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

3

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 theoreti-
cally 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 inter-personal 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 research-ers 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 under-stand better why perceivers were accurate (e.g., they used multiple tar-gets, 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