CHAPTER

16   The Accuracy of Stereotypes About Personality 🔓

Lee Jussim, Sean T. Stevens, Nate Honeycutt

Abstract

This chapter first reviews general conceptual and methodological issues in the study of the (in)accuracy of stereotypes. First, the authors define stereotype accuracy. Second, different types of accuracy are discussed and standards for considering a stereotype belief accurate or inaccurate are presented. Finally, the chapter reviews the empirical evidence produced thus far that bears on the (in)accuracy of stereotypes regarding personality. The strongest evidence regarding the (in)accuracy of stereotypes regarding personality have focused on gender, age, and national character stereotypes. Weaker evidence relying primarily on self-reports of individual differences with respect to racial stereotypes and miscellaneous stereotypes (e.g., jazz vs. ballet dancers) is also reviewed. The chapter concludes by developing hypotheses that could help explain the pattern of (in)accuracy in stereotypes regarding personality.

Keywords: stereotype, accuracy, gender, age, national character

Subject: Social Psychology, Psychology

Series: Oxford Library of Psychology

p. 245   Stereotypes are one of the most fundamental social psychological phenomena, because they involve how people perceive other people. As such, they have been studied for nearly one hundred years. For most of that time, scholars defined them as inaccurate or declared that the empirical evidence indicated that they are inaccurate (see Jussim, 2012, for a historical review). Such a conclusion, however, was premature, and not based on much empirical evidence. Rigorous empirical assessments of the accuracy of stereotypes began in the late 1970s, accelerated in the 1990s, and, now, with over 50 empirical studies published, has yielded a surprising conclusion: highly inaccurate stereotypes are the exception; moderate to high accuracy is common. These issues have been amply reviewed elsewhere (e.g., Campbell, 1967; Judd & Park, 1993; Jussim, 2012; Jussim, Cain, Crawford, Harber, & Cohen, 2009; Jussim, Crawford, Anglin, Chambers, Stevens, & Cohen, 2016; Jussim, Crawford, & Rubinstein, 2015; Jussim, Stevens, & Honeycutt, 2018; Lee, Jussim, & McCauley, 1995; Mackie, 1973; Ryan, 2002), and is not a main focus of the present chapter, though we do draw heavily on our prior work for this review of the evidence regarding the accuracy of stereotypes about

personality. Rather than re-list all the reviews just cited each time they are relevant, we hence refer to those reviews with the term "the general stereotype accuracy literature."

The unique contribution of this chapter is that it reviews the evidence specifically regarding the (in)accuracy of stereotypes regarding personality. As such, it is the first review to do so of which we are aware. First, we provide a conceptual overview of what stereotype accuracy is, and then describe how it can be assessed. Second, we identify what types of studies are included and excluded from this review, and explain why. The third section is the core of the chapter—it reviews every study of the (in)accuracy of stereotypes regarding personality that we were able to discover. We end the chapter with some speculative conclusions about what explains patterns of accuracy and inaccuracy in stereotypes regarding personality.

What Is Stereotype Accuracy?

Answering this question requires three steps: (1) defining stereotype, (2) defining accuracy, (3) combining them. Our definition of stereotype is based on that of Ashmore and Del Boca (1981): people's beliefs about groups and their individual members. This leaves the accuracy of stereotypes as an open empirical question.

We use the term "accuracy" to refer to correspondence between a belief and a credible criterion. Stereotype accuracy, therefore, refers to degree of correspondence between a belief about a group and some criterion against which that belief can be evaluated. In the general stereotype accuracy literature, common criteria include Census and government data, meta-analyses of group differences, behavioral observations, and group self-reports. Potential criteria in the general stereotype accuracy literature span a vast variety of characteristics, including academic achievement, beliefs and attitudes, preferences, wealth and income, interpersonal skills, and personality (which is the exclusive focus of the present chapter). The general stereotype accuracy literature has addressed controversies and limitations to various criteria (see Jussim, 2012, for the most extended discussion). In short, self-reports are often considered the weakest of these criteria; however, some self-report criteria have been subject to extensive validation (such as the NEO-PI-R, which is used in many of the studies reviewed herein).

How to Assess the Accuracy of Stereotypes

The general stereotype accuracy literature has long articulated what is necessary to empirically investigate the accuracy of stereotypes: (1) Assess beliefs about one or more groups, (2) Identify criteria for evaluating the validity of those beliefs, and (3) Compare them. To assess the accuracy of a stereotype a *single individual* holds, one needs to: (1) Assess that person's belief about one or more groups, (2) Identify an appropriate criterion for that belief, and (3) Compare them. One can also assess *consensual* accuracy by taking these steps: (1) Assess the average belief held by a group of perceivers about a target group, (2) Identify criteria, and (3) Compare mean perceptions to criteria. Correspondence with criteria is typically assessed with correlations; discrepancies from criteria are typically assessed with difference scores.

Stereotype accuracy is usually assessed with either *discrepancies* or *correlations*. Discrepancies assess how close to criteria people's beliefs come. They can be computed in either of two ways, if the stereotype and criterion are assessed on the same scale. The simplest is direct comparison of stereotype to criteria by subtraction (stereotype minus criteria equals discrepancy). These discrepancy scores can be viewed as assessing: "How close to perfect accuracy are people's stereotypes?" Sometimes, however, researchers compare standard deviations (SDs); if the SD among stereotype perceptions is larger than that among criteria, it indicates that stereotypes exaggerate real differences; if smaller, it indicates they underestimate real differences. This latter method has the benefit of providing an overall assessment of exaggeration/underestimation across all groups and attributes, and the limitation of not permitting more

fine-grained analyses regarding *which* attributes are accurately versus inaccurately perceived. Regardless of the method used, however, smaller discrepancies indicate greater accuracy.

Furthermore, componential approaches can be used to test hypotheses about particular sources of inaccuracy (Judd & Park, 1993; see Jussim, 2012, for a review and critique of several componential approaches). Componential approaches treat discrepancy scores as best understood through analysis of variance. In Judd and Park's (1993) approach, components are divided into effects attributable to perceiver groups (do some groups of perceivers consistently over- or underestimate targets' standing on attributes); target groups (are some groups more likely to be over- or underestimated by perceivers), attributes (are some types of attributes more likely to be over- or underestimated than others), and all possible two-way interactions and the three-way interaction. The meaning and interpretation of the two-way and three-way interactions hinge on the nature of the attributes. If attributes are chosen such that they are stereotypical for one group and counterstereotypical for the other, the three-way interaction has particular significance, because it can reveal in-group bias. If, for example, stereotype exaggeration (overestimation of the difference between stereotypical and counterstereotypical attributes) occurs more when people judge the other group than when they judge their own group (i.e., more when Group A judges Group B and when Group B judges Group A than when Group A judges Group A and Group B judges Group B), one would have evidence of in-group bias.

Correlations assess how well people's beliefs about groups correspond to criteria. Higher correlations indicate greater correspondence—that is, higher accuracy. Correlations assess how sensitive people's stereotypes are to *variability* in the actual differences between criteria.

The different types of information about accuracy provided by discrepancy scores versus correlations can be seen with a simple hypothetical example. A perceiver estimates the average height of Nepalese, ↵ Norwegian, and Nigerian men as 60, 62, and 64 inches. If the real average heights are 72, 74, and 76 inches, the discrepancy is a full foot off—quite inaccurate. And yet the perceiver's estimates correlate 1.0 with the real differences. This occurs because, even though the estimates of average height are far off, they are nonetheless perfectly sensitive to variations in height.

Whether one assesses accuracy via discrepancy scores or correlations, one can assess two types of stereotypes: consensual and personal stereotypes. Consensual (or aggregate) stereotypes are shared by the members of a culture or a particular sample. They have so far always been sample means (e.g., the mean belief about women's height in a sample is the best estimate of the consensual stereotype for women's height for the group sampled; though it is imaginable that other metrics of shared-ness could be justified, such as a sample median). Personal stereotypes are the beliefs about a group held by a single individual, whether or not that belief is shared.

Standards for Considering a Stereotype Accurate or Inaccurate

There is no objective standard for accuracy and there is no established consensus in the literature regarding what constitutes an accurate or inaccurate belief. Indeed, this is both ironic and problematic, given the frequency with which many researchers have declared stereotypes to be "inaccurate." That is, without articulating any standard for (in)accuracy, it is not clear how anyone can justify declaring stereotypes to be either accurate or inaccurate. In order to redress this limitation, we have developed heuristic standards for doing so (see Jussim, 2012; Jussim et al., 2009; Jussim et al., 2016, for extended discussions).

Because there are two broad types of accuracy, discrepancy from perfection and correspondence with real differences, there needs to be two separate standards. For discrepancies, we consider judgments within 10% or 0.25 SDs of the criteria to be accurate, and within 20% and 0.50 SDs to be near misses. We consider judgments that are more than 20% or 0.50 SDs discrepant from criteria to be inaccurate.

For example, McCauley and Stitt (1978) examined the accuracy of racial stereotypes by comparing them to Census data (although they reported results primarily as diagnostic ratios, the original data can be found in McCauley, 1995). At the time, 39% of African Americans had completed high school. Three of five samples estimated that high school completion rate as 40%–48%; by our standard, all three were accurate because 39% +/-10% means anything from 29%–49% would be considered accurate. The other two sample estimates were 50% and 52%, and, by our standards, would be considered near misses (39% +/- 20% means anything outside the 29%–49% range but within 19%–59% would be considered a near miss). If criteria and stereotypes are measured in other sorts of units, SDs could be used in an essentially similar manner. For personality judgments, SDs are more likely to be used than percentages. For example, Rogers and Wood (2010), in a study described later in more detail, examined the accuracy of American regional stereotypes and reported both stereotypes and personality criteria using standardized scores (z scores). Alaskans were seen as quite low on extraversion (z = -1.88), which, in fact, they were (z = -1.63). Because the stereotype is within 0.25 SDs of the criteria, by our standard, the stereotype of Alaskans' extraversion is accurate.

How much correspondence should be considered "accurate"? Absent reason to do otherwise (and such reasons are usually absent), we use the same standards to which social scientists hold themselves. The standards first articulated by Cohen (1988) are that correlations of 0.4 and higher could be considered accurate because they are "large." Only 24% of social psychological effects exceed 0.3 (Richard et al., 2003). As per Rosenthal's (1991) binomial effect size display, a correlation of 0.4 translates into people being right 70% of the time. This means they are right more than twice as often as they are wrong; $r \geq 0.4$ is a strong standard.

Moderate correspondence reflects a mix of accuracy and inaccuracy. Following the same standards as science (Cohen, 1988; Richard et al., 2003), we characterize correlations between people's beliefs and criteria ranging from 0.25 to 0.4 as moderately accurate. Correlations below 0.25 are considered inaccurate.

These standards for accuracy should not be reified. They have been offered more in a spirit of reasonable heuristic defaults. It is possible that for some purposes different standards could be used and justified. We strongly encourage researchers who wish to make claims about either the accuracy or inaccuracy of stereotypes to articulate standards for doing so. In the absence of such articulation, scientists are in no position to make *any* claims about the (in)accuracy of stereotypes in general, or of particular stereotypes regarding particular groups held by particular people.

What Stereotype Accuracy Is Not

Stereotype accuracy is not the same as group differences or similarities. Stereotypes have sometimes been declared to be "inconsistent with" some stereotype or another, without referring to evidence that actually empirically compares' specific people's measured stereotypes to criteria (e.g., Ellemers, 2018; Hyde, 2014). The only way, however, to know whether group differences (or similarities) confirm or disconfirm people's stereotypes is to assess those stereotypes and compare them to group characteristics. If any piece is lacking —the assessment of the stereotype, the assessment of group characteristics, or the comparison—no conclusions about stereotype (in)accuracy are scientifically justified. Thus, all claims that "stereotypes are inaccurate" that appear in the scientific literature (Ellemers, 2018; Hyde, 2014; see Jussim, 2012; Jussim et al. 2016, for reviews) are themselves not justified because they have never been made on the basis of research that empirically assesses and compares stereotypes to group differences.

When Stereotype Accuracy Cannot Be Assessed

Accuracy of stereotypes cannot always be assessed for both empirical and conceptual reasons. If no empirical criteria for accuracy exist, the validity of the stereotype cannot be evaluated. Conceptually, accuracy can be assessed for descriptive or predictive beliefs, but not for prescriptive or proscriptive beliefs. Descriptive beliefs are supposed to describe reality in some way. "Ohio gets the most rain in April," describes the relative frequency of rain in April (implicitly compared to other months). Similarly, "The Yankees have won more World Series than any other baseball team" described the relative frequency of World Series victories (implicitly compared to other teams). Similarly, the accuracy of predictive beliefs can be evaluated for accuracy. If Jana says, "It will rain here tomorrow," either there will or will not be rain. Similarly, if in 2019 Musa said, "Eighty-five to ninety-five percent of African American voters will vote for the Democratic candidate for President in 2020," this will either turnout true or not.

Prescriptive beliefs describe (one's opinions about) what things or people *should* be like. "Good shoes have arch support" is a prescriptive belief, as is, "children should be seen and not heard." Only descriptive and predictive beliefs, but not prescriptive beliefs, can be evaluated for their accuracy. Whether good shoes always require arch support or children should be quiet is personal preference; and personal preferences cannot be evaluated for their objective accuracy. In contrast, descriptive and predictive beliefs, whether about rain in April or the frequency of World Series victories, can be evaluated for accuracy.

Similarly, personal preferences cannot be evaluated for accuracy regarding the things or people preferred. The accuracy of "I don't like Buddhists" is not a descriptive or predictive belief about Buddhists and therefore, as a statement about Buddhists, cannot be evaluated for accuracy. The accuracy of prescriptive stereotypes, such as "professional football players should be role models," or "girls should wear skirts," cannot be evaluated for accuracy. When stereotypes involve something *other* than descriptive or predictive beliefs, no claims about stereotype (in)accuracy are possible.

What Is Included and Excluded in This Chapter

This chapter focuses exclusively on stereotype accuracy regarding personality. Studies failing to meet the standards described previously, which require assessment of a stereotype, identification of criteria, and a comparison, are excluded. Thus, studies examining only the personality content of some stereotypes are excluded, even if the authors' reached conclusions about accuracy, because, in fact, they have no empirical grounds for doing so. Similarly, studies that merely assess group differences in personality, but not anyone's stereotypes about those groups, are excluded, again, because no empirical assessment of accuracy is possible.

Studies focusing on stereotype accuracy regarding characteristics other than personality are excluded. This includes, but is not restricted to, studies assessing stereotype accuracy regarding demographic characteristics (e.g., wealth, income, likelihood of using social services, likelihood of committing or being victimized by crime, etc.), political attitudes, personal preferences or tastes, academic achievement, and so on. We do include studies of demographic stereotypes *if and only if* those studies assessed stereotypes regarding personality (rather than, e.g., regarding attitudes, wealth, etc.).

One can also think of personality characteristics in terms of stereotypes. Do people have stereotypes of extraverts, introverts, neurotics, and so forth? They probably do, or, at least, could, if they stopped to think about it. However, we know of no research that has assessed the accuracy of such beliefs, so none is reviewed here.

Some prior reviews of stereotype accuracy have also addressed whether perceivers rely primarily on stereotypes or individuating information in person perception, and on whether relying on a stereotype

increases or reduces the accuracy of person perception judgments (Jussim, 2012; Jussim et al., 2018). Although these issues are related to issues of stereotype accuracy, they are also quite different. Finding out that, for example, Shakira has an accurate perception regarding Lois's or Mohamed's personality is interesting and important, but it tells us nothing about the overall accuracy of her gender, ethnic, or religious stereotypes. Thus, these issues are not included here (but see chapter by Rogers in this handbook for a discussion of the role of normative information in judgments of others).

The Empirical Evidence

We have discussed much, though not all, of the literature presented in what follows regarding national character, age, and gender stereotypes in our prior reviews of stereotype accuracy (Jussim, 2012; Jussim et al., 2009; Jussim et al., 2016; Jussim et al., 2015). Although more detail can be found in those reviews, our descriptions of the specifics of the studies presented here are often quite similar to that which appeared in one or more prior reviews. Nonetheless, this is the first time any of our reviews have focused specifically on the evidence regarding personality.

The (In)accuracy of National Character Stereotypes

Several studies have examined the accuracy of national or regional personality stereotypes (sometimes referred to as "national character stereotypes") using the NEO-PI-R, which is essentially an updating of what were once called Big Five Personality Inventories. These have been administered to thousands of people worldwide (e.g., Costa & McCrae, 2008; Costa, Terracciano, & McCrae, 2001; McCrae & Allik, 2002, McCrae & Terracciano, 2005), so there is now data on personality characteristics in many different countries. These can be and have been used to assess the accuracy of "national character (i.e., personality) stereotypes."

The first of such studies examined the accuracy of consensual national character *autostereotypes* (stereotypes of their own group—e.g., Italians' beliefs about Italians) in 49 cultures (*N* of almost 4,000) worldwide (Terracciano et al., 2005) which were compared to mean observer ratings of individuals and self-reports on the NEO-PI-R. Correlational accuracy (consensual stereotype accuracy correlations assessed as the intraclass correlation of autostereotypes with observer ratings and self-reports) averaged near zero. Furthermore, consensual stereotypes exaggerated real differences; although they did not directly assess discrepancy scores, they did show that perceived differences between national groups generally exceeded criterion differences. Personal stereotype accuracy was not assessed.

There were, however, several limitations to this research. First, the criterion samples used in this study were haphazard samples of convenience. This meant that low levels of correspondence between stereotype and criteria could have occurred, not because people were inaccurate, but because the criterion sample was not representative of the target population (see Judd & Park, 1993; Jussim et al., 2016 for more on the difficulties that arise when the target sample is mismatched to the stereotype assessed, also referred to as the mismatch limitation).

Heine, Buchtel, and Norenzayan (2008) identified another reason that people's national character stereotypes might have been more accurate than found by Terracciano et al. (2005). They argued that people rely on local cultural norms when arriving at perceptions of themselves (called "the reference group effect") which can distort measurement of cultural differences. To address this issue, they compared stereotypes to behavior, focusing exclusively on conscientiousness; thus their findings do not bear on the other Big Five characteristics (agreeableness, neuroticism, openness to experience, extraversion). Heine et al. (2008) operationalized conscientiousness behaviorally (GDP, longevity, walking speed, clock accuracy,

and postal worker speed). The correlations between consensual national autostereotypes and behavior averaged about 0.60, indicating high correspondence accuracy. They assessed neither personal stereotype accuracy nor discrepancy scores.

But the controversy did not end there. A replication (McCrae et al., 2013, *N* > 3,000) of Terracciano et al. (2005) addressed some of the issues raised by Heine et al. (2008), though it still used self-reports on the Big Five as criteria. McCrae et al. (2013) discussed several reasons to be skeptical of low accuracy correlations resulting from the reference group effect. For example, foreigners have shown a pattern similar to that of their national compatriots, strongly suggesting that people are not using local norms as their basis for social judgments. The McCrae et al. (2013) replication, like the original (Terracciano et al., 2005), found

almost no evidence of accuracy. These findings are consistent with research showing ↳ little accuracy in the personality stereotypes held by over 2000 Austrian, Czech, German, Polish, and Slovak college students regarding each other and their own country (Hrebickova & Graf, 2014). As is typical in this area, the criterion samples were large and international, but not random or representative.

In contrast, another study (Allik, Alyamkina, & Mescheryakov, 2015) of the consensual stereotype accuracy correlations held by two Finno-Urgic ethnic groups (Erzians, Mokshans) and Russians regarding one another's personalities (as measured by the NEO-PI-R), in the Republic of Mordovia (Russian Federation) found: (1) at least moderate accuracy in autostereotypes (correlations of 0.38, 0.51, 0.25, respectively); (2) substantial accuracy in Ezrians' stereotypes regarding Mokshans (*r* = .39) and Mokshans' stereotypes of Ezrians (*r* =. 51); (3) considerable accuracy in Ezrians' (*r* = .51) but not Mokshans' (*r* = .07) stereotypes of Russians; and (4) stereotype accuracy correlations near zero for Russian perceivers regarding Erzians and Mokshans. The accuracy levels obtained in this study may be underestimates, however, because the criterion samples were small (100 for each group) haphazard samples of convenience.

Another study focused on the accuracy of six eastern European ethnic groups' (Finns, Estonians, Latvians, Lithuanians, Poles, and Belarussians) autostereotypes and stereotypes of Russians (Realo et al., 2009). There were 200 or more participants for each perceiver group (most were college students, but there were also subsamples of working adults among the Latvians and Estonians). Convenience samples constituted perceiver and target groups, whose personality and perceptions of national character were assessed on the NEO-PI-R. The evidence of personality stereotype accuracy was mixed. Consensual autostereotype correlations ranged from 0.07 (for Estonian and Latvian students) to 0.52 (for Russians). The only heterostereotype examined was regarding Russians, which was inaccurate (the median of eight consensual stereotype accuracy correlations was near zero).

Overall, then, the pattern of results regarding national character stereotype accuracy is both mixed and limited. We are aware of no studies that have even examined personal stereotype accuracy (either correlational or discrepancies). Nearly all of the remaining work has focused exclusively on correlational accuracy and has found a decidedly mixed pattern. When using the NEO-PI-R (Big Five) to assess traits, which is a self-report measure, albeit a highly validated one, there is much evidence of inaccuracy (both across traits and across groups), interspersed with some evidence of moderate accuracy. There is only the one study using behavioral criteria (Heine et al., 2008) and it only examined conscientiousness, but it found a high level of consensual stereotype correlational accuracy. What explains these differences in patterns of (in)accuracy is not clear. As such, it would seem to be fertile ground for empirical research that could provide deeper insights into personality, personality measurement, social perception, and stereotypes. We propose some speculative hypotheses about these issues at the end of this chapter.

A Mixed Pattern for American Regional Stereotypes

Rogers and Wood (2010) examined the accuracy of undergraduates' American regional stereotypes of personality (the Big Five). The country was divided into 18 clusters of states, plus Alaska and Hawaii (separately). Results were compared against an Internet sample of over 600,000 people from across the United States who had completed a Big Five Personality inventory.

Although the perceiver sample was small (84), the study is unique in several ways. In addition to being the only study of the accuracy of American regional stereotypes of which we are aware, it is one of few studies to assess consensual discrepancies, consensual accuracy correlations, and personal stereotype accuracy correlations (it did not assess the accuracy of personal discrepancies). There were 100 judgments (the Big Five traits by 20 clusters of states): 21 consensual stereotype discrepancies were accurate; 23 were near misses, and the remaining 56 were inaccurate.

For all regions, consensual stereotype accuracy was high for neuroticism (.59) and openness (.48), but near zero for extraversion, agreeableness, and conscientiousness. However, perceptions regarding Alaska and Hawaii were sufficiently inaccurate to be considered outliers. Without them, the correlation for openness shot up to 0.78; the previously near zero correlation for agreeableness became 0.56; the correlation for neuroticism was largely unchanged at 0.60; the near zero correlation for extraversion stayed near zero; and the previously near zero correlation for conscientiousness became –.55 (i.e., strongly in the wrong direction). Personal stereotype accuracy correlations were moderate for neuroticism and openness (r's = .31 and 0.25, respectively), but near zero for extraversion, agreeableness, and conscientiousness.

The small sample of perceivers means that all of these results should be interpreted as preliminary. Furthermore, although we wish we could tell some compelling narrative about the meaning of these results, there is no obvious explanation (at least to us) for the inconsistent pattern of accuracy. Additionally, because of the small sample, there probably is a great deal of uncertainty (wide confidence intervals) in the results and they probably should not be overinterpreted pending larger-scale assessments of US regional stereotype accuracy regarding personality.

The Accuracy of Age Stereotypes

We know of only one study of the accuracy of age stereotypes of personality. Chan et al. (2012) examined the accuracy of age stereotypes regarding personality in 26 countries among over 3,000 people. Criteria were again self-reports on NEO-PI-R (Big Five). Three patterns emerged. Consensual stereotype accuracy correlations ranged from 0.50 to 0.90. Discrepancy scores, however, showed that people consistently exaggerated real differences among adolescents (14- to 20-year-olds), adults (21–59), and the old (60+). For example, the perceived difference between adolescents and the old were $d = .90$, but the criteria differences were only $d = .54$. A similar pattern was found for extraversion, openness, and conscientiousness (there was a slight tendency to underestimate the real differences on agreeableness).

Last, Chan et al. (2012) assessed personal stereotype accuracy correlations, which averaged 0.34, indicating moderate accuracy. There was a great deal of consistency in these patterns across country, gender, and age of rater. Thus, these patterns may be universal.

Representative samples were available as criteria for some, but not all countries. Analyses showed that levels of accuracy did not vary much based on whether the criterion sample was representative or not. This is the only study we know of to empirically demonstrate that accuracy was essentially the same regardless of whether convenience or representative samples were the basis for criteria. Although representative samples remain the gold standard for criteria, these findings showing similar results regardless of sampling

procedure increase the credibility of findings from research on the accuracy of stereotypes regarding personality that has relied exclusively on convenience samples.

The Accuracy of Stereotypes Regarding the Rich

A recent study examined the accuracy of perceptions of personality regarding the rich, who were operationally defined as people with a net worth in excess of one million Euros (the study was conducted in Germany; Leckelt, Richter, Schroder, Kufner, Grabka, & Back, 2019). The personality attributes were the Big Five plus narcissistic admiration and locus of control. They examined (in)accuracy at both consensual and individual levels using both discrepancies and correlations.

This study focused on perceptions of differences. As such, they also obtained personality information from a representative sample of Germans. Perceived differences were then compared to the real differences (as indicated on self-reports on the BFI-S, the short form of the Big Five Inventory). Consensual discrepancies exaggerated real differences on four of the attributes (neuroticism, agreeableness, narcissistic admiration, and locus of control with exaggeration effects $0.23 \leq d < 1.02$). Perceived differences between the very rich and the general population were no more than three-tenths of a scale point (on a scale running from one to seven) for neuroticism, extraversion, openness, and conscientiousness; it was half a scale point for agreeableness. The consensual correlation was quite high ($r = .83$). At the personal (individual perceiver) level, people perceived personality differences between the rich and general population quite accurately, the average correlation between the two was $r = .56$.

Gender

Some of the clearest evidence regarding the accuracy of gender stereotypes regarding personality appeared in a paper strikingly titled, "Gender Stereotypes of Personality: Universal and Accurate?" (Löckenhoff et al., 2014). In this study, over 3,000 participants in 26 countries indicated their perceptions of males and females on the NEO-PI-R (Big Five) personality traits. Women were stereotyped as more agreeable, conscientious, open to experience, and neurotic than men, and as lower than men on extraversion.

These stereotypes were then compared to self-reports on the NEO-PI-R obtained from separate samples in the same countries and to observer reports of sex differences based on prior research. Although these were convenience samples, they totaled over 20,000 people from 26 countries. For all five personality traits, consensual stereotype discrepancies were accurate, ranging from SD differences of 0.01 to 0.22, regardless of whether self-reports or observer reports were used as criteria. There was no tendency to exaggerate differences. Löckenhoff et al. (2014, Table 2) also examined the accuracy of consensual gender stereotype

p. 252 correlations, separately for ↳ beliefs about young, adult, or old males and females. In general, these stereotypes met our standards for being considered accurate, ranging from 0.36 to 0.70, with a median of 0.47.

The criterion samples were large, but not representative, so this study suffers from the mismatch limitation. Because mismatch means that the stereotyped target group could differ from convenience criterion samples, the most likely implication is that results reflect a lower bound on accuracy.

Three older studies reported in two papers also assessed gender stereotype accuracy regarding personality. Martin (1987) assessed the accuracy of gender stereotypes in two studies. The first used a version of the Bem Sex Role Inventory (BSRI, Bem, 1981) modified to assess college students' stereotypes about "North American adult males and females" on 30 trait descriptors. The original BSRI was a self-report measure using a 7-point scale; the modification involved asking the college students to identify the percentage of men and women who had each trait. Although not all items were about personality (e.g., "defends beliefs"),

many were at least personality-like enough to include in this review (e.g., assertive, dominant, compassionate, affectionate, conscientious, etc.). Unfortunately, the criterion sample was possibly even less representative than a typical convenience sample. It was 94 male and 56 female adult visitors to a campus open house day (who were asked to rate themselves on the BSRI using a true/false format).

Martin (1987) presented the main results as *diagnostic ratios* (DRs, proportion of males/proportion of females, computed separately for the stereotype judgments and criterion self-reports). These tended to exaggerate real differences. For example, 54% of men and 34% of women self-described as dominant, producing a DR of 1.59 for the difference. However, the DR for the perceived difference was 1.98, indicating that people believed there was a greater sex difference in dominance than was reflected in the self-descriptions.

However, DRs have some computational quirks. For example, let's say a perceiver believes 5% of men and 1% of women are assertive, whereas the self-reports are, respectively, 20% and 16%. The perceived difference (4%) exactly equals the real difference. However, the DR will make it appear as if the perceiver exaggerates the real difference, because his DR is 5, whereas the DR for the criterion differences is merely 1.25. One way to view this limitation is that DRs become easily inflated by an overall tendency for people to make low estimates (which is what Martin, 1987 found), thereby leading to inflated estimates of exaggeration.

However, in the first study, Martin (1987) also presented the raw mean percentage estimates of men and women with each trait. When using these raw percentage estimates (Table 3, p. 493), perceptions of male/female differences were largely accurate using our standards of +/-10%. Specifically, perceptions of differences were accurate 24 times, there were 4 near misses, and 2 judgments were inaccurate. The accuracy could have occurred because: (1) male/female tendencies to self-inflate were comparable; (2) subtracting male minus female self-perceptions implicitly removes that bias; (3) what remains is largely valid; so that (4) perceptions of differences were not very discrepant from the implicitly self-inflation-adjusted real differences. Nonetheless, in absolute terms, 22 judgments were overestimates of differences and 8 were underestimates, indicating that exaggeration occurred more often than underestimation.

Although Martin (1987) did not report consensual stereotype accuracy correlations, they are computable from her data reported in Table 3. Consensual stereotype accuracy correlations were completely inaccurate for male targets ($r = -.10$), but quite accurate for female targets ($r = .72$) and for male/female differences ($r = .80$). Personal stereotype accuracy correlations were not reported.

Study 2 altered the procedure in two major ways. First, self-reports and stereotypes were assessed with a new composite measure, including 32 items from the Extended Personal Attributes Questionnaire (EPAQ; Spence, Helmreich, & Holahan, 1979), and eight from the BSRI. Second, instead of asking about "North American males and females," the stereotype measure asked about the "male and female students at this university" (University of British Columbia). Participants were merely described as 106 female and 33 male undergraduate volunteers. Thus, this is another convenience sample and suffers from a stereotype/criterion sample mismatch problem. Further, the male sample is small.

Martin (1987) did not report the raw percentage values for either the self-reports or stereotype judgments in the second study. The only results reported for the 40 judgments were DRs. Those showed mostly exaggeration of sex differences. However, to assess consensual correlational accuracy, it is easy enough to correlate the 40 DRs on the self-report criteria with the 40 stereotype judgments. That correlation is $r = .53$, indicating considerable accuracy in perceiving gender differences.

Allen (1995) performed a replication and extension of Martin's (1987) second study. One hundred introductory psychology students provided both stereotypes and self-reports on the same 40 items used by Martin (1987, Study 2). Unfortunately, Allen's (1995) method section did not explicitly articulate whom

participants were being asked about. However, in Martin's (1987) second study, they were asked about "students at this university." If, as a replication, Allen (1995) did the same, this study suffers from the stereotype/criterion mismatch.

Allen's (1995) main results were reported as DRs. Unfortunately, the raw percentage estimates were not reported. For both stereotype judgments and self-reports, the DRs ranged from 0.16 to 4.22, with one exception: the DR for *unprincipled* was almost 10. This was a bizarre outlier; nearly 10 times as many men as women described themselves as unprincipled. Because this outlier was so extreme, we excluded it from our subsequent summary, which is based on the other 39 attributes. Because Allen (1995) reported DRs rather than raw percentages, we cannot apply any of our standards for evaluating how accurate people were (even using his mismatched criterion sample). As with Martin (1987, both studies), the DR measure indicated exaggeration. Although Allen (1995) did not compute consensual accuracy correlations, they are readily computable from his Table II. That correlation was 0.61 (after excluding the outlier), indicating high accuracy.

Stereotypes of Ballet Versus Modern Dancers

One of the most unique studies of stereotype accuracy was conducted by Clabaugh and Morling (2004), which investigated the accuracy of stereotypes regarding modern dancers and ballet dancers. This study asked perceivers to rate modern dancers and ballet dancers in general, but then used the self-reports of the haphazard samples of dancers (recruited from what they described as seven preprofessional dance camps in Pennsylvania and New York City) in their study as criteria. The study examined stereotypes held by modern (*N* = 48) and ballet (*N* = 41) dancers attending a professional dance camp and by a sample of introductory psychology students (*N* = 34) regarding modern and ballet dancers.

Not all judgments were about personality, but enough were that we have included it in our review. Specifically, they assessed people's beliefs about the self-esteem, body image, physical condition, fear of negative evaluation, need for structure, and need for control regarding the different dancers, and used the dancers' self-reports on these same items as criterion. Unfortunately, the questionnaire assessed beliefs about modern and ballet dancers in general, rather than about those attending the camp, thereby creating mismatch between the stereotype assessed and the criteria.

Clabaugh and Morling (2004) reported the mean differences between groups in self-reports, and the mean perceived differences between the groups. Discrepancy scores showed high accuracy and some systematic error. At the consensual stereotype (aggregate) level, 11 of 18 perceived differences (six traits by three groups of perceivers) were accurate, six were near misses, and one was inaccurate. Of the seven inaccuracies (including the six near misses, four exaggerated the real difference, one underestimated the real difference, and there were two reversals (both groups of dancers believed that modern dancers had higher "body esteem" although ballet dancers reported higher body esteem). They did not report personal discrepancy scores.

Although Clabaugh and Morling (2004) did not report the correlation of perceived with criterion differences, it could be computed from their data. These consensual beliefs about group differences corresponded well with differences in the dancers' self-reports. Consensual stereotypes correlated with the self-reports 0.83, 0.90, and 0.79 for perceivers who were, respectively, ballet dancers, modern dancers, or introductory psychology students.

They also assessed the extent to which personal stereotypes corresponded with the dancers' self-reports. One set of analyses assessed how sensitive people were to different levels of each characteristic within each group (e.g., how sensitive are people to differences in the body image, self-esteem, etc., of ballet dancers). For each perceiver, Clabaugh and Morling (2004) computed the correlation between their perceived level of

each characteristic with the self-reported mean level. Among individual psychology students, these correlations indicated only modest sensitivity to variations in the traits of ballet dancers (average correlation between beliefs and criteria was 0.23). Among all other combinations of perceiver group (ballet, modern, intro psych student) and target group (ballet and modern), these correlations were substantial (the average correlations ranging from 0.48 to 0.63). This result is consistent with the *stereotypes as knowledge* hypothesis (Jussim et al., 2016, 2018), because people likely to have far more familiarity with dancers (other

dancers) were more accurate ↳ than were those less likely to be familiar with dancers (introductory psychology students).

They also assessed people's sensitivity to differences between ballet and modern dancers (how well do perceived differences correspond to the self-reported differences?). Again, these were personal stereotypes, because they computed these correlations for each perceiver. These average correlations were strikingly high: 0.67 for ballet perceivers, 0.71 for modern dance perceivers, and 0.62 for the introductory psychology students.

Summary of the Literature

Given how long social scientists have been studying stereotypes and, especially, making claims about their inaccuracy (Lippmann, 1991[1922]), there has been a surprising paucity of empirical research on stereotype accuracy until fairly recently (see reviews by Jussim, 2012; Jussim et al., 2018), and only a fraction of that literature has assessed the accuracy of stereotypes regarding personality.

Even when there is a substantial literature, as is the case with national character stereotypes, it has so many limitations and/or contradictory findings that it is not possible at this time to tell a simple compelling narrative about the data. National character stereotypes, when measured by Big Five and related instruments, have usually been found to be highly inaccurate, though there is some evidence of moderate accuracy in some studies. Moreover, the one study that used behavioral rather than self-report measures as criteria (Heine et al., 2008) found high consensual correlational accuracy for conscientiousness but did not assess any other trait or type of accuracy. Because there is ample evidence for the validity of the NEO-PI-R (McCrae & Terracciano, 2005), our view is that a breezy dismissal of the difference cannot be easily attributed to "it's just self-report." It is self-report, but when self-report scales have been highly validated, breezy dismissals are not justified. American regional stereotypes are probably conceptually similar to national character, and also showed evidence of moderate accuracy, though these have been the subject of only a single, small-scale study (Rogers & Wood, 2010). On the other hand, there were considerably higher levels of accuracy in stereotypes of personality regarding age, gender, and wealthy people.

To their credit, most of the national character studies have had very large samples and have examined stereotype accuracy cross-culturally. On the other hand, most used convenience criterion samples, which risks underestimating accuracy because of mismatch between stereotype and criterion, although the one study (Chan et al., 2012) that included both haphazard and representative samples had similar patterns of accuracy. Furthermore, far more work has examined consensual than personal stereotypes.

Although the *exaggeration* hypothesis has long and venerable roots in social psychology (e.g., Allport, 1954; Campbell, 1967), the general stereotype accuracy literature provides little systematic support for it, except in the case of political stereotypes (Jussim et al., 2016). The research on stereotypes about personality provides weak and inconsistent support for exaggeration. One national character study found exaggeration (Terracciano et al., 2005; others did not assess it); the one study of age stereotypes of personality also found exaggeration (Chan et al., 2012). Studies of gender stereotypes of personality (with larger and more diverse samples, and better match of stereotype to criteria) have tended not to find exaggeration (Löckenhoff et al., 2014; Swim, 1994), though an early study using mismatched samples did find exaggeration (Martin, 1987).

There was also a modest tendency toward exaggeration in the study of stereotypes of ballet versus modern dancers (Clabaugh & Morling, 2004).

Some Speculative Hypotheses

In this section, we offer three speculative hypotheses regarding stereotypes about personality that are inspired by the existing pattern of findings. We say "inspired by" rather than "empirically demonstrated" because, though there are data consistent with these hypotheses, our view is that the state of the data is currently too thin and uncertain to declare these hypotheses confirmed. As such, testing these hypotheses in future studies would be valuable. We refer to these as "wisdom of crowds," "stereotypes as knowledge," and the "more accuracy for stereotypes of differences" hypotheses, and discuss each next.

Wisdom of Crowds

When both consensual and personal stereotypes are assessed, both in the literature reviewed here and in the general stereotype accuracy literature, consensual stereotype accuracy is usually higher. This is almost certainly a wisdom of crowds effect, whereby combining independent judgments or predictions is well-established at producing higher accuracy than nearly all of the individual judgments (e.g., Surowiecki, 2004; see Jussim, 2012, for an extended application to stereotype accuracy). Evidence of accuracy strongly

suggets that a major source of at ↳ least some stereotypes is *social reality*—people's beliefs about groups are often strongly linked to what those groups are actually like. The simplest explanation for this is that sometimes people learn quite a lot about what many groups are like.

Stereotypes as Knowledge Hypothesis

Wisdom of crowds effects are also consistent with a *stereotypes as knowledge* hypothesis. Wisdom of crowds effects can only occur if there is some element of validity in most people's judgments. If stereotypes are (at least in part) a form of everyday knowledge, then more informed people should hold more accurate stereotypes. Consistent with this, stereotypes held by multicultural migrants (ethnic Finns who had lived in Russia but who had migrated back to Finland; Russian emigrants to Finland) were considerably more accurate than many other national stereotypes (Lönnqvist et al., 2012). This increased accuracy might have occurred because the migrants had more direct experience with—and therefore, increased knowledge of—both cultures. Outside of personality, other research has found increased racial stereotype accuracy among the highly educated (Kaplowitz et al., 2003) and among those encouraged to recognize, as opposed to ignore, group differences (Wolsko et al., 2000).

The flip side of *stereotypes as knowledge* is that, absent knowledge, stereotypes are likely to be inaccurate. This may explain the pervasive inaccuracy of national character stereotypes of personality (as measured by the Big Five), inasmuch as most people have little direct contact with many individuals from lots of other nations. In contrast, most people do have extensive experience with men and women, and with the young and old—which may help explain the accuracy found by so many studies of gender stereotypes (Hall & Carter, 1999; Löckenhoff et al., 2014; McCauley et al., 1988), and age stereotypes (Chan et al., 2012).

More Accuracy for Stereotypes of Differences

Most evidence for inaccuracy comes from the work on national character stereotypes, and much (though not all) of that is derived from work on autostereotypes when compared against self-report criteria such as the NEO-PI-R/Big Five. The reference group effect and lack of knowledge about people in distant cultures may therefore at least partially account for low levels of accuracy in such autostereotypes. In contrast, heterostereotypes (those of other groups) of personality, especially perceptions of *differences*, often moderately to strongly correlated with real differences (age: Chan et al., 2012; the rich: Leckelt et al., 2019; gender: Allen, 1995, Löckenhoff et al., 2014, Martin, 1987; ballet dancers: Clabaugh & Morling, 2004). It is distinctly possible that people are better at detecting *differences between groups* than they are at detecting the *absolute levels of traits* within groups. We know of no obvious theoretical basis for predicting this pattern. Thus, it would be useful for future research to develop or identify relevant theories making alternative predictions, and then tested as an a priori (and preferably preregistered) hypothesis.

Conclusion

Our review indicates that stereotypes of personality are a mixed bag. Most, but not all, of the evidence regarding national character stereotypes shows low accuracy. Much, but not all, of the rest of the evidence regarding gender, age, and other stereotypes, shows high correlational accuracy and more mixed results regarding discrepancies. Although these mixed results provide no basis for supporting the blanket emphasis on inaccuracy and exaggeration that once was common in the stereotype literature (see Jussim, 2012, for a review), they do provide a strong justification for the need for additional research. Stereotype (in)accuracy goes to the heart of several long-standing major issues in psychology: the (ir)rationality of judgment; the extent to which social beliefs and attitudes create versus reflect social realities; and the role of cognition in discrimination. Research on stereotype accuracy regarding personality has already contributed to these areas by its very mixed nature, thereby providing no easy support for absolutist views of social judgment either as fundamentally flawed or as fundamentally sound. For decades, the error and bias paradigm was dominant in social cognition (e.g., Fiske & Taylor, 1991; Ross, Lepper, & Ward, 2010). Testaments to the power of error and bias are legion (see Gigerenzer, 2018; Jussim, 2012, for reviews). Nonetheless, there has been a slow-moving revolution in both cognitive (Gigerenzer, 2018) and social psychology (Jussim, 2012), the latter driven in part by research repeatedly showing some, and sometimes a great deal, of accuracy in stereotypes (the present review; see also reviews by Hall & Goh, 2017; Jussim et al., 2016; Jussim et al., 2015). Some of the strongest evidence for accuracy occurred for demographic stereotypes, such as age and sex (and, outside of personality, also for race and ethnicity, see Jussim et al., 2016). This strongly suggests that, p. 256 despite whatever truth there is to the claim that stereotypes are ↳ cognitive sources of prejudice and discrimination (Fiske & Neuberg, 1990), a simplistic storyline of bad and irrational stereotypes producing nothing but bias is no longer justified. It remains possible, that, in the fullness of time, as the results of many more studies of stereotype accuracy of personality and other attributes come in, they will more clearly support one view or the other. For now, however, such strong conclusions would be premature.

References

Allen, B. P. (1995). Gender stereotypes are not accurate: A replication of Martin (1987) using diagnostic vs. self-report and behavioral criteria. *Sex Roles,* 32, 583–600. http://dx.doi.org/10.1007/BF01544213
Google Scholar   WorldCat

Allik, J., Alyamkina, E., & Meshcheryakov, B. (2015). The personality stereotypes of three cohabiting ethnic groups: Erzians, Mokshans, and Russians. *Cross-Cultural Research*, *49*(2), 111–134. https://doi.org/10.1177/1069397114540861
Google Scholar   WorldCat

Allport, G. W. (1954/1979). *The nature of prejudice*. Reading, MA: Addison-Wesley.
Google Scholar   Google Preview   WorldCat   COPAC

Ashmore, R. D., & Del Boca, F. K. (1981). Conceptual approaches to stereotypes and stereotyping. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 1–35). Hillsdale, NJ: Erlbaum.
Google Scholar   Google Preview   WorldCat   COPAC

Bem, S. L. (1981). *The Bem Sex-Role Inventory: A professional manual*. Palo Alto, CA: Consulting Psychologists Press.
Google Scholar   Google Preview   WorldCat   COPAC

Campbell, D. T. (1967). Stereotypes and the perception of group differences. *American Psychologist*, 22, 817–829. http://dx.doi.org/10.1037/h0025079
Google Scholar   WorldCat

Chan, W., McCrae, R. R., De Fruyt, F., Jussim, L., Löckenhoff, C., De Bolle, M., & Terracciano, A. (2012). Stereotypes of age differences in personality traits: Universal and accurate? *Journal of Personality and Social Psychology,* 103, 1050–1066. https://doi.org/10.1037/a0029712
Google Scholar   WorldCat

Clabaugh, A., & Morling, B. (2004). Stereotype accuracy of ballet and modern dancers. *Journal of Social Psychology,* 144, 31–48. https://doi.org/10.3200/SOCP.144.1.31-48
Google Scholar   WorldCat

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York, NY: Academic Press.
Google Scholar   Google Preview   WorldCat   COPAC

Costa, P. T., Jr., & McCrae, R. R. (2008). The revised NEO Personality Inventory (NEO-PI-R). In G. J. Boyle, G. Matthews, & D. H. Salclofske (Eds.), *The Sage handbook of personality theory, and assessment: Personality measurement and testing* (Vol. 2, pp. 179–198). London: Sage Publications. http://dx.doi.org/10.4135/9781849200479.n9
Google Scholar   Google Preview   WorldCat   COPAC

Costa, P. T., 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. http://dx.doi.org/10.1037/0022-3514.81.2.322
Google Scholar   WorldCat

Ellemers, N. (2018). Gender stereotypes. *Annual Review of Psychology*, *69*(1), 275–298. https://doi.org/10.1146/annurev-psych-122216-011719
Google Scholar   WorldCat

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

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

Gigerenzer, G. (2018). The bias bias in behavioral economics. *Review of Behavioral Economics,* 5, 303–336.
http://dx.doi.org/10.1561/105.00000092
Google Scholar   WorldCat

Hall, J. A., & Carter, J. D. (1999). Gender-stereotype accuracy as an individual difference. *Journal of Personality and Social Psychology,* 77, 350–359. http://dx.doi.org/10.1037/0022-3514.77.2.350
Google Scholar   WorldCat

Hall, J. A., & Goh, J. X. (2017). Studying stereotype accuracy from an integrative social-personality perspective. *Social and Personality Psychology Compass*, *11*(11), e12357. https://doi.org/10.1111/spc3.12357

Heine, S. J., Buchtel, E. E., & Norenzayan, A. (2008). What do cross-nation comparisons of personality traits tell us? The case of conscientiousness. *Psychological Science,* 19, 309–313. https://doi.org/10.1111/j.1467-9280.2008.02085.x
Google Scholar   WorldCat

Hrebickova, M., & Graf, S. (2014). Accuracy of national character stereotypes in central Europe: Outgroups are not better than ingroup in considering personality traits of real people. *European Journal of Personality,* 28, 60–72.
https://doi.org/10.1002/per.1904
Google Scholar   WorldCat

Hyde, J. S. (2014). Gender similarities and differences. *Annual Review of Psychology,* 65, 373–398.
https://doi.org/10.1146/annurev-psych-010213-115057
Google Scholar   WorldCat

Judd, C. M., & Park, B. (1993). Definition and assessment of accuracy in social stereotypes. *Psychological Review*, *100*, 109–128.
http://dx.doi.org/10.1037/0033-295X.100.1.109
Google Scholar   WorldCat

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

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

Jussim, L., Crawford, J. T., Anglin, S. M., Chambers, J. R., Stevens, S. T., & Cohen, F. (2016). Stereotype accuracy: One of the largest and most replicable effects in all of social psychology. In T. Nelson (Ed.), *Handbook of prejudice, stereotyping, and discrimination*, (2nd ed., pp. 31–63). New York, NY: Psychology Press.
Google Scholar   Google Preview   WorldCat   COPAC

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. https://doi.org/10.1177/0963721415605257
Google Scholar   WorldCat

Jussim, L., Stevens, S. T., & Honeycutt, N. (2018). Unasked questions about stereotype accuracy. *Archives of Scientific Psychology*, 6, 214–229. http://dx.doi.org/10.1037/arc0000055
Google Scholar   WorldCat

Kaplowitz, S. A., Fisher, B. J., & Broman, C. I. (2003). How accurate are perceptions of social statistics about blacks and whites? *Public Opinion Quarterly,* 67, 237–243. https://doi.org/10.1086/374400
Google Scholar   WorldCat

Leckelt, M., Richter, D., Schroder, C., Kufner, A. C. P., Grabka, M. M., & Back, M. D. (2019). The rich are different: Unraveling the perceived and self-reported personality profiles of high net-worth individuals. *British Journal of Psychology*, *110*(4), 669–689. https://doi.org/10.1111/bjop.12360
WorldCat

Lee, Y. T., Jussim, L., & McCauley, C. R. (Eds.). (1995). *Stereotype accuracy: Toward appreciating group differences*. Washington, DC: American Psychological Association.
Google Scholar   Google Preview   WorldCat   COPAC

Lippmann, W. (1991). *Public opinion*. Piscataway, NJ: Transaction Publishers. First published 1922.
Google Scholar   Google Preview   WorldCat   COPAC

p. 257 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. https://doi.org/10.1177/0022022113520075

Lönnqvist, J.-E, Yijälä, A., Jasinskaja-Lahti, I., & Verkasalo, M, (2012). Accuracy and contrast in national value stereotypes—A case study using Ingrian-Finns as bi-cultural experts. *International Journal of Intercultural Relations,* 36, 271–278. https://doi.org/10.1016/j.ijintrel.2011.08.002
Google Scholar   WorldCat

Mackie, M. (1973). Arriving at "truth" by definition: The case of stereotype inaccuracy. *Social Problems*, *20*, 431–447. https://doi.org/10.2307/799706

Martin, C. L. (1987). A ratio measure of sex stereotyping. *Journal of Personality and Social Psychology,* 52, 489–499. http://dx.doi.org/10.1037/0022-3514.52.3.489
Google Scholar   WorldCat

McCauley, C. R. (1995). Are stereotypes exaggerated? A sampling of racial, gender, academic, occupational, and political stereotypes. In Y. T. Lee, L. Jussim, & C. R. McCauley (Eds.), *Stereotype accuracy: Toward appreciating group differences* (pp. 215–243). Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/10495-009
Google Scholar   Google Preview   WorldCat   COPAC

McCauley, C., & Stitt, C. L. (1978). An individual and quantitative measure of stereotypes. *Journal of Personality and Social Psychology,* 36(9), 929–940. https://doi.org/10.1037/0022-3514.36.9.929
Google Scholar   WorldCat

McCauley, C., Thangavelu, K., & Rozin, P. (1988). Sex stereotyping of occupations in relation to television representations and census facts. *Basic and Applied Social Psychology*, *9*, 197–212. https://doi.org/10.1207/s15324834basp0903_3
Google Scholar   WorldCat

McCrae, R. R., & Allik, J. (Eds.). (2002). *The five-factor model of personality across cultures*. New York, NY: Kluwer Academic/Plenum Publishers. https://doi.org/10.1007/978-1-4615-0763-5_2
Google Scholar   Google Preview   WorldCat   COPAC

McCrae, R. R., Chan, W., Jussim, L., De Fruyt, F., Löckenhoff, C. E., De Bolle, M., & Avdeyeva, T. V. (2013). The inaccuracy of national character stereotypes. *Journal of Research in Personality,* 47, 831–842. https://doi.org/10.1016/j.jrp.2013.08.006
Google Scholar   WorldCat

McCrae, R. R., & Terracciano, A. (2005). Universal features of personality traits from the observer's perspective: Data from 50 cultures. *Journal of Personality and Social Psychology*, *88*, 547–561. https://doi.org/10.1037/0022-3514.88.3.547
Google Scholar   WorldCat

Realo, A., Allik, J., Lonnqvist, J., Verkasalo, M., Kwiatkowska, A., Koots, L., & Renge, V. (2009). Mechanisms of the national character stereotype: How people in six neighboring countries of Russia describe themselves and the typical Russian. *European Journal of Personality,* 23, 229–249. http://dx.doi.org/10.1002/per.719

Google Scholar   WorldCat

Richard, F. D., Bond, C. F., Jr., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology,* 7, 331–363. http://dx.doi.org/10.1037/1089-2680.7.4.331

Google Scholar   WorldCat

Rogers, K. H., & Wood, D. (2010). Accuracy of United States regional personality stereotypes. *Journal of Research in Personality,* 44, 704–713. https://doi.org/10.1016/j.jrp.2010.09.006

Google Scholar   WorldCat

Rosenthal, R. (1991). Effect sizes: Pearson's correlation, its display via the BESD, and alternative indices. *American Psychologist,* 46, 1086–1087. http://dx.doi.org/10.1037/0003-066X.46.10.1086

Google Scholar   WorldCat

Ross, L. D., Lepper, M., & Ward, A. (2010) History of social psychology: Insights, challenges, and contributions to theory and application. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (5th ed., Vol. 1, pp. 3–50). Hoboken, NJ: Wiley. https://doi.org/10.1002/9780470561119.socpsy001001

Google Scholar   Google Preview   WorldCat   COPAC

Ryan, C. S. (2002) Stereotype accuracy. *European Review of Social Psychology*, *13*, 75–109. https://doi.org/10.1080/10463280240000037

Google Scholar   WorldCat

Spence, J. X., Helmreich, R. L., & Holahan, C. K. (1979). Negative and positive components of psychological masculinity and femininity and their relationships to self-reports of neurotic and acting out behaviors. *Journal of Personality and Social Psychology,* 37, 1673–1682. http://dx.doi.org/10.1037/0022-3514.37.10.1673

Google Scholar   WorldCat

Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations*. New York, NY: Doubleday.

Google Scholar   Google Preview   WorldCat   COPAC

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. http://dx.doi.org/10.1037/0022-3514.66.1.21

Google Scholar   WorldCat

Terracciano, A., Abdel-Khalek, A. M., Adam, N., Adamova, L., Ahn, C., Ahn, H. N., & Meshcheriakov, B. (2005). National character does not reflect mean personality trait levels in 49 cultures. *Science,* 310, 96–100. https://dx.doi.org/10.1126%2Fscience.1117199

Google Scholar   WorldCat

Wolsko, C., Park, B., Judd, C. M., & Wittenbrink, B. (2000). Framing interethnic ideology: Effects of multicultural and color-blind perspectives on judgments of groups and individuals. *Journal of Personality and Social Psychology*, *78*, 635–654. http://dx.doi.org/10.1037/0022-3514.78.4.635

Google Scholar   WorldCat