

ARTICLE

Studying Stereotype Accuracy from an Integrative Social-Personality Perspective

Judith A. Hall  | Jin X. Goh

Northeastern University

Correspondence

Judith A. Hall, Dept. of Psychology, 125 NI,
Northeastern University, Boston, MA 02115.
Email: j.hall@northeastern.edu

Abstract

This article proposes an integrative framework for understanding the accuracy and inaccuracy of stereotypes. Specifically, we highlight research issues and traditions from social and personality psychology that do not often intersect, but which can be mutually informative. Within this framework, the social psychologist's interest in the accuracy of group stereotypes is conceptually much like a personality psychologist's interest in the accuracy with which perceivers can identify types of individuals, for example extraverts. Both fields make use, implicitly or explicitly, of personal attributes and behaviors (cues) in assessing accuracy of beliefs about group or individual traits. By using Brunswik's lens model perspective in combination with concepts from signal detection theory, judgments of stereotypes can be discovered to be accurate or inaccurate depending on how perceivers judge or use the cues. In drawing on research traditions and theoretical frameworks from both social and personality psychology, researchers can go beyond an all-or-nothing stance regarding stereotype accuracy to achieve a more nuanced understanding of when, how, and to what extent stereotypes are accurate.

KEYWORDS

Stereotype accuracy, interpersonal accuracy, lens model, signal detection theory

1 | INTRODUCTION

Are stereotypes accurate? This question has generated a wealth of debate and disagreement among psychologists and sociologists. In *The Nature of Prejudice*, Allport (1954/1979) proposed that a stereotype "is an exaggerated belief associated with a category" but that it can hold "a kernel of truth" (pp. 190-191). This foundational statement left room for extreme conclusions: stereotypes are inaccurate versus stereotypes are accurate, at least in their core. The middle ground—of finding out when there is a kernel of truth and how big that kernel is—has sometimes been neglected. Certainly very few researchers would argue that all stereotypes, for all groups, or perhaps even for any group, are completely false or true; however, the possibility that stereotypes might be somewhat, or even quite, accurate has received different degrees of acceptance.

Some take the position that stereotypes are generally inaccurate and if not wholly so, they are largely exaggerated (Bargh & Chartrand, 1999). In contrast, some have argued and/or found that stereotypes can be far more accurate than presumed (for reviews, see Eagly, 1995; Jussim, 2017; Jussim, Crawford, & Rubinstein, 2015; Ryan, 2003). It is important to point out that researchers' ambivalence is not about measuring group stereotypes; it is about measuring actual group differences, especially those relating to racial and ethnic groups (e.g., Campbell, 1967; Jussim, 2017; Jussim et al., 2015; Mackie, 1973; Ryan, 2003). Without documentation of group characteristics, however, it is not possible to conduct research into whether beliefs about them are correct. While no one would doubt there is scientific value in understanding the validity of stereotypes, the problem for researchers comes when contemplating the potential social risks and benefits of such an investigation. Whether the social benefit of disconfirming elements of a negative stereotype would outweigh the negative effect of confirming some of its elements cannot be known *a priori* and would be a difficult question to answer even after data are in hand. However, without the investigation scholars cannot even have the conversation, much less make efforts to alter negative beliefs that are false or remediate the actuality that underlies negative beliefs that are correct.

Having noted social psychologists' ambivalence about studying stereotype accuracy, we can point to another tradition where, interestingly, ideological ambivalence is absent. We refer to the study of interpersonal accuracy (Hall & Bernieri, 2001; Hall, Schmid Mast, & West, 2016). This interdisciplinary field addresses the question of how accurate people's beliefs, perceptions, and judgments of individuals are, sometimes also examining the behaviors or cues that facilitate or impede accuracy. Here the emphasis is on judging individuals rather than entire groups. In this endeavor, personality psychologists have been particularly active, although social psychologists as well as communication scientists also study judgments of individuals in domains such as emotion, power, sexual orientation, religion, and truthfulness. Researchers in this tradition generally argue that people are fairly accurate at judging the states and traits of other people.

We suggest that the distinction between stereotypes about social groups and beliefs about individuals is slight and mainly semantic. This is because asking perceivers to rate individual people on a characteristic is only a finer-grained way of asking them about whole groups of people (extraverts vs. introverts, sad people vs. happy people, high-power people vs. low-power people, etc.). In both cases, the issues are parallel: the beliefs that perceivers hold about the target people's traits and behaviors are the *stereotypes*; how the target people differ in terms of traits or behaviors is the *actuality*; and the extent to which various traits or behaviors believed by perceivers to characterize the target people in fact describe the target people is *stereotype accuracy*. However, despite these similarities between stereotype accuracy (the term used when whole target groups are the focus) and interpersonal accuracy (the term used when individual targets are the focus), these two fields have not been brought together in an integrative conceptual framework. We suggest that research on stereotype accuracy can be reconsidered in a broader theoretical and empirical light. Specifically, we integrate research traditions from interpersonal accuracy more broadly with those regarding stereotyping by using Brunswik's lens model perspective in combination with concepts from signal detection theory. This framework demonstrates that stereotypes (and perception of people generally) can be both accurate and inaccurate.

2 | STEREOTYPE ACCURACY

Studies that actually measure stereotype accuracy generally yield effect sizes of medium to large magnitude (Jussim, 2017; Jussim et al., 2015), with specific examples applying to stereotypes about racial and ethnic groups (Ashton & Esses, 1999; McCauley & Stitt, 1978; Ryan, 1996), gender (Diekmann, Eagly, & Kulesa, 2002; Hall & Carter, 1999; Löckenhoff et al., 2014; Swim, 1994), age (Chan et al., 2012), and people in different occupations (Imada, Fletcher, & Dalessio, 1980). Despite such results, a content analysis of textbooks on prejudice and stereotyping showed that most authors do not discuss or acknowledge the accuracy of stereotypes (Jussim et al., 2015). Diekmann, Eagly, and Johnston (2010) argued that "if stereotypes follow from correspondent inferences from role behavior to personal attributes, they invariably have a degree of accuracy at the group level... [A]ccuracy has proven much more impressive than implied by Allport's acknowledgement" (p. 217).

Stereotypes often fail to fit the *individuals* being evaluated; a good fit would occur only if there is no variation within the group and therefore guessing the base rate would be correct by definition, or if the individual being judged falls at exactly their group's mean. In other words, people use stereotypes when judging an individual from a particular group but that particular individual may not fit the stereotypes of their group, therefore contributing to inaccurate perception. Nonetheless, the interpersonal accuracy tradition argues that perception at the individual, interpersonal level can be surprisingly accurate.

3 | INTERPERSONAL ACCURACY

In the following, we will use the terms *perceiver* (the person perceiving and making judgments) and *target* (the person being perceived and judged). This tradition focuses on the accuracy with which perceivers can judge specific target people on some kind of personal or group characteristic such as those named earlier (Hall et al., 2016).

Researchers in the interpersonal accuracy tradition often celebrate the high accuracy rate that perceivers can obtain in judging others (e.g., Connelly & Ones, 2010; Efenbein & Ambady, 2002). Accuracy rates can be high even when the judgments are based on very limited information such as short video clips or very briefly exposed photographs of people (e.g., Ambady & Rosenthal, 1992; Matsumoto et al., 2000; Tskhay & Rule, 2013). Sometimes the accuracy rate is not very high but it is still above chance, which, considering the brevity and subtlety of the stimuli being judged, is also a reason to celebrate as it indicates the surprising ability of people to form accurate first impressions based on extremely minimal cues. This situation is epitomized by research in judging ambiguous identities that, in the stimuli presented, have no blatant external manifestations; these include sexual orientation, religious affiliation, and political orientation. Accuracy for judging these groups has an overall accuracy of 64.5% when the guessing level would be 50% (Tskhay & Rule, 2013). At any rate, although there are exceptions such as weak accuracy in lie detection (see below), the bottom line in this tradition is very often that perceivers are thought to be accurate at judging the attributes of target people, which logically implies correct *beliefs* about what information is and is not valid diagnostically—in other words, that perceivers hold accurate stereotypes about the characteristics being judged.

Clearly, there are differences between interpersonal accuracy and stereotype accuracy research traditions. In the interpersonal accuracy tradition, perceivers judge actual individuals whom they see or hear (in video clips, for example) on a given attribute (for example, extraversion) rather than judge the entire social category on a set of behaviors or traits. Also, researchers in the interpersonal accuracy tradition are not reluctant to gather more information on targets than simply how perceivers judge them on various traits or characteristics. They might, for example, measure smiling and loud voice and correlate those cues with perceivers' judgments of extraversion. Non-zero correlations indicate a belief (stereotype) about extraverted people. In this paradigm, these stereotypical beliefs about extraverts are measured implicitly because perceivers are not actually asked whether certain cues are associated with the trait in question, but rather their belief is inferred from the correlation between their extraversion judgments and the targets' measured cues. One can, of course, also obtain stereotypes about extraverts explicitly, by asking people to state their beliefs about extraverts as a group.

3.1 | Brunswik's Lens Model

Research on interpersonal accuracy allows for a statement about the degree of accuracy, but even when the accuracy rate is high, researchers are not always able to say *how* accuracy is achieved because not all research analyzes specific behavioral or appearance cues that might be relevant (or not) in the judgment process. Research using concepts from Brunswik's lens model (Brunswik, 1956) allows just that. The most robust tradition of this sort concerns the judgment of personality traits (Back & Nestler, 2016).

The relevant concepts from the lens model are *criterion* (continuing with this example: the targets' actual extraversion, as operationally defined), *cue validities* (how targets' actual extraversion, as operationally defined, is correlated

with measured aspects of the targets' appearance or behavior—together called cues), and *cue utilization* (how perceivers' ratings of the targets' extraversion correlate with the measured cues). Published lens models address a wide range of constructs, including personality traits, psychopathy, emotions, self-esteem, academic achievement, status, pain experience, and rapport (e.g., Bernieri, Gillis, Davis, & Grahe, 1996; Gosling, Ko, Mannarelli, & Morris, 2002; Hirschmüller, Schmukle, Krause, Back, & Egloff, in press; Laukka et al., 2016; Ruben & Hall, 2016; Sahoo & Sahoo, 2017; Schmid Mast & Hall, 2004; ten Brinke et al., 2017).

3.2 | Signal Detection Theory

Signal detection theory has been adopted to study accuracy in social perception (West & Kenny, 2011). There are four possible categories in signal perception (Wickens, 2001), analogous to types of judgment accuracy and inaccuracy in the present context: Hit, Correct rejection, Miss, and False alarm. These four categories can be applied as follows, using the perception of sexual orientation as an example (e.g., Rule, Ambady, & Hallett, 2009; Rule, Ishii, Ambady, Rosen, & Hallett, 2011; Rule, Rosen, Slepian, & Ambady, 2011): A Hit would be correctly identifying a gay person as gay; a Correct rejection would be *not* categorizing a straight person as gay; a Miss would be erroneously categorizing a gay person as straight; and a False alarm would be calling a straight person gay (using the categorization of gay identity as the reference). When cues are added into consideration, in the intervening spot between criterion and judgment, the categories of signal detection can be applied to the question of stereotype accuracy, as explained next.

4 | INTEGRATIVE FRAMEWORK

We have argued that making judgments about another person requires the explicit or implicit application of a belief about them and therefore qualifies as stereotyping, whether it be about their hair, their facial expression, their voice tone, their dress, or any other attribute, feature, or behavior the target person might display or possess. The key element that underlies this proposed integrative framework is the *content of those beliefs*, whether they are called stereotypes or cue utilizations. Returning to the lens model approach, comparison of cue validities and the associated cue utilizations opens the window into improved understanding of stereotype accuracy and inaccuracy. We propose that the lens model can potentially reveal four types of paths from the criterion, through the cues, to the perception. These four paths correspond to the prototypical categories discussed in signal detection theory (Wickens, 2001). We will continue with the example of extraversion, using vocabulary from signal detection theory and showing hypothetical results of a lens model analysis (see Figure 1).

The four possibilities of Hit, Correct rejection, Miss, and False alarm apply (hypothetically) as follows when cues are added to the model. A Hit occurs when a given cue (louder voice) is correlated positively both with actual extraversion

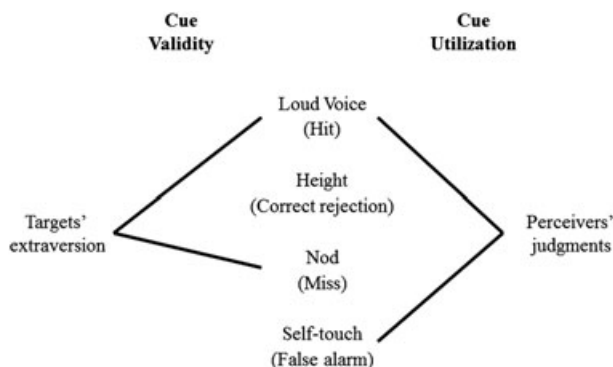


FIGURE 1 An example lens model on the hypothetical attributes of extraversion.

and perceivers' ratings of extraversion, meaning perceivers correctly used that cue in judging extraversion. (A Hit would equivalently occur if the cue were softer voice and its correlations with the criterion and the judgments were both negative.) A Correct rejection occurs when a given cue (in this example, height) is correlated with neither actual nor perceived extraversion. Although often not emphasized by researchers in this field, both Hits and Correct rejections reflect accurate beliefs—about cues that are and are *not* actually associated with extraversion. Inaccuracy is indexed by the other two paths: Miss (when nodding is correlated with actual extraversion but perceivers do not rate nodding targets as more extraverted) and False alarm (when self-touch is not correlated with actual extraversion but perceivers think it is). These four different paths demonstrate that beliefs about extraverted people can be simultaneously accurate and inaccurate because accuracy depends on how perceivers use these cues to guide their judgments.

Funder (1995), in describing his Realistic Accuracy Model of interpersonal perception, added several process components to the prototypical lens model. He proposed that for accurate judgment to occur, a perceiver needs first to be presented with cues that are both relevant to the construct being judged (where the construct is instantiated in a criterion measurement of some kind) and that absence of relevance precludes accuracy. While for Hits these conditions cannot be argued with, Funder's model offers an incomplete picture of accuracy because, as noted above, a cue that is *irrelevant* to the construct can contribute to accuracy in the form of a Correct rejection. For example, if wearing glasses is not relevant to extraversion and perceivers do not think it is, then they are correct with respect to this cue.

Funder (1995) also proposed several moderators of accuracy that could reveal when accuracy will be higher or lower. Individual differences in perceivers' general skill in judging people is one such moderator. An example in the context of stereotype accuracy is Hall and Carter's (1999) study on accuracy of gender stereotypes. Perceivers who scored higher on a test of accurately decoding the meanings of nonverbal cues were more accurate in their beliefs about how men and women differ on psychological variables—suggesting that the accuracy of their beliefs about social groups depended on whether they were accurate observers of people. Indeed, in the same study a tendency to rely on stereotypes was *negatively* correlated with accurate beliefs about gender differences. Thus stereotyping as an individual difference trait actually seemed to hurt one's stereotype accuracy. We do not claim this would always be the case because there are situations where falling back on stereotypes (or base rates) might sometimes improve accuracy (Lewis, Hodges, Laurent, Srivastava, & Biancarosa, 2012). In fact, in some ways of measuring accurate personality judgment, making use of stereotypes (i.e., using normative knowledge) is an integral part of achieving accuracy (Hall et al., in press).

We now offer concrete examples of how concepts from interpersonal accuracy, lens model, and signal detection theory could be integrated to inform our understanding of the accuracy of stereotypes about social groups such as African Americans, women, Jewish people, or Asians. In Figure 2, extraversion is replaced with a specific racial group (Asians) and the cues are hypothetical measured attributes of Asians (see Judd & Park, 1993, on how to measure attributes and stereotypes of social groups). Any correspondence between Asians' actual measured attributes and perceivers' judgments would be a Hit such as, hypothetically, in the case of intelligence (stereotype accuracy); likewise, any attributes that are uncorrelated with being Asian (for example, being curious) and not used by perceivers would be a Correct rejection (still a form of accuracy). On the other hand, if misalignment exists between cue validity (Asians' actual possession of attributes) and cue utilization (perceivers' judgments of Asians), then there is stereotype inaccuracy. If Asians are actually very trustworthy but perceivers do not judge them as such, this is a Miss; if Asians are actually not communal but perceivers judge them as such, this is a False alarm.

Figure 3 illustrates how the proposed framework can be used to understand gender stereotypes based on existing published findings (i.e., the Figure 3 results are not hypothetical). A Hit pathway supports a prevailing assumption that women are generally more communal and interpersonally oriented than men (Costa, Terracciano, & McCrae, 2001) and shows that perceivers hold these stereotypes as well (Löckenhoff et al., 2014). A Correct rejection is represented by the finding that women are not more or less active (as a personality trait) than men, and perceivers do not typically associate such a trait with women (Löckenhoff et al., 2014). A Miss is exemplified by the finding that women perform at equal level to men and sometimes outperform men on quantitative tasks and math tests (Hyde, 2014), yet people generally believe that women are not as mathematically capable as their male counterparts (Cheryan, Ziegler, Montoya, & Jiang, 2017).

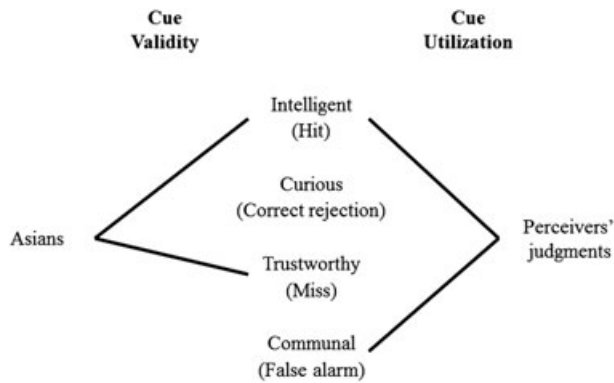


FIGURE 2 An example lens model on the hypothetical attributes of Asians.

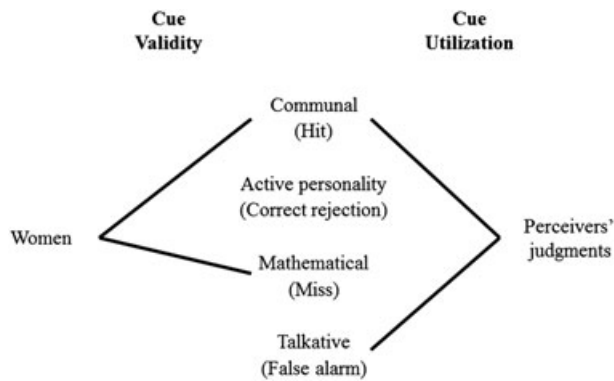


FIGURE 3 An example lens model on the attributes of women based on existing literature.

Finally, a False alarm pathway is evident when evidence shows that women are not necessarily more talkative than men, yet people believe they are (Mehl, Vazire, Ramírez-Esparza, Slatcher, & Pennebaker, 2007).

All of these paths provide insight into the relative accuracy and inaccuracy of different beliefs (stereotypes). Researchers should not focus preferentially on one type of path over another, as one would do if presuming inaccuracy or presuming accuracy. And even when there is no starting presumption, it is easy to attend to Hits, which stand out because both the cue validity correlation and the cue utilization correlation are consistent and non-zero, while paying little attention to Correct rejections, where neither path is significant or at any rate negligible in magnitude. Furthermore, Hits are easily prioritized over Misses and False alarms, although errors in cue utilization are of obvious importance. Typically, errors gain attention when they are very unexpected (as in the finding that smiling was not a cue to dyadic rapport even though perceivers thought it should be; Bernieri et al., 1996). Similarly, in research on lie detection, the common belief that shifting eyes and gaze aversion are valid cues to lying appears to be wrong and this has received wide discussion within the lie detection field (DePaulo et al., 2003; Hartwig & Bond, 2011). The more strongly held the belief, the more memorable is its disconfirmation.

Through this framework, we urge researchers to pay more attention to the possible combinations of accurate and inaccurate pathways. Some researchers have compared meta-analyses for just this purpose. Hall, Coats, and Smith LeBeau (2005) compared a meta-analysis on people's nonverbal display of dominance/power to a separate meta-analysis on perceptions of dominance/power based on perceivers' ratings of nonverbal displays. This allowed comparison of how powerful people actually behave to perceivers' implicit beliefs (stereotypes) about how they behave. Many more behaviors were believed to signal high dominance/power than were actual cues to dominance/power.

Demonstrating inaccurate stereotypic beliefs is important because they obviously can have a life of their own, influencing people's future perceptions and responses to other people along stereotypic lines. The potentially self-fulfilling effects of stereotypical beliefs have often been alluded to (e.g., Darley & Fazio, 1980; Zebrowitz, Hall, Murphy, & Rhodes, 2002). Finding that more cues are used incorrectly than correctly is not unique to the power and behavior literature. Ruben and Hall (2016), for example, found this to be dramatically true in lens models of the expression and detection of acute pain, as did Gifford (1994) in research on the judgment of personality traits.

In a lens model, Misses and False alarms most commonly consist of a cue validity correlation that is significant and a cue utilization correlation that is not (or vice versa), but it can be the case that the cue validity (actuality) and the cue utilization (belief or stereotype) go in *opposite* directions. This divergence was found in a study where perceivers listened to excerpts of one person's part of a conversation and then guessed the gender of that person's unseen conversational partner (Hall & Braunwald, 1981). When women spoke more dominantly, they were perceived as speaking to another woman but in fact women spoke more dominantly to men than to women.

Looking at the ratio of correct cue utilizations (Hits and Correct rejections) to incorrect cue utilizations (Misses and False alarms) could be helpful in reaching a more substantive and balanced understanding of the real world in which accuracy is neither perfect nor nonexistent. However, any such ratio must be considered in light of how the particular set of measured cues was selected. In a well-studied research area, a researcher might pick cues that have high *a priori* likelihood to be both valid and to be correctly perceived, thus predisposing a finding of stereotype accuracy. This was the case in a lens model of acoustic properties associated with emotional communication in music (Laukka, Eerola, Thingujam, Yamasaki, & Beller, 2013), where the choice of measured acoustic cues was strongly grounded in previous research on acoustic correlates of different emotional states. The pattern of correct and incorrect cue utilizations showed a very high preponderance of Hits and Correct rejections.

On the other hand, in a less explored research area, a researcher might measure a variety of cues with the goal of *discovering* what cues are valid and correctly perceived (or not), with the result that few or perhaps none of the measured cues are correctly utilized. The nature of Misses and False alarms would, in such a case, illuminate which aspects of stereotypes are unfounded and how (because Misses and False alarms represent opposite kinds of incorrectness).

Even when there are few Hits, a lens model could still show good accuracy of perceivers judging the criterion. This suggests that perceivers made use of correct stereotypes (beliefs) about cues that were present in the stimuli but that the researcher *did not measure*; this is exactly what happened in Ruben and Hall's (2016) lens model of judgments of exaggerated acute pain, where the constellation of measured facial movements produced only Misses and False alarms even though accuracy of judging the targets' pain was high.

In gaining a picture of the overall pattern of correct and incorrect cue utilizations (stereotypes), a researcher could calculate the ratio of correctly to incorrectly used cues, keeping in mind that the analysis is limited to the particular collection of cues measured in the given study. One can also consider the magnitude of the paths that are uncovered. A False alarm path with a correlation of .70 between judgment and cue indicates a stereotype that is plainly inaccurate, whereas the same path with a correlation of only .30 indicates a stereotype that is more weakly held by perceivers. Of course, there are many variables determining the size of correlations so such interpretations would need confirmation.

Another way to look at stereotype accuracy is to correlate the vector of cue validities with the vector of cue utilizations. These vectors are simply the list of cue validity correlations and the corresponding list of cue utilization correlations, which can then be correlated together. A high correlation would mean that cues with the strongest stereotypes are also the most valid. Gosling et al. (2002), Hall et al. (2005), Hartwig and Bond (2011), and Hirschmüller, Egloff, Nestler, and Back (2013) all found positive evidence for stereotype accuracy using this kind of analysis in their analyses of judging personality from people's rooms, judging power/dominance, judging deception, and judging personality from video clips, respectively, meaning that perceivers weighted the more valid cues higher than the less valid cues when making their judgments.

The preceding paragraph gave examples of judges' sensitivity to the strength of cue validities in studies from the interpersonal accuracy tradition in which perceivers made judgments of many individual people. The same approach can be applied when whole social categories are the targets of judgment. This was done by Swim (1994) and by Hall

and Carter (1999). Hall and Carter compared 77 documented gender differences to participants' estimates of gender differences on the same 77 traits and behaviors. Correlations between the actual differences and participants' estimates of the differences were remarkably high, indicating a high degree of awareness of the relative sizes of the various gender differences; Swim (1994) also found this kind of correspondence. Also, Swim (1994) found that estimated gender differences did not generally exceed actually documented gender differences in magnitude.

5 | CONCLUSION

Stereotypes can be harmful, however benign they may seem (Czopp, Kay, & Cheryan, 2015). Nonetheless, stereotypes, conceived as beliefs about the behaviors and traits of groups, are normal cognitive processes that warrant full understanding. As previously advocated by others (e.g., Jussim, 2017; Ryan, 2003), understanding the conditions under which stereotypes can be accurate is important for theory development as well as prejudice interventions. If the working theory and dominant definition posit without much empirical support that all stereotypes are inaccurate, then effective change could be hindered by faulty assumptions. Furthermore, multicultural perspectives and appreciation of group differences (as opposed to a colorblind approach) have been shown to be particularly effective in combatting racial prejudice (Richeson & Sommers, 2016) but exaggerated beliefs about group differences are obviously problematic (Zell, Strickhouser, Lane, & Teeter, 2016). Understanding the delineation between group differences and stereotypes is an important development that could be advanced by adopting an integrative perspective on accuracy and inaccuracy. In order to fully comprehend the antecedents, development, and consequences of stereotyping, we urge researchers to consider the simultaneity of stereotype accuracy and inaccuracy.

ORCID

Judith A. Hall  <http://orcid.org/0000-0001-5840-2895>

REFERENCES

- Allport, G. W. (1954). *The nature of prejudice*. Reading, MA: Addison-Wesley.
- Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, *111*, 256–274.
- Ashton, M. C., & Esses, V. M. (1999). Stereotype accuracy: Estimating the academic performance of ethnic groups. *Personality and Social Psychology Bulletin*, *25*, 225–236.
- Back, M. D., & Nestler, S. (2016). Judging personality. In J. A. Hall, M. Schmid Mast, & T. V. West (Eds.), *The social psychology of perceiving others accurately* (pp. 98–124). Cambridge, UK: Cambridge University Press.
- Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, *54*, 462–479.
- Bernieri, F. J., Gillis, J. S., Davis, J. M., & Grahe, J. E. (1996). Dyad rapport and the accuracy of its judgment across situations: A lens model analysis. *Journal of Personality and Social Psychology*, *71*, 110–129.
- Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley: University of California Press.
- Campbell, D. T. (1967). Stereotypes and the perception of group differences. *American Psychologist*, *22*, 817–829.
- Chan, W., McCrae, R. R., De Fruyt, F., Jussim, L., Löckenhoff, C. E., De Bolle, M., ... Nakazato, K. (2012). Stereotypes of age differences in personality traits: Universal and accurate? *Journal of Personality and Social Psychology*, *103*, 1050–1066.
- Cheryan, S., Ziegler, S. A., Montoya, A. K., & Jiang, L. (2017). Why are some STEM fields more gender balanced than others? *Psychological Bulletin*, *143*, 1–35.
- Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychology Bulletin*, *136*, 1092–1122.
- Costa, P. Jr., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: robust and surprising findings. *Journal of Personality and Social Psychology*, *81*, 322–331.
- Czopp, A. M., Kay, A. C., & Cheryan, S. (2015). Positive stereotypes are pervasive and powerful. *Perspectives on Psychological Science*, *10*, 451–463.

- Darley, J. M., & Fazio, R. H. (1980). Expectancy confirmation processes arising in the social interaction sequence. *American Psychologist*, 35, 867–881.
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, 129, 74–118.
- Diekman, A. B., Eagly, A. H., & Johnston, A. M. (2010). Social structure. In J. F. Dovidio, M. Hewstone, P. Glick, & V. M. Esses (Eds.), *The SAGE handbook of prejudice, stereotyping, and discrimination* (pp. 209–224). New York: SAGE.
- Diekman, A. B., Eagly, A. H., & Kulesa, P. (2002). Accuracy and bias in stereotypes about the social and political attitudes of women and men. *Journal of Experimental Social Psychology*, 38, 268–282.
- Eagly, A. H. (1995). The science and politics of comparing women and men. *American Psychologist*, 50, 145–158.
- Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, 128, 203–235.
- Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, 102, 652–670.
- Gifford, R. (1994). A lens-mapping framework for understanding the encoding and decoding of interpersonal dispositions in nonverbal behavior. *Journal of Personality and Social Psychology*, 66, 398–412.
- 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., Back, M. D., Nestler, S., Frauendorfer, D., Schmid Mast, M., & Ruben, M. A. (in press). How do different ways of measuring individual differences in zero-acquaintance personality judgment accuracy correlate with each other? *Journal of Personality*.
- Hall, J. A., & Bernieri, F. J. (Eds) (2001). *Interpersonal sensitivity: Theory and measurement*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hall, J. A., & Braunwald, K. G. (1981). Gender cues in conversations. *Journal of Personality and Social Psychology*, 40, 99–110.
- 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., Coats, E. J., & LeBeau, L. S. (2005). Nonverbal behavior and the vertical dimension of social relations: A meta-analysis. *Psychological Bulletin*, 131, 898–924.
- Hall, J. A., Schmid Mast, M., & West, T. V. (Eds) (2016). *The social psychology of perceiving others accurately*. Cambridge, UK: Cambridge University Press.
- 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.
- Hirschmüller, S., Eglhoff, B., Nestler, S., & Back, M. D. (2013). The dual lens model: A comprehensive framework for understanding self-other agreement of personality judgments at zero acquaintance. *Journal of Personality and Social Psychology*, 104, 335–353.
- Hirschmüller, S., Schmukle, S. C., Krause, S., Back, M. D., & Eglhoff, B. (in press). Accuracy of self-esteem judgments at zero acquaintance. *Journal of Personality*.
- Hyde, J. S. (2014). Gender similarities and differences. *Annual Review of Psychology*, 65, 373–398.
- Imada, A. S., Fletcher, C., & Dalessio, A. (1980). Individual correlates of an occupational stereotype: A reexamination of the stereotype of accountants. *Journal of Applied Psychology*, 65, 436–439.
- Judd, C. M., & Park, B. (1993). Definition and assessment of accuracy in social stereotypes. *Psychological Review*, 100, 109–128.
- Jussim, L. (2017). Précis of social perception and social reality: Why accuracy dominates bias and self-fulfilling prophecy. *Behavioral and Brain Sciences*, 40, 1–20.
- Jussim, L., Crawford, J. T., & Rubinstein, R. S. (2015). Stereotype (in)accuracy in perceptions of groups and individuals. *Current Directions in Psychological Science*, 24, 490–497.
- Laukka, P., Eerola, T., Thingujam, N. S., Yamasaki, T., & Beller, G. (2013). Universal and cultural-specific factors in the recognition and performance of musical affect expressions. *Emotion*, 13, 434–449.
- Laukka, P., Elfenbein, H. A., Thingujam, N. S., Rockstuhl, T., Iraki, F. K., Chui, W., & Althoff, J. (2016). The expression and recognition of emotions in the voice across five nations: A lens model analysis based on acoustic features. *Journal of Personality and Social Psychology*, 111, 686–705.
- 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.
- Löckenhoff, C. E., Chan, W., McCrae, R. R., De Fruyt, F., Jussim, L., De Bolle, M., ... Terracciano, A. (2014). Gender stereotypes of personality: Universal and accurate? *Journal of Cross-Cultural Psychology*, 45, 675–694.

- Mackie, M. (1973). Arriving at "truth" by definition: The case of stereotype inaccuracy. *Social Problems*, 20, 431–447.
- Matsumoto, D., LeRoux, J., Wilson-Cohn, C., Raroque, J., Kooken, K., Ekman, P., ... Amo, L. (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.
- McCauley, C., & Stitt, C. L. (1978). An individual and quantitative measure of stereotypes. *Journal of Personality and Social Psychology*, 36, 929–940.
- Mehl, M. R., Vazire, S., Ramírez-Esparza, N., Slatcher, R. B., & Pennebaker, J. W. (2007). Are women really more talkative than men? *Science*, 317, 82.
- Richeson, J. A., & Sommers, S. R. (2016). Toward a social psychology of race and race relations for the twenty-first century. *Annual Review of Psychology*, 67, 439–463.
- Ruben, M. A., & Hall, J. A. (2016). A lens model approach to the communication of pain. *Health Communication*, 31, 934–945.
- 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., 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., 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.
- Ryan, C. S. (1996). Accuracy of Black and White college students' in-group and out-group stereotypes. *Personality and Social Psychology Bulletin*, 22, 1114–1127.
- Ryan, C. S. (2003). Stereotype accuracy. *European Review of Social Psychology*, 13, 75–109.
- Sahoo, F. M., & Sahoo, K. (2017). The perceived predictors of achievement in management students: An innovative use of lens model. *Journal of the Indian Academy of Applied Psychology*, 43, 34–40.
- 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.
- Swim, J. K. (1994). Perceived versus meta-analytic effect sizes: An assessment of the accuracy of gender stereotypes. *Journal of Personality and Social Psychology*, 66, 21–36.
- ten Brinke, L., Porter, S., Korva, N., Fowler, K., Lilienfeld, S. O., & Patrick, C. J. (2017). An examination of the communication styles associated with psychopathy and their influence on observer impressions. *Journal of Nonverbal Behavior*, 41, 269–287.
- 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.
- West, T. V., & Kenny, D. A. (2011). The truth and bias model of judgment. *Psychological Review*, 118, 357–378.
- Wickens, T. D. (2001). *Elementary signal detection theory*. New York: Oxford University Press.
- 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.
- Zell, E., Strickhouser, J. E., Lane, T. N., & Teeter, S. R. (2016). Mars, Venus, or Earth? Sexism and the exaggeration of psychological gender differences. *Sex Roles*, 75, 287–300.

Judith A. Hall, University Distinguished Professor at Northeastern University, received her BA and PhD degrees from Harvard University. Her research is on interpersonal processes including antecedents and consequences, especially with regard to nonverbal communication.

Jin X. Goh completed his PhD in social and personality psychology at Northeastern University, and he is currently a postdoctoral research associate at the University of Washington. His research focuses on the relationship between prejudice and nonverbal behavior.

How to cite this article: Hall JA, Goh JX. Studying Stereotype Accuracy from an Integrative Social-Personality Perspective. *Soc Personal Psychol Compass*. 2017;11:e12357. <https://doi.org/10.1111/spc3.12357>