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

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