Carton, 1993), Emotion Recognition Index (ERI; Scherer & Scherer, 2011), and the Geneva Emotion Recognition Test (GERT; Schlegel et al., 2014).

Could I respectfully suggest that we do not need any new emotion recognition accuracy (ERA) measures? Boone and Schlegel (Chapter 18) demonstrate that ERA measures correlate among each other with a magnitude of about $r = .30$, on average. While not an overwhelmingly strong relationship, this effect size was one of the strongest relationships among various correlations within IPA domains, which suggests that these measures are assessing the same underlying construct (namely, ERA). Given that several standard ERA measures demonstrate validity and strong psychometric properties, it seems reasonable to suggest that those measuring ERA should prudently consider using existing standard measures and certainly avoid the construction of a nonstandard "one-off" measure, unless there is some specific research-related reason for doing so. ERA researchers have choices in which measure to use and without further compelling evidence that a new measure needs to be developed or used, I urge ERA researchers to employ existing, preferably standard, ERA measures. I am also inclined to recommend that dynamic stimuli be given preference over static images for ecological validity reasons but concur with Bänziger's reasoning that in some cases, ecological validity may need to be sacrificed if a researcher intends to study emotional expressions in isolation, outside of any other interpersonal information (e.g., situational context, verbal information, etc.).

Beyond recognizing emotions, other standard IPA measures more broadly construe IPA as accurate evaluations of interpersonal situations or states, such as the Profile of Nonverbal Sensitivity (PONS; Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979), the Interpersonal Perception Task (IPT; Costanzo & Archer, 1989), Reading the Mind in the Eyes test (Baron-Cohen et al., 2001), and the standard stimulus paradigm of empathic accuracy (Ickes, Chapter 3). The newer standard measures include the Patient Emotion Cue Test (PECT; Blanch-Hartigan, 2011) and the Test of Accurate Perception of Patients' Affect (TAPPA; Hall et al., 2014), both designed to specifically investigate IPA in clinical settings. One advantage to using standard measures is possible comparisons between studies that use the same measure. Ickes (Chapter 3) notes that the advantage of the standard stimulus paradigm in empathic accuracy allows for comparisons between perceivers and further investigation into matters such as cross-stimulus consistency. It is fairly striking that there are no standardized measures of lie detection, given the interest and obvious relevance and utility of such a test. In addition, as noted in several chapters, there is no standard measure of personality judgments. This may be due in part to the wide array of personality traits, but why not have a standard measure of perceiving the Big Five? There is a research idea for the taking.

And then there are the many, many nonstandardized IPA tests designed to assess anything from accurate personality perception to doctors' accurate inferences about patients (e.g., Murphy, Hall, & Colvin, 2003; Ruben, Chapter 14). In tallying various IPA measures, Hall et al. (2008) found 95 of 109 IPA studies involved nonstandard formats, which certainly indicate the popularity of nonstandard measures. Hodges and Wise (Chapter 17) note that researchers tend to pick one measurement paradigm and stick with it. As a researcher who frequently employs nonstandard measures to assess IPA (e.g., Murphy, 2007; Murphy et al., 2003; Murphy, Lehrfeld, & Isaacowitz, 2010), I understand the rationale behind a nonstandard measure. Typically, there simply is no standard measure of the construct of interest. For example, in studying whether people can accurately assess intelligence in others, I was not interested in general IPA skills but the specific ability to assess intelligence. Thus, the creation of a nonstandard measure was the most reasonable option. Perhaps in an ideal universe, researchers could devote time and effort to the development and validation of a measure for their specific construct of interest, but the reality of research often negates that possibility, particularly if the measure may only be used once or twice.

Nevertheless, the distinction between a standard and nonstandard test is meaningful on several levels. The use of nonstandard measures certainly introduces variance into IPA measurement. There is evidence to suggest that standard measures yield higher accuracy rates than nonstandard measures (Hall et al., 2008). Nonstandard measures yield results that are usually not directly comparable across studies (Bänziger, Chapter 2), unless authors convert scores into interpretable metrics, a point I elaborate upon later in this chapter. Griffiths and Ashwin (Chapter 9) noted that the use of nonstandard IPA measures hampered progress in understanding the relationship between emotion recognition and schizophrenia, as methodological limitations of early ERA (nonstandard) measures clouded whether ERA deficits were truly characteristic of schizophrenia. Boone and Schlegel (Chapter 18) convincingly argue that nonstandard measures might explain low correlations among IPA measures, given that nonstandard measures may not be validated and authors tend not to report psychometric properties of such measures. On the other hand, nonstandard tests may provide higher ecological validity or allow for specificity in studying a specific construct (Ruben, Chapter 14).

Boone and Schlegel (Chapter 18) provide several concrete suggestions regarding future test development that would involve consideration of standard or nonstandard measures. Before creating a test, the test developer should have a clear definition of what construct the measure is intended to assess, and clear rationale for the selection of stimuli, number of response options, and accuracy criteria. Those who use nonstandard measures should report the internal consistency of their measure, and perhaps consider dropping items that do not correlate well with other items. I also would add that researchers present accuracy findings into translatable metrics (e.g., percent accuracy) that allow comparison of accuracy rates across IPA constructs and measures.

Context matters and training works. If context is broadly construed as situational variables, including the states and traits of targets or perceivers, then we know that many such variables are associated with IPA (e.g., prosociality, empathy, gender, intelligence, and mood) (Hall, Andrzejewski, & Yopchick, 2009; Hodges & Wise, Chapter 17; Murphy & Hall, 2011). And people can achieve accuracy in all sorts of contexts, from a doctor's office to a job interview. In a cultural context, cross-cultural accuracy in emotion recognition is generally considered universal, though not without controversy (Luckman & Elfenbein, Chapter 16). Yet, these are fairly broad operationalizations of context. If we narrow context to specific elements, then accuracy rates clearly do vary depending on the context.

If channel of communication is considered a context, then clearly context matters (Murphy, 2012). Emotion recognition is affected by the channel of communication, as well as the specific emotion being conveyed. For example, "survival" emotions such as anger and disgust were best recognized from the face, but social status emotions such as pride and embarrassment were best recognized from the body (App, McIntosh, Reed, & Hertenstein, 2011). In a variety of studies and paradigms, IPA tends to increase when cues are available through multiple communication channels, such as audiovisual versus visual alone (e.g., Murphy et al., 2003; Wallbott & Scherer, 1986). Depending on the specific IPA construct being measured, verbal information may be particularly salient (Ickes, Chapter 3; Hall & Schmid Mast, 2007). Other elements of context include situational context such as the background of target stimuli or the setting in which a social interaction takes place (Hess & Hareli, 2015). For example, accuracy was higher when targets were engaged in a free conversation in comparison to targets in a structured interaction and when targets were interacting in groups of three (Letzring, 2008; Letzring, Wells, & Funder, 2006). Thus, context undoubtedly plays a role in IPA.

We also know that training can increase IPA. Several meta-analyses examining training and improved deception detection demonstrate small-to-medium effects (Driskell, 2012; Frank & Feeley, 2003; Hauch, Sporer, Michael, & Meissner, 2014). As summarized by Blanch-Hartigan, Andrzejewski, and Hill (Chapter 12), numerous training studies have improved IPA in clinical populations such as individuals with schizophrenia, autism, and learning disabilities. Blanch-Hartigan, Andrzejewski, and Hill's own meta-analysis (2012) on training to improve IPA in nonclinical populations showed similar results, whereby training was effective. Significant moderators included stronger improvement when the training involved both practice and feedback and when it was delivered individually in small groups (<10). Another significant moderator was the IPA domain where comprehension perception (accuracy in perceiving others' comprehension) showed the strongest effect (though this effect was based on four studies), followed by empathic accuracy (i.e., accurately inferring the thoughts and feelings of others; Ickes, Chapter 3), and deception detection. The domain of "other" person perception accuracy (e.g., judging another's status, interpersonal rapport, personality, etc.) showed the smallest effect.

The moderator findings in the Blanch-Hartigan et al. (2012) meta-analysis seem to conceptually map onto the domains identified in Boone and Schlegel's meta-analysis (Chapter 18), which investigated relationships between IPA measures. Boone and Schlegel conclude that IPA measures of emotion, deception, and situational affect tend to "hang together" but IPA measures of personality and social affordances (i.e., social attributes, such as sexual orientation) are rather disparate. Boone and Schlegel's findings provide strong evidence that IPA may not be one global skill that involves a set of distinct yet correlated abilities. Perhaps IPA measures of personality are too disparate or the domain of "personality judgment" is too broad, which might explain why Blanch-Hartigan et al. found the weakest effects of training and Boone and Schlegel found low correlations between personality IPA measures with IPA measures in other domains.

We need more research. Perhaps the most consistent theme emerging across the collected chapters was that we need more research. I concur with the authors on this uncontroversial point. Below I elaborate on areas in which there remains ambiguity in IPA research and expound on possible future directions.

What *don't* we know about interpersonal accuracy? (or where is there ambiguity in IPA research?)

How accurate is accurate? Though people may achieve accuracy in person perception, reports of such results are often unclear as to what "accuracy" actually means. That is, how accurate is accurate? Researchers tend to use the criterion of "better-than-chance" results but chance levels are dependent on the test measure and score criteria (Hall et al., 2008). Without appropriate conversion of findings into an understandable metric such as percent accuracy, interpreting an accuracy score of $r = .28$ as better than chance does little more than simply assure us our results are not a fluke. Those who translate their findings into a percent accuracy metric offer some clarity, provided that the chance level specific to the finding also is reported. With such information, better comparisons between accuracy measures and domains are possible and get us closer to answering the "how accurate is accurate" question.

Just as effect sizes provide practical information as to the size of an effect, understandable metrics to decipher accuracy rates should be reported. I encourage IPA researchers to present their findings in understandable terms that allow a reader to easily comprehend accuracy rates, regardless of the number of stimuli or response options within a particular IPA test. Hall et al. (2008) presented examples of how IPA scores can easily be converted to a common metric using the proportion index (*pi*; Rosenthal & Rubin, 1989) or the binominal effect size display (BESD; Rosenthal & Rubin, 1982). These conversions, which are relatively simple to calculate, essentially account for various scoring techniques, whether a correlational approach or a proportion-correct approach. With common metrics of accuracy rates, we can compare accuracy rates between tests and domains and consider larger implications of accuracy rates, such as whether a given accuracy rate is impressive or disconcerting. For example, Ickes (Chapter 3) notes that empathic accuracy chance rates are 5% and empathic accuracy rates generally are in the 20% range (between strangers) with notable "room for improvement to the theoretical maximum of 100% accuracy" (p. 55).

What role does motivation play in IPA? The motivation question lingers. Does motivation increase IPA? Some say yes, others say no. West (Chapter 7) proposes that motivation is a key element, either directly or indirectly related to the accurate assessment of group-based attitudes. Ickes strongly advocates that motivation matters a great deal, and his empathic accuracy model includes motivation as a central component (Ickes & Simpson, 1997, 2001). And Isaacowitz, Vicaria, and Murray (Chapter 10) detail proposed theoretical perspectives that explain declines in IPA with age as coinciding with changes in motivational goals.

By contrast, a number of other studies detail either no or even detrimental effects of motivation on IPA (e.g., Forrest & Feldman, 2000; Hall et al., 2008; Hall, Gunnery, & Horgan, Chapter 15; Klein & Hodges, 2001). Schmid (Chapter 11) summarizes the inconsistent findings regarding motivation and IPA (see also Hodges & Wise, Chapter 17). Schmid's summary highlights how motivation, much like IPA itself, is a broad construct, which might explain the inconsistent findings. That is, depending on the operationalization of motivation (e.g., gender-role motives, affiliative motives, and monetary incentives), motivation may or may not be related to IPA. Similarly, if IPA measures are not all measuring one underlying construct, but instead represent a correlated set of skills, the disparate findings regarding motivation and IPA perhaps make more sense. The studies that find motivation relates to accuracy may be measuring a different IPA skill than studies that find no relationship between IPA and accuracy. So trying to answer the question of the role of motivation in IPA leads to the unsatisfactory conclusion that "it depends." In the end, I concur with Schmid's assessment that there are clearly moderators at work. Ickes' conclusion also resonated that "in people's everyday lives, their motivation to be accurate or not is far more complicated and variable, stemming both from people's long-standing motives and from the more transient, situationally induced motives that can quickly appear and then, just as quickly, vanish" (p. 62).

Making the distinction between expresser and perceiver effects. The measurement of IPA, whether standard or nonstandard, is typically conceptualized from a trait perspective, as a measurable construct within a given individual. Indeed, almost all of the aforementioned standard and nonstandard measures are designed under this pretext.[2] Perhaps this is just a terminology issue such that "perception" is specifically about perceiver abilities, as opposed to "sensitivity" which could involve both perception and expressivity (Snodgrass, 2001). Indeed, target expressivity contributed a large portion of variance to overall interpersonal sensitivity scores (Snodgrass, Hecht, & Ploutz-Snyder, 1998). A variety of studies indicate that the expressivity of a target is a factor worth considering in the measurement of IPA. Letzring (2008) found that perceivers' accuracy related to the number of targets with high IPA (i.e., "good judges" of personality) in the interaction. She concluded that good judges can elicit relevant information from the target, also suggesting that the expressivity of the target matters. A motivated deceiver (target) increased the likelihood of accurate deception detection by a perceiver (when the perceiver has access to nonverbal channels) (Burgoon & Dunbar, Chapter 4; DePaulo & Kirkendol, 1989). Perceiver variance tends to be small in comparison to target variance in empathic accuracy paradigms (Ickes et al., 2000). Other studies demonstrate higher IPA may be due at least in part to idiosyncratic expressive styles and expressivity of targets (Hall, Rosip, Smith LeBeau, Horgan, & Carter, 2006; Wallbott & Scherer, 1986).

If we do not account for the expressivity of a target, is the interpersonal missing from our IPA research? We have IPA measures and "good judge" research but perhaps we also need more "good target" research. Ruben (Chapter 14) indicates that measuring the expressivity of a target is a laborious process, which might explain why target expressivity is typically not addressed by IPA researchers. (However, if the measurement of IPA involves more than one target, then the target expressivity may be less of a concern, as levels of expressivity among the targets presumably vary.) More sophisticated statistical models do exist that may help disentangle

[2] The original version of the Diagnostic Analysis of Nonverbal Ability (DANVA; Nowicki & Duke, 1994) measured individual differences in both emotion *expressivity* (how well individuals expressed various discrete emotions) and emotion *receptivity* (how well individuals recognize the emotions of others, i.e., emotion recognition accuracy). However, the revised DANVA2 only contains receptive measures of accurate emotion recognition (Nowicki & Carton, 1993; Nowicki & Duke, 2001).

perception and expressivity. For example, Biesanz's Social Accuracy Model (SAM) of interpersonal perception (Biensanz, 2010) allows for the measurement of both perceiver and target effects in IPA measurement (i.e., perceptive accuracy and expressive accuracy), among other components in the accuracy paradigm. This represents a progression in IPA measurement, though SAM requires more sophisticated statistical modeling than do proportion or correlational accuracy scoring approaches.[3] So far, SAM has primarily been applied in personality IPA judgments, though the application of this model into other IPA domains appears feasible.

The struggle of theory in IPA. As a whole, there was relatively little discussion within the chapters about possible theories in IPA. Isaacowitz et al. (Chapter 10) expanded on developmental theoretical perspectives as possible explanations to age-related changes in IPA, though the findings are relatively mixed in determining one overarching explanation in the lifespan domain. Back and Nestler (Chapter 5) elaborated on Brunswik's lens model (Brunswik, 1956) and Funder's Realistic Accuracy Model (RAM; Funder, 1995) as context for understanding accurate personality perception. Bänziger (Chapter 2) explained various theoretical explanations for accurate emotion recognition, though such theories are solidly linked to the functions of emotions themselves. Likewise, Interpersonal Deception Theory (IDT; Buller & Burgoon, 1994; Burgoon & Buller, 2015) is a thorough explanation of how deception detection unfolds (Burgoon & Dunbar, Chapter 4). Yet again, this theory is specifically about a particular IPA domain and not proposed as an account for IPA as a whole.

There are many existing theories in which IPA research could be placed. As I have summarized elsewhere (Murphy, 2012), many social cognitive theories of person perception exist beyond the aforementioned Brunswik's lens model (1956) and RAM (Funder, 1995). Other relevant theories include the ecological theory of social perception (based on Gibson's theories of object perception [Gibson, 1979]) (McArthur & Baron, 1983; Zebrowitz & Collins, 1997), Patterson's (1995) parallel process model, Kenny's Social Relations and PERSON models (Kenny, 2004; Kenny & LaVoie, 1984), the Social Accuracy Model (SAM; Biensanz, 2010), and the Self–Other Asymmetry Model (SOKA; Vazire, 2010).

Yet, even these preceding theories take relatively narrow views, usually limited to the specific construct or domain (e.g., accurately perceiving

[3] Kenny's Social Relations Model (SRM; Kenny, 1994) also accounts for perceiver, target, and relationship (dyad) effects. However, SRM primarily concerns agreement or consensus among judges, rather than accuracy by an external criterion (Kenny, 2002).

personality, emotion, etc.). At a very broad level, an evolutionary perspective would maintain IPA serves an adaptive function to allow individuals to function effectively in social interactions (Bänziger, Chapter 2; Ickes, Chapter 3; Schaller, 2008). Indeed, an evolutionary explanation is one proposed explanation for gender differences in IPA (see Hall et al., Chapter 15). And some have gone as far as proposing that accurate interpersonal judgments of others evolved as a function of survival and reproduction (Haselton & Funder, 2006). Another grander theory of relevance is embodiment theory; just as the embodiment theory has been applied to cognition and emotion, perhaps it is also applicable to accurate person perception (Winkielman, Niedenthal, Wielgosz, Eelen, & Kavanagh, 2015) (for further description see Bänziger, Chapter 2; and Schmid, Chapter 11).

Yet, the question remains as to whether we need one specific, yet unifying, theory of IPA. Boone and Schlegel conclude from their meta-analysis on relationships between IPA measures that IPA is an expansive construct involving a group of loosely related skills conceptually linked through person perception processes. If so, then it may be unrealistic to posit one unifying theory under which all IPA research operates. Even if one argues against one unifying IPA theory, I would urge researchers to specifically identify the theory or theories under which their research endeavors fall.

Other areas of ambiguity. For the sake of brevity, I will only mention a few other areas where more research is needed to disentangle IPA ambiguity. A few authors mentioned the distinction between IPA judgments made by observers outside an interaction, as compared to in vivo judgments, where perceivers may also serve as targets (as in a round-robin design). The most ecologically valid approach would be in vivo judgments, as this is typically how judgments are formed in everyday life. Likewise, dynamic presentation of stimuli (e.g., video) is more realistic to everyday judgments yet a substantial portion of emotion recognition research uses static images. These distinctions (perceiver orientation, dynamic vs. static stimuli) are relevant to the larger questions regarding definitions and levels of accuracy in IPA.

Another perplexing IPA area is the verticality issue of power, dominance, and status, which remains complicated, though that should not be an impediment to further study (Hall, Schmid Mast, & Latu, 2015; Schmid Mast & Latu, Chapter 13). Finally, the contributions of verbal and nonverbal components to IPA need further examination. Ickes (Chapter 3) provides some guidance here by presenting analyses detailing the percentage of verbal, paralinguistic, and nonverbal cues to empathic accuracy. But more research into verbal and nonverbal contributions to IPA in other domains is needed.

The future of interpersonal accuracy research

Predictions of the future are notoriously unreliable. Even estimates about the relatively near future can be wildly inaccurate. For instance, I recently read that in 2007, Microsoft CEO Steve Ballmer estimated that there was no chance the iPhone would get a significant market share (Lieberman, 2007).[4] Psychological research on affective forecasting repeatedly demonstrates how poorly we predict our own feelings and behaviors (e.g., Balcetis & Dunning, 2013; Meyvis, Ratner, & Levav, 2010; Wilson & Gilbert, 2003). Thus, I will not engage the futile endeavor of predicting the future in regard to IPA research. Instead, allow me to ponder a fantasy world, where I speculate on possible fruitful avenues of future IPA research.

Social outcomes and social settings of IPA. Traditionally, IPA investigations tend to be examinations of IPA correlates. In many cases, a researcher will measure IPA and other possibly related constructs such as affective states or personality traits. This investigation of IPA correlates has its place, particularly when a researcher is demonstrating the validity of an IPA measure. The original PONS monograph summarized results from more than 133 samples and most analyses involved measuring external correlates (Hall, 2001; Rosenthal et al., 1979). Moreover, the research examining the correlates of IPA established the relevance of IPA to social interaction and social functioning. Through such research, we know that those with higher IPA tend to be smarter, less depressed, and more socially competent, empathic, tolerant, conscientious, in addition to a host of other miscellaneous positive traits (Hall et al., 2009; Murphy & Hall, 2011; Nowicki & Carton, 1993; Nowicki & Mitchell, 1998). On the other end of the spectrum, numerous psychopathologies are associated with lower IPA (Griffiths & Ashwin, Chapter 9).

While the research on IPA correlates is necessary, as the field develops, I would prefer to learn more about social outcomes related to IPA. Research investigating the social outcomes of IPA, particularly outcomes that are not self-reported by the perceiver, would give us a more complete picture of IPA. Ruben (Chapter 14) detailed various studies where high IPA in clinicians was associated with important clinical variables such as higher patient satisfaction, and better patient appointment keeping.

[4] As of 2015, over 40% of smart phone owners in the United States used an iPhone (Hahn, 2015).

Higher IPA was associated with better supervisor performance ratings for female managers (Byron, 2008). In sales settings, higher IPA in salespeople related to higher sales rates, salary increases, and customers' ratings of service quality (Byron, Terranova, & Nowicki, 2007; Puccinelli, Andrzejewski, Markos, Noga, & Motyka, 2013). As such research illustrates, the socially relevant and positive aspects of IPA are associated with both intra-individual (e.g., traits, affective states) and interpersonal outcomes.

Likewise, I would encourage more IPA research in applied settings. There are several established lines of research investigating IPA in applied settings such as the aforementioned medical and salesperson research, as well as the IPA workplace and leadership research (Ruben, Chapter 14; Schmid Mast & Latu, Chapter 13). Legal and law enforcement domains are other domains with systematic research into IPA, particularly regarding lie detection (Burgoon & Dunbar, Chapter 4). And while smattering of studies have delved into other IPA-relevant domains, such as psychotherapy (Bachrach, Luborsky, & Mechanick, 1974; Kleiman & Rule, 2013; Livingston, 1981), parenting (Otani, Suzuki, Shibuya, Matsumoto, & Kamata, 2009; Suveg, Jacob, & Payne, 2010), education (Bernieri, 1991), and close relationships (Noller, 2006), further programmatic research into such IPA-relevant domains is certainly pertinent and appealing. For example, I would be interested in knowing whether psychotherapists who are high in IPA have more satisfied clients or better success in treating their clients; or if parents with high IPA have better adjusted children; or whether high IPA in teachers results in better student learning outcomes.[5] Given that higher IPA is (typically) associated with positive social functioning and IPA training works, continuing to document IPA-associated social outcomes and relevant domains may even increase general psychological functioning.

IPA processes and mechanisms. We need a better understanding of IPA processes. How does IPA unfold? What are the possible processes or mechanisms behind IPA? Some researchers propose that IPA is an automatic and nonconscious process (Alaei & Rule, Chapter 6). As detailed by Schmid (Chapter 11), researchers have used cognitive load paradigms to test the automaticity of IPA. If IPA is automatic, then participants under cognitive load would have higher accuracy compared to participants in no-load or deliberation conditions, as the

[5] It is worth noting that that Robert Rosenthal, who may be considered the forefather to IPA research, had early work regarding expectancy effects in the classroom (Rosenthal, Baratz, & Hall, 1974; Rosenthal & Jacobson, 1968). Later, his research grew well beyond classroom and educational settings. Yet today, relatively little IPA research exists documenting the effects of teacher or student IPA on learning outcomes.

cognitive load would presumably prevent participants from deliberating, forcing them to rely on automatic processing. Yet, findings are mixed, with various studies showing increased or decreased accuracy, or even no effect, under cognitive load (e.g., Ambady & Gray, 2002; Patterson & Stockbridge, 1998; Phillips, Channon, Tunstall, Hedenstrom, & Lyons, 2008; Tracy & Robins, 2008a). While there is robust research on processes associated with impression formation and person perception (e.g., Todorov & Mende-Siedlecki, 2014), more work could be done to expand and apply those processes to accurate person perception and IPA as a whole. In recent work, perceptual fluency predicted judgments of individuals from various social groups, whereby salient features that were processed easily led to more positive evaluations than features that were harder to identify or notice and subsequently led to disrupted processing (Lick & Johnson, 2013; Lick, Johnson, & Rule, 2015). Such research illustrates how perceptual systems may contribute to or inhibit accurate person perception.

Correspondingly, the neuroscience and neurobiology of IPA is in its infancy. The amygdala is clearly implicated in emotion recognition (Heberlein & Adolphs, 2007; Lundqvist & Öhman, 2005). Griffiths and Ashwin (Chapter 9) noted research demonstrating different brain regions associated with different domains of interpersonal perception (e.g., Heberlein & Saxe, 2005). With better understanding of the brain regions and neurobiology of IPA, we may get closer to knowing whether IPA is one global construct or a set of interrelated constructs. We are only at the beginning stages of comprehending mechanisms and processes of IPA, and I look forward to more research in these areas.

Concluding remarks

As Blanch-Hartigan et al. (Chapter 12) briefly summarized, person perception research has a long history within the relatively brief annals of psychology. In 1932, Jenness reviewed the then-current state of literature on recognizing emotional expressions and concluded:

> In view of the large number of contradictions in the data and conclusions, not to mention the superficial nature of some of the experiments, one might at first decide that the chances of arriving at a satisfactory solution of any major problem in this field are extremely slight. The reviewer feels, however, that these beginnings, rather than being fruitless, have pointed to the necessity for new and better techniques of research and for more thorough consideration of the fundamental problems involved. *(Jenness, 1932, p. 345)*

Almost 100 years later, I am here to do a similar task in summarizing the current state of affairs in IPA research. And I must make an equivalent verdict. After reading the previous chapters and writing my own, I feel buoyed by all the exciting directions of IPA research, including those already undertaken as well as the ones yet to come. IPA is an exciting topic, perhaps driven by human curiosity in understanding social relationships and our need to belong. Whatever the reasons behind the continuing and growing interest in IPA research, we have come a long way and more illuminating times are sure to come. Let's get to it!

References

Ambady, N., & Gray, H. M. (2002). On being sad and mistaken: Mood effects on the accuracy of thin-slice judgments. *Journal of Personality and Social Psychology*, 83, 947–961.

App, B., McIntosh, D. N., Reed, C. L., & Hertenstein, M. J. (2011). Nonverbal channel use in communication of emotion: How may depend on why. *Emotion*, 11, 603–617.

Bachrach, H., Luborsky, L., & Mechanick, P. (1974). The correspondence between judgements of empathy from brief samples of psychotherapy, supervisors' judgements and sensitivity tests. *British Journal of Medical Psychology*, 47, 337–340.

Balcetis, E., & Dunning, D. (2013). Considering the situation: Why people are better social psychologists than self-psychologists. *Self and Identity*, 12, 1–15.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 42, 241–251.

Bernieri, F. J. (1991). Interpersonal sensitivity in teaching interactions. *Personality and Social Psychology Bulletin*, 17, 98–103.

Biesanz, J. C. (2010). The Social Accuracy Model of interpersonal perception: Assessing individual differences in perceptive and expressive accuracy. *Multivariate Behavioral Research*, 45, 853–885.

Blanch-Hartigan, D. (2011). Measuring providers' verbal and nonverbal emotion recognition ability: Reliability and validity of the Patient Emotion Cue Test (PECT). *Patient Education and Counseling*, 82, 370–376.

Blanch-Hartigan, D., Andrzejewski, S. A., & Hill, K. M. (2012). The effectiveness of training to improve person perception accuracy: A meta-analysis. *Basic and Applied Social Psychology*, 34, 483–498.

Bond, C. F., Jr., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, 10, 214–234.

Borkenau, P., Mauer, N., Riemann, R., Spinath, F. M., & Angleitner, A. (2004). Thin slices of behavior as cues of personality and intelligence. *Journal of Personality and Social Psychology*, 86, 599–614.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Los Angeles: University of California Press.

Buller, D. B., & Burgoon, J. K. (1994). Deception: Strategic and nonstrategic communication. In J. A. Daly & J. M. Wiemann (Eds.), *Strategic interpersonal communication* (pp. 191–224). Hillsdale, NJ: Erlbaum.

Burgoon, J. K. & Buller, D. B. (2015). Interpersonal deception theory: Purposive and interdependent behavior during deceptive interpersonal interactions. In D. O. Braithwaite & P. Schrodt (Eds.), *Engaging theories in interpersonal communication, 2e* (pp. 349–362). Los Angeles, CA: Sage.

Byron, K. (2008). Differential effects of male and female managers' non-verbal emotional skills on employees' ratings. *Journal of Managerial Psychology*, 23, 118–134.

Byron, K., Terranova, S., & Nowicki, S. J. (2007). Nonverbal emotion recognition and salespersons: Linking ability to perceived and actual success. *Journal of Applied Social Psychology*, 37, 2600–2619.

Costanzo, M., & Archer, D. (1989). Interpreting the expressive behavior of others: The Interpersonal Perception Task. *Journal of Nonverbal Behavior*, 13, 225–245.

DePaulo, B. M., & Kirkendol, S. E. (1989). The motivational impairment effect in the communication of deception. In J. C. Yuille (Ed.), *Credibility assessment* (pp. 51–70). New York: Kluwer Academic/Plenum Publishers.

Driskell, J. E. (2012). Effectiveness of deception detection training: A meta-analysis. *Psychology, Crime and Law*, 18, 713–731.

Du, S., & Martinez, A. M. (2011). The resolution of facial expressions of emotion. *Journal of Vision*, 11, 1–13.

Ekman, P., & Friesen, W. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17, 124–129.

Forrest, J. A., & Feldman, R. S. (2000). Detecting deception and judge's involvement: Lower task involvement leads to better lie detection. *Personality and Social Psychology Bulletin*, 26, 118–125.

Frank, M. G., & Feeley, T. H. (2003). To catch a liar: Challenges for research in lie detection training. *Journal of Applied Communication Research*, 31, 58–75.

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, 102, 652–670.

Funder, D. C. (2012). Accurate personality judgment. *Current Directions in Psychological Science*, 21, 177–182.

Gawronski, B. (2004). Theory-based bias correction in dispositional inference: The fundamental attribution error is dead, long live the correspondence bias. *European Review of Social Psychology*, 15, 183–217.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Hahn, J. (2015, March 8). Apple starts 2015 as the top U.S. smartphone maker, Android as the top smartphone OS. *Digital Trends.* Retrieved from www.digitaltrends.com/mobile/apple-starts-2015-as-the-top-u-s-smartphone-maker-android-as-the-top-smartphone-os/#ixzz3VcoYIRx8

Hall, J. A. (2001). The PONS test and the psychometric approach to measuring interpersonal sensitivity. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement* (pp. 143–160). Mahwah, NJ: Erlbaum.

Hall, J. A., Andrzejewski, S. A., Murphy, N. A., Schmid Mast, M., & Feinstein, B. A. (2008). Accuracy of judging others' traits and states: Comparing mean levels across tests. *Journal of Research in Personality*, 42, 1476–1489.

Hall, J. A., Andrzejewski, S. A., & Yopchick, J. E. (2009). Psychosocial correlates of interpersonal sensitivity: A meta-analysis. *Journal of Nonverbal Behavior*, 33, 149–180.

Hall, J. A., & Bernieri, F. J. (2001). *Interpersonal sensitivity: Theory and measurement*. Mahwah, NJ: Lawrence Erlbaum Associates.

Hall, J. A., Rosip, J. C., Smith LeBeau, L., Horgan, T. G., & Carter, J. D. (2006). Attributing the sources of accuracy in unequal-power dyadic communication: Who is better and why? *Journal of Experimental Social Psychology*, 42, 18–27.

Hall, J. A., & Schmid Mast, M. (2007). Sources of accuracy in the empathic accuracy paradigm. *Emotion*, 7, 438–446.

Hall, J. A., Schmid Mast, M., & Latu, I. M. (2015). The vertical dimension of social relations and accurate interpersonal perception: A meta-analysis. *Journal of Nonverbal Behavior*, 39, 131–163.

Hall, J. A., Ship, A. N., Ruben, M. A., Curtin, E. M., Clever, S., Roter, D. L., Smith, C., & Pounds, K. (2014). The Test of Accurate Perception of Patients' Affect (TAPPA): An ecologically valid tool for assessing interpersonal perception accuracy in clinicians. *Patient Education and Counseling*, 94, 218–223.

Haselton, M. G., & Funder, D. C. (2006). The evolution of accuracy and bias in social judgment. In M. Schaller, J. A. Simpson, & D. T. Kenrick (Eds.), *Evolution and social psychology* (pp. 15–37). Madison, CT: Psychosocial Press.

Hauch, V., Sporer, S. L., Michael, S. W., & Meissner, C. A. (2014). Does training improve the detection of deception? A meta-analysis. *Communication Research*. Advance online publication.

Heberlein, A. S., & Adolphs, R. (2007). Neurobiology of emotion recognition: Current evidence for shared substrates. In E. Harmon-Jones & P. Winkielman (Eds.), *Social neuroscience: Integrating biological and psychological explanations of social behavior* (pp. 31–55). New York: Guilford Press.

Heberlein, A. S., & Saxe, R. R. (2005). Dissociation between emotion and personality judgments: Convergent evidence from functional neuroimaging. *NeuroImage*, 28, 770–777.

Hess, U., & Hareli, S. (2015). The role of social context for the interpretation of emotional facial expressions. In M. K. Mandal, & A. Awasthi (Eds.), *Understanding facial expressions in communication* (pp. 119–141). New Delhi: Springer.

Ickes, W., Buysse, A., Pham, H., Rivers, K., Erickson, J. R., Hancock, M., & … Gesn, P. R. (2000). On the difficulty of distinguishing "good" and "poor" perceivers: A social relations analysis of empathic accuracy data. *Personal Relationships*, 7, 219–234.

Ickes, W., & Simpson, J. (1997). Managing empathic accuracy in close relationships. In W. Ickes (Ed.), *Empathic accuracy* (pp. 218–250). New York: Guilford Press.

Ickes, W., & Simpson, J. (2001). Motivational aspects of empathic accuracy. In G. J. O. Fletcher & M. S. Clark (Eds.), *Interpersonal processes: Blackwell handbook in social psychology* (pp. 229–249). Oxford, UK: Blackwell.

Jenness, A. (1932). The recognition of facial expressions of emotion. *Psychological Bulletin*, 29, 324–350.

Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis*. New York: Guilford.

Kenny, D. A. (2002, January 24). Nine basic questions [website]. Retrieved from http://davidakenny.net/ip/laing.htm

Kenny, D. A. (2004). PERSON: A general model of interpersonal perception. *Personality and Social Psychology Review*, 8, 265–280.

Kenny, D. A., & LaVoie, L. (1984). The social relations model. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 142–182). Orlando, FL: Academic Press.

Kleiman, S., & Rule, N. O. (2013). Detecting suicidality from facial appearance. *Social Psychological and Personality Science*, 4, 453–460.

Klein, K. J. K., & Hodges, S. (2001). Gender differences, motivation, and empathic accuracy: When it pays to understand. *Personality and Social Psychology Bulletin*, 27, 720–730.

Letzring, T. D. (2008). The good judge of personality: Characteristics, behaviors, and observer accuracy. *Journal of Research in Personality*, 42, 914–932.

Letzring, T. D., Wells, S. M., & Funder, D. C. (2006). Quantity and quality of available information affect the realistic accuracy of personality judgment. *Journal of Personality and Social Psychology*, 91, 111–123.

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

Lick, D. J., Johnson, K. L., & Rule, N. O. (2015). Disfluent processing of nonverbal cues helps to explain anti-bisexual prejudice. *Journal of Nonverbal Behavior*. Advance online publication.

Lieberman, D. (2007, April 30). CEO Forum: Microsoft's Ballmer having a "great time." *USA Today*. Retrieved from http://usatoday30.usatoday.com/money/companies/management/2007-04-29-ballmer-ceo-forum-usat_N.htm

Lippa, R. A., & Dietz, J. K. (2000). The relation of gender, personality, and intelligence to judges' accuracy in judging strangers' personality from brief video segments. *Journal of Nonverbal Behavior*, 24, 25–43.

Livingston, S. A. (1981). Nonverbal communication tests as predictors of success in psychology and counseling. *Applied Psychological Measurement*, 5, 325–331.

Lundqvist, D., & Öhman, A. (2005). Caught by the evil eye: Nonconscious information processing, emotion, and attention to facial stimuli. In L. F. Barrett, P. M. Niedenthal, & P. Winkielman (Eds.), *Emotion and consciousness* (pp. 97–122). New York: Guilford Press.

Matsumoto, D., LeRoux, J., Wilson-Cohn, C., Raroque, J., Kooken, K., Ekman, P., . . . Goh, A. (2000). A new test to measure emotion recognition ability: Matsumoto and Ekman's Japanese and Caucasian Brief Affect Recognition Test (JACBART). *Journal of Nonverbal Behavior*, 24, 179–209.

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

Meyvis, T., Ratner, R. K., & Levav, J. M. (2010). Why don't we learn to accurately forecast feelings? How misremembering our predictions blinds us to forecasting errors. *Journal of Experimental Psychology: General*, 139, 579–589.

Murphy, N. A. (2007). Appearing smart: The impression management of intelligence, person perception accuracy, and behavior in social interaction. *Personality and Social Psychology Bulletin*, 33, 325–339.

Murphy, N. A. (2012). Nonverbal perception. In S. T. Fiske & C. N. Macrae (Eds.), *Handbook of social cognition* (pp. 196–215). London: Sage.

Murphy, N. A., & Hall, J. A. (2011). Intelligence and interpersonal sensitivity: A meta-analysis. *Intelligence*, 39, 54–63.

Murphy, N. A., Hall, J. A., & Colvin, C. R. (2003). Accurate intelligence assessments in social interactions: Mediators and gender effects. *Journal of Personality*, 71, 465–493.

Murphy, N. A., Lehrfeld, J. M., & Isaacowitz, D. M. (2010). Recognition of posed and spontaneous dynamic smiles in young and older adults. *Psychology and Aging*, 25, 811–821.

Noller, P. (2006). Nonverbal communication in close relationships. In V. Manusov & M. L. Patterson (Eds.), *The Sage handbook of nonverbal communication* (pp. 403–420). Thousand Oaks: Sage Publications, Inc.

Nowicki, S., Jr., & Carton, J. (1993). The measurement of emotional intensity from facial expressions: The DANVA FACES 2. *Journal of Social Psychology*, 133, 749–751.

Nowicki, S., & Duke, M. P. (1994). Individual differences in the nonverbal communication of affect: the Diagnostic Analysis of Nonverbal Accuracy scale. *Journal of Nonverbal Behavior*, 18, 9–34.

Nowicki, S., & Duke, M. P. (2001). Nonverbal receptivity: The Diagnostic Analysis of Nonverbal Accuracy (DANVA). In J. A. Hall & F. J. Bernieri (Eds.) *Interpersonal sensitivity: Theory and measurement* (pp. 183–198). Hillsdale, NJ: Lawrence Erlbaum Associates.

Nowicki, S. J., & Mitchell, J. (1998). Accuracy in identifying affect in child and adult faces and voices and social competence in preschool children. *Genetic, Social, and General Psychology Monographs*, 124, 39–59.

Otani, K., Suzuki, A., Shibuya, N., Matsumoto, Y., & Kamata, M. (2009). Dysfunctional parenting styles increase interpersonal sensitivity in healthy subjects. *Journal of Nervous and Mental Disease*, 197, 938–941.

Patterson, M. L. (1995). A parallel process model of nonverbal communication. *Journal of Nonverbal Behavior*, 19, 3–29.

Patterson, M. L., & Stockbridge, E. (1998). Effects of cognitive demand and judgment strategy on person perception accuracy. *Journal of Nonverbal Behavior*, 22, 253–263.

Phillips, L. H., Channon, S., Tunstall, M., Hedenstrom, A., & Lyons, K. (2008). The role of working memory in decoding emotions. *Emotion*, 8, 184–191.

Puccinelli, N. M., Andrzejewski, S. A., Markos, E., Noga, T., & Motyka, S. (2013). The value of knowing what customers really want: The impact of salesperson ability to read affect on service quality. *Journal of Marketing Management*, 29, 356–373.

Rosenthal, R., Baratz, S. S., & Hall, C. M. (1974). Teacher behavior, teacher expectations, and gains in pupils' rated creativity. *The Journal of Genetic Psychology: Research and Theory on Human Development*, 124, 115–121.

Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., & Archer, D. (1979). *Sensitivity to nonverbal communication: The PONS test*. Baltimore: The Johns Hopkins University Press.

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom: Teacher expectation and pupils' intellectual development*. New York: Holt, Rinehart & Winston.

Rosenthal, R., & Rubin, D. B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology*, 74, 166–169.

Rosenthal, R., & Rubin, D. B. (1989). Effect size estimation for one-sample multiple-choice-type data: Design, analysis, and meta-analysis. *Psychological Bulletin*, 106, 332–337.

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.

Rule, N. O., & Ambady, N. (2008). Brief exposures: Male sexual orientation is accurately perceived at 50 ms. *Journal of Experimental Social Psychology*, 44, 1100–1105.

Samochowiec, J., Wanke, M., & Fiedler, K. (2010) Political ideology at face value. *Psychological and Personality Science*, 1, 206–213.

Schaller, M. (2008). Evolutionary bases of first impressions. In N. Ambady & J. J. Skowronski (Eds.), *First impressions* (pp. 15–34). New York: Guilford Publications.

Scherer, K. R., & Scherer, U. (2011). Assessing the ability to recognize facial and vocal expressions of emotion: Construction and validation of the Emotion Recognition Index (ERI). *Journal of Nonverbal Behavior*, 35, 305–326.

Schlegel, K., Grandjean, D., & Scherer, K. R. (2014). Introducing the Geneva Emotion Recognition Test: An example of Rasch-based test development. *Psychological Assessment*, 26, 666–672.

Schmid Mast, M., & Hall, J. A. (2006). Women's advantage at remembering others' appearance: A systematic look at the why and when of a gender difference. *Personality and Social Psychology Bulletin*, 32, 353–364.

Schmid Mast, M., Hall, J. A., Murphy, N. A., & Colvin, C. R. (2003). Judging assertiveness. *Facta Universitatis*, 2, 731–744.

Snodgrass, S. E. (2001). Correlational method for assessing interpersonal sensitivity within dyadic interaction. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement* (pp. 201–218). Mahwah, NJ: Lawrence Erlbaum Associates.

Snodgrass, S. E., Hecht, M. A., & Ploutz-Snyder, R. (1998). Interpersonal sensitivity: Expressivity or perceptivity? *Journal of Personality and Social Psychology*, 74, 238–249.

Suveg, C., Jacob, M. L., & Payne, M. (2010). Parental interpersonal sensitivity and youth social problems: A mediational role for child emotion dysregulation. *Journal of Child and Family Studies*, 19, 677–686.

Todorov, A., & Mende-Siedlecki, P. (2014). The cognitive and neural basis of impression formation. In K. N. Ochsner & S. M. Kosslyn (Eds.), *The Oxford handbook of cognitive neuroscience, Vol. 2: The cutting edges* (pp. 188–203). New York: Oxford University Press.

Tracy, J. L., & Robins, R. W. (2008a). The automaticity of emotion recognition. *Emotion*, 8, 81–95.

Tracy, J. L., & Robins, R. W. (2008b). The nonverbal expression of pride: Evidence for cross-cultural recognition. *Journal of Personality and Social Psychology*, 94, 516–530.

Vazire, S. (2010). Who knows what about a person? The self-other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology*, 98, 281–300.

Wallbott, H. G., & Scherer, K. R. (1986). Cues and channels in emotion recognition. *Journal of Personality and Social Psychology*, 51, 690–699.

Wehrle, T., Kaiser, S., Schmidt, S. & Scherer, K. R. (2000). Studying the dynamics of emotional expression using synthesized facial muscle movements. *Journal of Personality and Social Psychology*, 78, 105–119.

Wilson, T. D., & Gilbert, D. T. (2003). Affective forecasting. *Advances in Experimental Social Psychology*, 35, 345–411.

Winkielman, P., Niedenthal, P., Wielgosz, J., Eelen, J., & Kavanagh, L. C. (2015). Embodiment of cognition and emotion. In M. Mikulincer, P. R. Shaver, E. Borgida, & J. A. Bargh (Eds.), *APA handbook of personality and social psychology, Volume 1: Attitudes and social cognition* (pp. 151–175). Washington, DC: American Psychological Association.

Zebrowitz, L.A. (2001). Groping for the elephant of interpersonal sensitivity. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement* (pp. 333–350). Hillsdale, NJ: Lawrence Erlbaum Associates.

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.

Index

Note: Some extremely common terms, such as *emotion recognition*, are selectively indexed.