RUNNING HEAD: THE ACCURACY OF DEMOGRAPHIC STEREOTYPES

The Accuracy of Demographic Stereotypes

Lee Jussim

Rutgers, The State University of New Jersey – New Brunswick

Sean T. Stevens

New York University – Stern School of Business

Nathan Honeycutt

Rutgers, The State University of New Jersey – New Brunswick

The article was published in Chinese, in *The Journal of South China Normal University, May 2019, pp 5-18.* This is the English version.

Address correspondence to:

Lee Jussim

Rutgers, The State University of New Jersey – New Brunswick

Tillett Hall, Room 619

53 Avenue E

Piscataway, NJ, 08854

jussim@rci.rutgers.edu

**Abstract**

 This paper reviews evidence on the accuracy of demographic stereotypes. Doing so requires first discussing how to define stereotypes and how to assess their accuracy. Evidence is then reviewed showing that race, gender, and age stereotypes tend to be moderately to highly accurate. Recent work on “implicit bias” also shows that implicit stereotypes primarily reflect social realities rather than prejudices. Evidence on person perception is also reviewed showing that people tend to rely on the most diagnostic information when judging others and, in general, that means relying on individuating information. Nonetheless, relying on a stereotype has generally been found to increase rather than reduce the accuracy of person perception. Because prior theory and research has focused primarily on sources of inaccuracy in stereotypes, it seems to have largely missed the mark. Therefore, we also review research on several processes that might help explain these relatively high levels of stereotype accuracy. Last, limitations to existing research on stereotype accuracy are discussed, as are directions for future research.

For nearly 100 years, psychologists and other social scientists have declared stereotypes inaccurate (e.g., Aronson, 2011; Lipmann, 1922). Stereotypes have been variously described as rigid (Lippmann, 1922), rationalizations of prejudice and inequality (Jost & Banaji, 1994; La Piere, 1936), exaggerations based on small “kernels of truth” (Allport, 1954/1979), and out of touch with reality (Bargh & Chartrand, 1999). In this paper, we explain why, logically such claims are not justified, and we review the now considerable empirical evidence demonstrating accuracy in stereotypes regarding race, gender, and age (for a review of the evidence regarding other stereotypes, see Jussim, Crawford, Anglin, Chambers, Stevens & Cohen, 2016; Jussim, Crawford & Rubinstein, 2015b). We focus on these kinds of stereotypes to limit the scope of the present review and because these sorts of demographic stereotypes lie at the heart of many of the social justice concerns that have motivated much research.

**What is a Stereotype?**

 **Prior definitions.** Many prior definitions declare that stereotypes are inaccurate (e.g., Aronson, 2008; Bargh & Chartrand, 1999; Lippmann, 1922) or, at least, that they are exaggerations of real differences (Allport, 1954/1979; Campbell, 1967). In an early broad-ranging review that highlighted the lack of data that, at the time, had actually addressed the stereotype accuracy question, Brigham (1971, p. 31) defined stereotypes as “An ethnic stereotype is a generalization made about an ethnic group, concerning a trait attribution, which is considered to be unjustified by an observer.”

 This is a terrific definition of the phenomenology of how people *use* the term “stereotypes,” which is often to condemn or denounce some belief about groups as reflecting bias, prejudice, or bigotry. Indeed, as we have argued elsewhere, even when scientists define stereotypes in a neutral manner, they often bring an implicit assumption of inaccuracy in through an intellectual back door. This can easily be accomplished by starting with a neutral definition, and then *only* discussing research that demonstrates bias (see Jussim, 2012; Jussim et al., 2015b for reviews and concrete examples).

 There is, however, a more serious problem with this definition. It is purely subjective. As a result, the exact same phenomenon can both be and not be a stereotype. This would occur whenever one perceiver considers the generalization about an ethnic group to be unjustified (and therefore a stereotype) while another considers the same generalization justified (and therefore not a stereotype). Although as a phenomenological description this works just fine, as a scientific definition, it fails, because science cannot define its constructs in this sort of logically incoherent manner, a manner that does not fully consider the empirical evidence about a phenomenon, and wherein some phenomena simultaneously is and is not an example of that construct.

 **The black hole problem.** There is a second problem with many prior statements about the inaccuracy of stereotypes. Often, it is not clear whether they are meant to be definitions or *empirical statements.* This difference is important. Something that is true by definition is inherently true because of that definition, not because of empirical evidence. “Blood” is the red fluid that courses in our veins, not because surveys show that most people have red fluid coursing through their veins, but because the characteristics of that fluid is what defines “blood.” In contrast, the claim “men are, on average, taller than women” is not true by definition. It is an empirical statement based on comparing the average height of men and women. All men are not taller than all women, and it is possible to imagine a future where this example reverses (e.g., there are many species where females are larger than males).

 As empirical statements, however, claims that “stereotypes are inaccurate” almost always suffer from what we call the “black hole” problem (Jussim et al, 2016). That is, nearly all fit either of two patterns, both of which violate one of the most fundamental norms of scientific scholarship – that a statement of fact requires supporting empirical evidence. Although this is so obvious that it should go without saying, this norm has been routinely violated with respect to claims about the alleged “inaccuracy” of stereotypes. Many cite no evidence whatsoever in support of the claim. For example:

*"... stereotypes are maladaptive forms of categories because their content does not correspond to what is going on in the environment" (Bargh & Chartrand, 1999, p. 467).”*

There is simply no reference to anything empirical here, a pattern common to such claims (e.g., Allport, 1954; Aronson, 2008; see Jussim et al., 2015b for a review).

 The second black hole pattern does involve including a citation – to some paper that may *declare* stereotypes to be inaccurate, but this citation neither provides nor reviews evidence of inaccuracy. For example:

"The term *stereotype* refers to those interpersonal beliefs and expectancies that are both widely shared and generally invalid (Ashmore & Del Boca, 1981)." (Miller & Turnbull, 1986, p. 233, [*Annual Review of Psychology*]).

There is a citation here – to Ashmore and Del Boca (1981). Although Ashmore and Del Boca (1981) did review how prior researchers *defined* stereotypes, they did not review or provide empirical evidence that addressed the accuracy of stereotypes. Furthermore, they concluded that the only part of the definition that was broadly shared was that stereotypes were “…beliefs about the personal attributes of a social group” (p. 21). Thus, the Miller and Turnbull (1986) quote also ends in an empirical black hole, a pattern that characterizes many other papers that do include citations (e.g., American Psychological Association, 1991; Heilman, 1983).

 **Logical issues in defining stereotypes.** However, perhaps we are wrong in interpreting such statements as empirical. Perhaps, instead, they were *definitions*.One does not need empirical evidence to support a definition, so, in that case, there would be no violation of any scientific standard.

 But there is a problem here of a different type. Scholars have rarely, if ever, stated precisely what it means to define stereotypes as inaccurate. One possibility is that this definition means that all beliefs about groups are stereotypes (and all are therefore inaccurate). Another is that it means that only inaccurate beliefs about groups are stereotypes; accurate beliefs about groups exist, but they are not stereotypes. The problem is that, either way, this definition is logically incoherent.

 Consider the first meaning. All beliefs about all groups cannot possibly be inaccurate. That is because it would be inaccurate to believe two groups are different, and it would be inaccurate to believe they are not different. It is impossible for both statements to be true. Therefore, this definition can be rejected out of hand.

 Consider the second definition. It is logically possible to define inaccurate beliefs about groups as stereotypes and, therefore, to not “count” accurate beliefs about groups as stereotypes. However, if “inaccuracy” is a defining feature of a stereotype, then one cannot know whether one is studying a stereotype, unless one first shows that it is inaccurate. We are not aware of any published research claiming to study “stereotypes” that has first established the inaccuracy of the belief under study. By this standard, therefore, there is no research on any known “stereotypes,” because none has first established inaccuracy. Consequently, according to this definition, the entire body of social science literature purporting to address stereotypes, and their inaccuracy, would need to be discarded. Put differently, it is logically incoherent to define stereotypes this way, and then refer to any specific evidence as addressing “stereotypes,” because none of it first established inaccuracy.

 **A neutral definition.** Avoiding these logical and empirical problems is simple. “Stereotypes” cannot be defined subjectively (as did Brigham, 1971); and they cannot be *defined* as inaccurate. Our working definition, therefore, is adapted from that of Ashmore & Del Boca (1981): A stereotype is a belief about a group. This simple, neutral definition allows stereotypes to be either accurate, inaccurate, or of unknown accuracy. It allows stereotypes to include beliefs of unknowable accuracy or for which there is no possibility of accuracy (e.g., prescriptive beliefs involving how different groups “should” or “should not” behave). It permits but does not require stereotypes to lead to biases, self-fulfilling prophecies, and any of a myriad of social evils (or goods).

 This neutral definition eliminates problems of logical incoherence. But once stereotypes are no longer presumed to be inaccurate by definition, it raises an empirical question: How (in)accurate are stereotypes? Before we can even begin to review the evidence that bears on this question, however, we need to address two prior questions: What is social perceptual accuracy and how can it be assessed?

**What is Stereotype Accuracy?**

**Descriptive beliefs can be accurate; prescriptive ones cannot.**  Given what we have reviewed so far, it follows that only descriptive or predictive beliefs about a group can be assessed for their accuracy.  The accuracy of a belief such as “rich Americans vote for Republicans” can be evaluated – by analyzing the vote totals for Republican and non-Republican candidates and considering an individual voter’s socioeconomic status. There are, however, no accuracy criteria for stereotypes that prescribe or proscribe behaviors. “Men should not wear skirts,” and “children should be seen and not heard” are examples of such stereotypes. Their accuracy cannot be evaluated because there is no objective criterion (or criteria) to assess, they are simply matters of opinion. Such stereotypes merely express the stereotype-holder’s preferences or moral values. One can agree or disagree, but there is no basis for characterizing them as either accurate or inaccurate.

**Criteria.** Although establishing criteria to assess the accuracy of an individual’s beliefs regarding physical events, such as the amount of rainfall or the speed of a moving vehicle, is fairly easy, identifying the criteria for establishing accuracy in social perception is more challenging.  This is because the criteria for evaluating beliefs *about people* are often not very clear or objective.  How do we evaluate friendliness or intelligence?  However, over the last several decades, rigorous methods for assessing intelligence, personality, and many other characteristics have been administered, and even meta-analyzed (e.g., Hyde, 2005). Thus, there are now strong scientific criteria for many group differences, such as nationally representative surveys, meta-analyses, and national census data.

 **Types of accuracy.** Accuracy is a multidimensional construct (e.g., Judd & Park, 1993; Kenny, 1994). Stereotype accuracy has been commonly assessed in either of two ways in the scientific literature. *Discrepancy scores* assess how close to perfection people’s beliefs come. If the average woman’s height is 5 feet 6 inches, and a stereotype belief is 5 feet 5 inches, the stereotype is imperfect; it is off by an inch. Discrepancy scores assess precisely how close or distant people’s beliefs are to perfectly accurate as indicated by one’s criteria. Thus, we sometimes refer to these as “*discrepancies from perfection”* because *any* discrepancy, however minute, will not be a perfect bull’s-eye.

 Research on stereotype accuracy has also used *correlations* to assess how well people’s beliefs about groups correspond to what those groups are like, in a given environment or social context. Stereotype beliefs can be correlated with criteria (e.g., people’s ratings of women’s average height, wealth, and aggressiveness, could be correlated with criteria for women’s height, wealth, and aggressiveness). Higher correlations indicate greater correspondence of the stereotype with criteria – i.e., higher accuracy (for a review, see Jussim, 2012).

Discrepancy scores and correlations have both been used to assess two types of stereotypes: consensual and personal stereotypes. Consensual (or aggregate) stereotypes refer to the extent to which a stereotype is shared by the members of a culture, or a particular sample, and are usually assessed by 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). Personal stereotypes are simply any individual’s beliefs about a group, regardless of whether that belief is shared by others. Thus, our empirical review presents results for four aspects of stereotype (in)accuracy: consensual discrepancies, personal discrepancies, consensual correlations, and personal correlations.

 **Standards for accuracy and inaccuracy.**  Statistical significance provides essentially no information about accuracy, because: 1) even tiny discrepancies will significantly differ from zero if the sample is large enough; and 2) even very large discrepancies may not significantly differ from zero, if the sample is small enough. Accuracy should be a function of closeness to criterion, not sample size.

 Therefore, for discrepancy scores, we consider judgments that are within 10% or .25 standard deviations of the criteria to be accurate, and within 20% and .50 standard deviations to be near misses (for a more in-depth discussion of these cutoff points, see Jussim, 2012). Judgments more than 20% or .50 standard deviations (SDs) off are considered inaccurate. So, for example, if 50% of Whites score above 520 on the verbal SAT, we would characterize a stereotype estimate of anywhere between 40% and 60% as accurate, stereotype estimates of between 30-39% and between 61-70% as near misses, and anything below 30% or above 70% as inaccurate (see Jussim, 2012; Jussim et al. 2009 for extended discussions and justifications for these standards).

For correlational accuracy, Jussim (2012), drawing heavily on Rosenthal and Rubin’s (1982) binomial effect size display, Cohen’s (1988) standards for small, medium and large effect sizes, and common sense, suggested that a high degree of correspondence between a belief and reality is indicated by a correlation of .40 or higher.  Moderate correspondence is indicated by a correlation between .30 and .40.  The former would correspond to being right about 70% of the time, while the latter would be accurate about two-thirds of the time.  In other words, accuracy is not an all or none phenomenon, and instead is usually a matter of degree.

**Level of Analysis**

 When assessing accuracy, the stereotype must be assessed at the same level of analysis as the criterion. Violation of this principle has led to a great deal of unnecessary confusion about what constitutes accuracy and inaccuracy in stereotypes and stereotyping. For example, this sort of analysis has periodically appeared in the social psychological literature (e.g., Allport, 1954; APA, 1991; Fiske, 1998; Hamilton, et al., 1990; Stangor, 1995):

***Even if it can be successfully shown that perceivers accurately judge two groups to differ on some attribute:***

***1. Perceivers cannot assume that their stereotypes of the group automatically fit all members of the group;***

***2. Perceivers cannot apply their belief about the group when judging individuals;***

***3. If perceivers do apply their belief about the group when judging individuals, they are likely to be wrong much of the time because few members perfectly fit the stereotype.***

 This criticism has some validity, but that validity depends, in part on what this type of statement means. To the extent that the “perceivers cannot” statements represent moral injunctions, rather than statements about accuracy, they are beyond the scope of a consideration of stereotype accuracy. However, if “perceivers cannot” means “they would reach inaccurate judgments if they did” these arguments are a central focus of this paper.

 This line of reasoning’s suggestion – that all stereotypes are inaccurate because most members of a group fail to fit a stereotype – is partially justified. It is true that most members of a group will fail to perfectly fit a stereotype. This, however, does not mean that the stereotype is inaccurate. To understand why requires understanding how this reasoning confounds two different levels of analysis and how considerably greater conceptual clarity can be brought to understanding stereotype accuracy by clearly distinguishing among these levels of analysis. Table 1 presents an analytic breakdown of different levels of analysis at which accuracy can be assessed.

Table 1

Identifying the Appropriate Level of Analysis in Studies of Social Perceptual Accuracy

|  |  |  |
| --- | --- | --- |
| Level of Analysis:  | Social Belief is a: | Level of Criteria for Assessing Accuracy of That Social Belief: |
| POPULATIONThis level assesses the accuracy of a stereotype about a group.Research examples:Judd, et al, 1995McCauley & Stitt, 1978Swim, 1994 | STEREOTYPE REGARDING AN ENTIRE POPULATIONExamples:1. An introductory psychology student believes that White Americans are wealthier than African-Americans. 2. A high school teacher believes that teenage boys are better at math than are teenage girls. | POPULATION 1. Income of White Americans and African-Americans in a nationally representative sample or in the U.S. Census. 2. Meta-analyses of hundreds of studies assessing sex differences in teenagers’ math performance. |
| PERSON PERCEPTIONThis level assesses the accuracy of beliefs about differences between specific individual targets belonging to different groups. This corresponds to what is frequently termed “stereotypes and person perception.”Research examples:Brodt & Ross, 1998Clarke & Campbell, 1955Madon, et al, 1998 | PERCEPTION OF DIFFERENCES BETWEEN SPECIFIC INDIVIDUAL MEMBERS OF SOCIAL GROUPSExamples:1. An introductory psychology student sees little difference between the wealth of African-American and White students in his class. 2. A high school teacher believes the girls in her class are doing better at math than are the boys in her class. | INDIVIDUALS1. The wealth (net worth; yearly income) of the African-American and White students in that student’s introductory psychology class. 2. Performance in class and on standardized tests of the boys and girls in this teacher’s class. |

 **Stereotypes as perceptions of populations.** The first row in Table 1 refers to stereotypes: beliefs (or generalizations) about whole populations (typically, but not always, large demographic groups). The level at which one must measure the criterion for assessing the accuracy of beliefs about groups is the population that comprises that group. Claims about the characteristics of New Yorkers (or women or African-Americans or librarians) should be compared to the characteristics of a representative sample or the whole population of New Yorkers (or women or African-Americans or librarians, respectively). It is not possible to evaluate the accuracy of a belief about Asians in general by using as a criterion the characteristics of a particular Asian target. To do so would be equivalent to evaluating the claim that “Alaska is cold” by measuring the temperature at noon on July 4th in Anchorage.

 Census figures, results from randomly selected samples, and meta-analyses of hundreds of studies have all been justifiably used as criteria against which to compare the accuracy of people’s stereotypes. However, such research cannot, and was never intended, to evaluate the accuracy of people’s perceptions of individuals from different groups, which requires a level of analysis below that of whole populations.

 **Stereotypes and person perception.** The second row of Table 1 presents a second level of analysis for assessing accuracy: That of perceptions of differences between individuals belonging to different groups. This is generally referred to in the scholarly literature as “stereotypes and person perception,” and is represented by studies that have people rate one or more individuals belonging to different social groups.

 Let’s say a fourth-grade teacher assigns higher grades to the girls than to the boys in their class. One might claim that this teacher stereotypes their girls as achieving more highly, but assigning higher grades to girls is itself not a claim about whole populations of boys and girls. Whether this is accurate or biased cannot be determined by comparison to the mean achievement of nationally representative samples of fourth grade boys and girls. Instead, determining the accuracy of this teacher’s higher grading of girls requires comparison of their ratings to the achievement of the particular girls and boys in their class. This level of analysis addresses the role of stereotypes in causing systematic inaccuracy in perceivers’ judgments about individuals they know personally. Such claims occur at a different, smaller, level of analysis than do claims about differences between whole populations.

 Assessing the accuracy of the perceived difference at this level of analysis must be accomplished by comparing the perceived mean difference between individual targets from differing groups to the actual mean difference.

 **Stereotype accuracy and level of analysis: Conclusion.** Claims suggesting that stereotypes are inaccurate because they do not apply to all individual members of a group (Allport, 1954/1979; APA, 1991; Fiske, 1998; Hamilton, et al., 1990; Stangor, 1995) are both true and false. The claim that stereotypes cannot possibly apply to all individual members of a group is completely true. The suggestion that this renders stereotypes inaccurate is, however, unjustified because it confounds levels of analysis (population and either small group, individual, or both). A claim about a population cannot be evaluated against the characteristics of an individual, or even small groups of individuals. Consistency between the level of the perception and the level of the criterion must be maintained when assessing accuracy by comparing: beliefs about populations (stereotypes) to characteristics of those population groups, or beliefs about differences between small groups of individuals to the actual differences between those small groups of individuals.

 **The one exception: Absolutist stereotypes.** Absolutist stereotypes – beliefs that all members of a group have some attribute – will indeed almost always be false, because there are almost always wide variations among individuals. A single exception invalidates an absolutist belief. Just as a belief that the temperature in all locations in Alaska is always below freezing will be disconfirmed by a single reading of 33 degrees Fahrenheit in Juneau on July 15th at 1pm, a belief that all Germans are efficient will be disconfirmed by discovery of a single inefficient German.

 The vast accumulated empirical evidence on stereotypes, however, has yet to report a single person who holds absolutist stereotypes. Instead, the evidence indicates that most stereotypes are quantitative and probabilistic, not absolute (e.g., Judd, et al., 1995; McCauley & Stitt, 1978; Swim, 1994). Probabilistic stereotypes, which permit many exceptions and wide variability, can only be evaluated by comparison to population-level criteria. People who hold absolutist stereotypes undoubtedly exist, and probably comprise significant portions of extremist groups such as the Ku Klux Klan and neo-Nazis. Nonetheless, such people are atypical of the participants in most scientific research on stereotypes.

**Accuracy of Stereotypes of Race, Gender, and Age**

 The current review is restricted to research on race, gender, and age stereotypes. A thorough review of all stereotype accuracy research would require a much longer paper (for more comprehensive reviews, see Jussim, 2012; Jussim et al., 2016). Furthermore, these sorts of demographic stereotypes typically are central to the social sciences’ longstanding concern with social problems and social justice (see e.g., Jost & Kay, 2010; Tetlock, 1994; Tetlock, Kristel, Elson, Green, & Lerner, 2000).

 We also focus only on studies with criterion samples well-matched to the stereotype under study. This is a problem well-known in the stereotype accuracy literature (e.g., Judd & Park, 1993; Jussim, 2012). If one assesses sex stereotypes in a particular country, the appropriate criteria involve characteristics of men and women as assessed either by national census data or based on nationally representative samples. Another reasonable criterion would be meta-analyses of large numbers of studies.

Sometimes, however, researchers have simply used convenience samples as their criterion (see Table 2.1 from Jussim et al., 2016). Such methods underestimate accuracy, because of mismatch between the stereotype (about people in a country), and the haphazard and unrepresentative criterion sample (see Jussim et al., 2016). Studies with mismatched stereotypes and criterion samples are not included here.

**Racial and Ethnic Stereotypes**

 Four studies assessed the accuracy of racial and ethnic stereotypes for specific judgments. Consensual discrepancies were accurate for a plurality or majority of all judgments (Ashton & Esses, 1999; Kaplowitz, Fisher, & Broman, 2003; McCauley & Stitt, 1978; Ryan, 1996). There was more evidence of underestimating than of exaggerating real differences in three studies (Kaplowitz et al., 2003; McCauley & Stitt, 1978; Wolsko, Park, Judd, & Wittenbrink, 2000). One found more evidence of exaggerating real differences (Ryan, 1996).

 Only one study reported personal discrepancies from perfection. Ashton and Esses (1999) found that, by our standards, 36 participants’ stereotypes were accurate, 33 exaggerated real differences, and 25 underestimated real differences (although they also used several other standards, which differed from ours, this pattern of slightly more exaggeration than underestimation occurred no matter what standard they used).

 The original studies rarely reported consensual correlations. However, we computed them *if* the original papers reported: 1) Sample mean estimates on different attributes; 2) The criterion score for those attributes; and 3) The variables were all commensurate (e.g., all measured as percentages or on the same 1-7 scale). It was then simple to compute consensual correlations by correlating the sample means on the stereotype judgments with the criteria (and recall that, based on the binomial effect size display, we consider *r* > .40 highly accurate and .25 < *r* < .40 moderately accurate).

 Two studies reported results from which consensual correlations could be readily computed (McCauley & Stitt, 1978, though the original means were more clearly presented in McCauley, 1995; Ryan, 1996). In Ryan’s (1996) research, the consensual stereotypes correlated from about .50 to .80 with self-report criteria. In the McCauley and Stitt (1978) study, consensual stereotypes correlated .27 to .96 with U.S. Census data (median r = .83, in this chapter, we generally report median r’s because they provide a clearer sense of just how high *typical* stereotype accuracy correlations are, whereas averages based on small numbers of correlations are vulnerable to being skewed by one or two unusual correlations).

 Two studies also assessed personal stereotype accuracy correlations. Ryan (1996) found such correlations to average about .40, whereas Ashton and Esses (1999) found them to average .69.

 In general, stereotype accuracy was *higher* when the criteria were objective (Census data and Canadian achievement data, respectively, for McCauley & Stitt, 1978 and Ashton & Esses, 1999) than when they were self-reports (Ryan, 1996). Researchers often justifiably raise concern about self-reports as criteria (Fiske, 1998; Stangor, 1995). These results suggest that the use of self-reports, which can themselves be tainted by all sorts of biases – including social desirability biases (see e.g., Tourangeau & Yan, 2007) to avoid appearing prejudiced (Mendelberg, 2001) – probably leads to underestimates of accuracy.

**Gender Stereotypes**

 Eight studies described in four papers (Briton & Hall, 1995; McCauley, Thangavelu & Rozin, 1988; McCauley & Thangavelu, 1991; Swim, 1994) examined the accuracy of gender stereotypes. Each found that a majority of consensual discrepancies were accurate. A ninth study reported 18 accurate consensual gender stereotype judgments, 21 near misses, and 9 inaccurate judgments (Beyer, 1999). No study provided evidence that inaccuracies consistently exaggerated real differences, and, as with racial stereotypes (see above), the most common pattern was *underestimating* real differences.

 Four studies reported in two papers (Cejka & Eagly, 1999; Diekman, Eagly, & Kulesa, 2002) assessed accuracy by averaging not only across participants, but across judgments, as well. The main pattern at this level of aggregation was accuracy and near misses (which were underestimations of male/female differences). One study (Hall & Carter, 1999) did not provide any information on consensual discrepancies. Across 14 studies, the consensual correlations, which were either reported or computable, ranged from .35 to .98, with all but two over .60, and a median of .79.

 Although none of the studies assessed personal stereotype discrepancies on a judgment by judgment basis, Diekman et al. (2002) did so by averaging over all judgments. The average discrepancy over all judgments was a near miss. Five studies described in three papers provided results on personal stereotype accuracy correlations (Beyer, 1999; Diekman et al., 2002; Hall & Carter, 1999). Personal stereotype accuracy correlations ranged from -.04 to .60, which makes them seem more variable than they really were. Six of eight were over .40, and the median (across studies) of the average personal stereotype accuracy correlations (within studies) was .45.

 Another study (Halpern, Straight, & Stephenson, 2011) was titled, “Beliefs about Cognitive Gender Differences: Accurate for Direction, Underestimated for Size.” Cognitive gender differences referred to academic and intellectual accomplishments of males and females (both children and adults). Using our standards for judgments of males, consensual stereotypes were accurate four times, with three near misses, and three inaccurate stereotype judgments. Consensual stereotype discrepancies for judgments of females were accurate twice, with four near misses, and four inaccurate stereotypes. Consensual stereotypes about gender *differences* were accurate five times, with one near miss, and four inaccurate stereotypes. Inaccuracies consistently *underestimated* real gender differences. Halpern et al. (2011) did not report results for personal stereotype accuracy (either discrepancies or correlations), and consensual stereotype accuracy correlations could not be computed from the data they did report.

 Another recent study was 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 Big Five personality traits. In a pattern that was generally consistent across countries, women were stereotyped as higher than men on agreeableness, conscientiousness, openness to experience, and neuroticism, and as lower than men on extraversion.

 These were then compared to self-reports on the Big Five in the same countries and to observer reports of sex differences based on prior research. For all five personality traits, consensual stereotype discrepancies were accurate, 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 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 .36 to .70, with a median of .47. The criterion samples were large, but not representative, so this study suffers from the mismatch limitation.

**Age Stereotypes**

We are aware of only one study of the accuracy of age stereotypes. Chan et al. (2012) examined the accuracy of age stereotypes regarding personality in 26 countries among over 3,000 people. Criteria were self-reports on the Big Five personality traits. Three patterns emerged: Consensual stereotype accuracy correlations were consistently very high, ranging from about .50 to .90, depending primarily on how the correlation was assessed. However, people also consistently exaggerated real differences among the young, adults, and the old, as the standard deviation of the stereotype perceptions averaged 1.3 to 1.7 times larger than that of the criterion. Last, this was one of few studies to actually assess personal stereotype accuracy correlations, which averaged .34, indicating moderate accuracy at the individual level. Interestingly, there was a great deal of consistency in these patterns across country, gender, and age of rater. Thus, these patterns appeared 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 of which we are aware to empirically demonstrate that use of convenience criterion samples did not alter the results.

**Stereotypes and Person Perception**

**Process**

 To be as accurate as possible, people should judge others on the basis of the most diagnostic information available. Although occasionally this may be stereotypes, usually it is individuating information (e.g., Crawford et al., 2011; Jussim, 1991; Rubinstein, Jussim, & Stevens, in press). This area of research has been highly controversial, with many researchers emphasizing the power of stereotypes to bias judgments (Devine, 1995; Fiske & Neuberg, 1990; Fiske & Taylor, 1991; Jones, 1986; Jost & Kruglanski, 2002) and others emphasizing the relatively modest influence of stereotypes and the relatively large role of individuating information (Jussim, Eccles, & Madon 1996; Kunda & Thagard, 1996).

 Fortunately, literally hundreds of studies have now been performed that address this issue, and, even more fortunately, multiple meta-analyses have been performed summarizing their results. Effects of stereotypes on person judgments, averaged over hundreds of experiments, range from 0 to .25 (see reviews by Jussim, 2012; Jussim et al., 2009). Averaging over the meta-analyses produces an overall estimate of the mean effect of stereotypes at about r= .10 (Jussim, 2012). The few naturalistic studies of the role of stereotypes in biasing person perception have yielded similarly small effects (e.g., Clarke & Campbell, 1955; Jussim et al., 1996; Madon et al, 1998).

 In contrast, when arriving at person perception judgments, people rely heavily on individuating information. This is one the largest average effects found in social psychology, *r* = .71 (Jussim, 2012; Kunda & Thagard, 1996; see also Richard, Bond Jr., & Stokes-Zoota, 2003). In other words, people seem to be relying on individuating information far more than stereotypes during person perception.

 But what about accuracy? Most of the studies examining these issues have examined experimentally created fictitious targets who had no “real” attributes, so that there were no criteria with which to assess accuracy. Therefore, the next section reviews the very small handful of stereotype and person perception studies that have actually addressed the accuracy issue.

**Accuracy**

 **Accuracy in perception of small group differences.**  Madon et al. (1998) examined the accuracy of 7th grade teachers’ perceptions of their students’ performance, talent, and effort at math about one month into the school year. Madon et al. assessed accuracy in the following manner. First, they identified the teachers' perceptions of group differences by correlating teachers' perceptions of individual students with the students' race, sex, and social class. This correlation indicated the extent to which teachers’ systematically evaluated individuals from one group more favorably than individuals from another group. Next, Madon et al. assessed actual group differences in performance, talent, and effort by correlating individual students’ final grades the prior year (before teachers knew the students), standardized test scores, and self-reported motivation and effort with students' race, sex, and social class. The teachers' accuracy was assessed by correlating the teachers' perceived differences between groups with the groups' actual differences.

 Madon et al. found that teachers were mostly accurate. The correlation between teachers' perceived group differences and actual group differences was *r* = .71. The one outlier was teachers’ perceptions of sex differences in effort, however, were highly inaccurate – they believed girls exerted more effort than boys, but there was no sex difference in self-reported motivation and effort. When this outlier was removed, the correlation between perceived and actual group differences increased to *r* = .96.

 We are aware of only two other studies that have addressed whether people systematically and unjustifiably favor or disparage individuals belonging to certain groups (Clarke & Campbell, 1955; Jussim et al.1996). Both yielded evidence of accuracy accompanied by small bias. All three studies (including Madon et al., 1998), however, were conducted in educational contexts – Jussim et al. (1996) addressed teachers’ perceptions of students, and Clarke and Campbell (1955) addressed students’ perceptions of one another. It remains an empirical question whether this pattern of accuracy and small bias in perceptions of demographic differences between individuals with whom one has extended contact is unique to classrooms, or characterizes social perception more broadly.

 This pattern of moderate to high accuracy in perceptions of differences between small groups can occur for either of two reasons. First, perceivers might have jettisoned their stereotypes completely, and judged targets primarily on the basis of relevant individuating information. Second, perceivers might not have jettisoned their stereotypes. If their stereotypes (e.g., “girls perform slightly higher in math classes than do boys”) were accurate (girls actually did perform slightly higher than boys), teachers could also have reached accurate perceptions of differences between boys and girls by applying their stereotype.

 The research described thus far does not distinguish between these explanations. Regardless of the explanation, however, this research does lead to one clear conclusion: In the few studies that have examined stereotypes and person perception under naturalistic conditions, there is no evidence of stereotypes powerfully and pervasively distorting social perception. There was some evidence of bias and distortion, but the far stronger pattern has been one of accuracy. The next section, therefore, reviews studies that have not only assessed accuracy, but have also assessed the sources of both accuracy and bias in person perception.

**Does Relying on a Stereotype Increase or Reduce Accuracy in Person Perception?**

 What does empirical research indicate about whether people’s reliance on stereotypes increases or reduces the accuracy of their judgments? Only a handful of studies provide data capable of addressing this issue, and they are discussed next.

 **Sex stereotypes: Jussim et al. (1996) and Madon et al. (1998).** Both Jussim et al. (1996) and Madon et al. (1998) examined the accuracy of teacher expectations (Madon et al., 1998 was described previously; Jussim et al., 1996, was similar, except that it was conducted in sixth grade rather than seventh grade, and it did not examine the accuracy of perceived differences between students from different demographic groups). Both found that, when controlling for individuating information (motivation, achievement, etc.), student social class and race/ethnicity had little or no effect on teacher expectations. Thus, teachers essentially jettisoned their social class and ethnic stereotypes when judging differences between children from different social class and ethnic backgrounds. Although this finding is in many ways laudable, teachers relying entirely on individuating information does not help address the question of whether relying on a stereotype increases or reduces accuracy.

 Both studies, however, found that sex stereotypes biased teachers’ perceptions of boys and girls performance (standardized regression coefficients of .09 and .10 for performance, and .16 and .19 for effort, for Madon et al. and Jussim et al. respectively). In both studies, teachers perceived girls as performing higher and exerting more effort than boys. Because these effects occurred in the context of models controlling for individuating information, they are best interpreted as stereotypes influencing teacher perceptions – bias effects, in traditional social psychological parlance.

 Did these sex stereotyping bias effects increase or reduce the accuracy of teachers’ perceptions? They did both. In the case of performance, the sex stereotype effect increased teacher accuracy. The real performance difference, as indicated by final grades the prior year was *r*= .08 and *r* = .10 (for the 1996 and 1998 studies, respectively, girls received slightly higher grades). The regression model producing the “biasing” effect of stereotypes yielded a “bias” that was virtually identical to the real difference. In other words:

*“The small independent effect of student sex on teacher perceptions (of performance) accounted for most of the small correlation between sex and teacher perceptions (of performance). This means that teachers apparently stereotyped girls as performing slightly higher than boys, independent of the actual slight difference in performance. However, the extent to which teachers did so corresponded reasonably well with the small sex difference in performance. In other words, teachers’ perceptions of differences between boys and girls were accurate because teachers relied on an accurate stereotype”* (Jussim et al., 1996, p. 348).

The same conclusion, of course, also characterizes the results for the 1998 study.

 On the other hand, the results regarding effort provided evidence of bias that reduced accuracy. There was no evidence that girls exerted more effort than boys. Therefore, the influence of student sex on teacher perceptions of effort – i.e., teachers’ reliance on a sex stereotype to arrive at judgments of effort – led teachers to perceive a difference where none existed. This is an empirical demonstration of something that, logically, has to be true. Relying on an *inaccurate* stereotype when judging individuals can only harm one’s accuracy.

 **Other stereotypes.** Similar results have been found with occupational stereotypes (Cohen, 1981) and stereotypes of college students living in different residences (Brodt & Ross, 1998). We are aware of no other research examining whether relying on a demographic stereotype increases or reduces the accuracy of person perception.

**Accuracy of Implicit Stereotypes**

 Implicit methods and unconscious prejudice has been a hot topic in social psychology at least since Devine’s classic paper (1989), titled “Stereotypes and prejudice: Their automatic and controlled components.” This emphasis on “implicit” beliefs and attitudes was then given another boost by Greenwald and Banaji’s (1995) paper on “implicit cognition,” and, especially, by the creation of the implicit association test (IAT, Greenwald, McGhee, & Schwartz, 1998). The IAT assessed differences in the strength of associations in memory between categories and attributes (e.g., how much Black and White are associated with smart and stupid or pleasant and unpleasant).

 This work has been most commonly framed as addressing “implicit bias,” and the early scholarship suggested implicit biases reflected unconscious prejudices and caused discrimination (e.g., Kang & Banaji, 2006). However, both early work on implicit cognition and a slow revolution in understanding the IAT suggest this emphasis was largely misplaced. For example, a review of the prior two decades of work on implicit learning (Reber, 1989, p. 219) concluded: “Implicit learning produces a tacit knowledge base that is abstract and representative of the environment … and it can be used implicitly to solve problems and make accurate decisions about novel stimulus circumstances.” Reber quite explicitly characterizes implicit learning as “accurate.”

 A full analysis of how this understanding theoretically morphed into one emphasizing bias and prejudice is beyond the scope of this paper. That latter emphasis, however, is consistent with the bias for bias that has long characterized social psychology (Jussim, 2005, 2012), and with political biases that lead the field to less intense skeptical scrutiny of narratives emphasizing the power of oppression, such as “implicit bias” (Jussim et al. 2015a).

 Nonetheless, there has been a slow-moving revolution in the understanding of exactly what is measured by the IAT and other implicit measures. Arkes & Tetlock (2004) first proposed that such measures capture cultural knowledge as much or more than individual bias. And, most recently, in a broad and sweeping reconceptualization, Payne, Vuletich, & Lundberg (2017, p. 242) conclude that, “…the reason that implicit bias is widespread in general is that the environment has a relatively constant level of disparities and systemic inequalities that repeatedly raise the accessibility of stereotype concepts.” Our view is that, although this is a vast improvement over presumptions that IAT scores reflect prejudice, it does not go far enough. Given the evidence reviewed by Reber (1989), it seems likely that implicit associations reflect the *realities of inequality in the environment at a given point in time* and not merely “activated stereotypes.”

 Recent work has shown just how this can work at the individual level (Rubinstein, et al., in press). In the absence of individuating information, people perceived a white college applicant to be smarter than a black college applicant, consistent not only with cultural stereotypes, but with large differences in the academic achievement of whites and African Americans (Ryan & Bauman, 2016). However, when people had clear information about the applicants’ high school records, race bias against the black applicant, both implicit and explicit, disappeared. Furthermore, strong versus weak high school records influenced both implicit and explicit evaluations more than did race. Although we cannot quite call this “accuracy,” because the targets were hypothetical rather than real people, the process uncovered – of judging people on their merits, when those merits were available, but relying on stereotypes without such information – strongly suggest that implicit beliefs are heavily influenced by social realities.

**What Processes Lead to Accuracy?**

For much of the last 50 years, social psychology has had an overweening emphasis on error and bias (Funder, 1987; Jussim, 2005, 2012). Combine with this with the erroneous scientific presumption that stereotypes were inaccurate, and it seems likely that many in the field believed (and many probably still do) that the abundant literature on error and bias provides a great deal of insight into stereotypes and stereotyping (e.g., Banaji & Greenwald, 2013; Fiske & Taylor, 1991, 2013; Nisbett & Ross, 1980). And if one is interested in understanding the processes that lead to bias, that literature is thoroughly well-developed in the above sources (and many others), so will not be reviewed here. Because of the (erroneous) presumption that stereotypes were largely inaccurate, social psychology has spent far less effort seeking to understand the processes that could produce accuracy in stereotypes and stereotyping. We being to address that gap by reviewing theory and research on how and why people may be likely to develop accurate stereotypes.

**Lens Model**

Brunswik’s (1952) Lens model is one of the earliest and most commonly used models for understanding social perceptual accuracy (Funder, 1995; Hall & Groh, 2017; Jussim, 2012). The idea is simple: People’s attributes (personality, competencies, values) manifest as behaviors “cues” in Lens Model parlance), and the more people are aware of and correctly utilize the cues, the more accurate they will be. When perceiving individuals, people only have a limited number of cues, and so, although person perception often is more accurate than once believed there is still ample evidence of error and bias in person perception as well (see, e.g., Funder, 1987, 1995; Jussim, 2012 for reviews). However, when a Lens Model analysis is combined with the well-known “wisdom of crowds” effect, something quite different seems to happen.

**Combining Multiple Sources of Information (Wisdom of Crowds and More)**

 **The wisdom of crowds.** Consensual accuracy correlations are quite high, often r=.70 and higher, and, as such, among the largest effects in all of social psychology (Jussim et al, 2016). Why? Most likely, this reflects a “wisdom of crowds” effect (Suriowecki, 2004; see also Payne et al., 2017).

 When people make independent judgments, they bring a broad diversity of knowledge, background, insight and intuition to bear on a problem. Some people’s knowledge and background might lead them to be fairly close, but to systematically overestimate the answer. Others’ might lead them to be fairly close but to systematically underestimate the answer. With many people, errors cancel out and all that is left is the part that is pretty close to the truth. Of course, some people’s estimates may be almost completely random, idiosyncratic, and clearly wrong. However, random, idiosyncratic errors are, by definition, just as likely to overestimate as to underestimate the real or best answer. An overestimate cancels out an underestimate, so that when you average them, the average will be closer to the truth than either estimate. When people have at least some degree of knowledge or expertise relevant to a question, despite their individual imperfections, their group judgments are frequently going to be more accurate than their individual judgments.

 This works, however, only if the average reflects the truth. So why are consensual stereotypes so valid? The only way they can become so valid is if social reality has a systematic influence on individuals’ beliefs about groups. This influence does not need to be large. If, however, social reality was completely unrelated to people’s beliefs, those beliefs, even when aggregated, would not correspond with reality. Of course, the more highly the individual beliefs correspond with reality, the (even) more highly consensual beliefs will correspond with reality. That influence may be direct, obtained through personal experience with individuals from different groups, or indirect, obtained through family socialization, the mass media, education, etc. But one way or another, social reality appears to be one of, if not, the major influence on stereotypes.

 So, the “wisdom of crowds” may help explain why consensual stereotype correlations are so stunningly high. The more important point, regardless of the explanation, is simply *that* consensual stereotype accuracy correlations are stunningly high.This result should constitute a dagger in the heart of: 1) Any modern definition of stereotypes as “inaccurate” or implicit assumption of “inaccuracy”; and 2) Any perspective suggesting that social stereotypes are generally *false cultural myths.* The *shared* component of stereotypes, rather than being some sort of false cultural myth, is not only the *most accurate component of stereotypes,* it is one of the very largest effects in all of psychology.

 **Individual accuracy and the wisdom of combining multiple sources of information.** When perceiving groups, people have potentially many different source of information: personal experience with individuals, personal experiences with groups (e.g., when visiting different neighborhoods, cultures, or countries), and a variety of social and indirect social sources (friends, news, mass media, nonfiction books, etc.). By combining information from multiple sources, especially independent sources, people may be able to reach far more accurate judgments about groups than about individuals (e.g., Suriowiecki, 2004; Tetlock & Gardner, 2015). Thus, one reason even individual level stereotypes may often be at least moderately accurate is that people combine information about groups from many different sources. Even if that information is imperfect and subject to various biases, if there is a core of reality to most of those sources of information, people’s stereotypes will likely end up at least reasonably accurate.

**Pattern Detection**

 The title of a recent paper communicates its main message: “Superior pattern detectors efficiently learn, activate, apply, and update social stereotypes” (Lick, Alter, & Freeman, 2017). And who is a superior pattern detector? People with greater intelligence. Why would pattern detection lead to more accurate learning and updating of stereotypes? Because stereotypes are, in essence, *patterns* of characteristics that covary with group memberships. Good pattern detectors (compared to bad ones) would be more likely to realize systematic ways in which groups differ from one another. This is all consistent with the “stereotypes as knowledge” hypothesis (Jussim et al., 2016), as was research showing that those with higher intelligence also held more accurate racial stereotypes (Kaplowitz, Fisher, & Broman, 2003). The research discussed previously, on how implicit associations (once believed to reflect bias and prejudice) largely reflect the realities of social inequalities, is further evidence that much of this ability to detect patterns occurs readily, possibly not even always requiring conscious awareness.

**Limitations and Directions for Future Research**

 Social psychology has focused on bias hypotheses for nearly a century, so there is more than ample theorizing with respect to how stereotypes sometimes go astray. In this section, we focus specifically on limitations to prior work with respect to accuracy, and directions for future work investigating the accuracy of stereotypes.

 Most work on stereotype accuracy has, so far, focused on North America and Europe. We therefore know very little about the (in)accuracy of stereotypes in Asia, African, and South America. To the extent that stereotypes are just like other forms of knowledge, then a natural prediction would be that stereotypes are less accurate (overall) in societies with lower levels of education. However, much of the earliest empirical research on stereotypes and intergroup perceptions from the 1940s-1960s found widespread agreement among different groups about the characteristics of particular target groups (including Arab, Chinese, Filipino, Korean, Pakistani, and Samoan groups; see Ottati & Lee, 1995, for a review). Although agreement does not guarantee accuracy, accuracy does cause agreement (if each of two perceivers accurately perceive something, they will agree). Ultimately, however, the accuracy of such stereotypes is an empirical question.

 Additional work on explicit and implicit sources of accuracy is also needed. The pattern recognition work by Lick et al. (2017) strongly suggests it, and intelligence, play important roles in stereotype accuracy. However, that research investigated experimentally-created targets, rather than real world target groups; therefore, it was not capable of directly assessing the role of pattern recognition in accuracy. Such research is needed not simply to test for the role of pattern recognition but to further address the role of social reality (rather than, or in addition to biases) in leading to stereotypes.

 Similarly, the rising tide of evidence that even stereotypical associations measured implicitly reflect social realities (Payne et al., 2017; Rubinstein et al., in press) strongly suggests that such implicit stereotypes are likely to be substantially accurate. However, this, too, is an empirical question. How strongly do implicit associations as indicated by measures such as the IAT correlate with real group differences as indicated in Census data or meta-analyses? The answer to this question will also contribute to understanding whether stereotypes, even implicit ones, mostly reflect reality or the perceiver’s own biases.

 Last, the role of stereotypes in increasing or reducing the accuracy of person perception judgments has only been assessed by a small number of studies. Although those have generally found that relying on a stereotype increases accuracy compared to not doing so, there are too few studies to justify strong conclusions. If social psychology wishes to understand the role of stereotypes in person perception, a most fundamental question is whether relying on them generally increases or reduces accuracy.

**Conclusion**

 Stereotypes regarding demographic groups in North America and Europe have, so far, proven to be far more accurate than once believed. Even implicit associations seem to reflect social reality more than prejudice. People judge others primarily on the basis of their personal characteristics, when they are known, not on stereotypes. This occurs both explicitly and implicitly. And the few studies that have examined the issue have generally found that relying on a stereotype can increase, rather than reduce, the accuracy of person perception. The historical emphasis on stereotype inaccuracy persists in many modern perspectives and requires scientific self-correction. This review has aimed to stimulate such self-correction by summarizing the extant evidence on stereotype accuracy.

References

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

American Psychological Association (1991). In the supreme court of the United States: Price Waterhouse v. Ann B. Hopkins (amicus curiae brief). *American Psychologist*, *46*, 1061–1070.

Arkes, H. R., & Tetlock, P. E. (2004). Attributions of implicit prejudice, or “Would Jesse Jackson ‘fail’ the Implicit Association Test?” *Psychological Inquiry, 15,* 257–279.

Aronson, E. (2008). *The social animal* (10th ed.). New York: Worth Publishers.

Aronson, E. (2011). *The social animal* (11th ed.). New York: Worth Publishers.

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.

Ashton, M. C., & Esses, V. M. (1999). Stereotype accuracy: Estimating the academic performance of ethnic groups. *Personality and Social Psychology Bulletin,* *25*, 225–36.

Banaji, M.R., & Greenwald, A.G. (2013). *Blindspot: Hidden biases of good people*. New York, NY: Delacorte Press

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, *54*, 462–479.

Beyer, S. (1999). The accuracy of academic gender stereotypes. *Sex Roles,* *40*, 787– 813.

Brigham, J. C. (1971). Ethnic stereotypes. *Psychological Bulletin,* *76*, 15–38.

Briton, N. J., & Hall, J. A. (1995). Beliefs about female and male nonverbal communication*. Sex Roles,* *32*, 79–90.

Brodt, S. E. & Ross, L. D. (1998). The role of stereotyping in overconfident social prediction. *Social Cognition, 16*, 225–52.

Brunswik, E. (1952). *The conceptual framework of psychology*. University of Chicago Press.

Campbell, D. T. (1967). Stereotypes and the perception of group differences. *American Psychologist,* *22*, 817–29.

Cejka, M. A., & Eagly, A. H. (1999). Gender-stereotypic images of occupations correspond to the sex segregation of employment. *Personality and Social Psychology Bulletin,* *25*, 413–23.

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.

Clarke, R. B., & Campbell, D. T. (1955). A demonstration of bias in estimates of Negro of ability. *Journal of Abnormal and Social Psychology*, *51*, 585-588.

Cohen, C. E. (1981). Personal categories and social perception: Testing some boundaries of the processing effects of prior knowledge. *Journal of Personality and Social Psychology,* *40*, 441–52.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.

Crawford, J. T., Jussim, L., Madon, S., Cain, T.R., & Stevens, S. T. (2011). The use of stereotypes and individuating information in political person perception. *Personality and Social Psychology Bulletin,* *37*, 529–42.

 Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*, 5–18.

 Devine, P. (1995). Prejudice and outgroup perception. In: *Advanced social psychology*, ed A. Tesser, pp. 467–524. McGraw-Hill.

 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.

Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. In: *The handbook of social psychology*, *vol. 2*, 4th edition, ed. D. Gilbert, S. T. Fiske & G. Lindzey, pp. 357–411. McGraw-Hill.

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*. Advances in Experimental Social Psychology,* *23*, 1–74.

Fiske, S.T., & Taylor, S.E. (1991). *Social cognition*, 2nd edition. McGraw-Hill.

Fiske, S.T., & Taylor, S.E. (2013). *Social cognition: From brains to culture*. Thousand Oaks, CA: Sage Publications

Funder, D. C. (1987). Errors and mistakes: Evaluating the accuracy of social judgment*. Psychological Bulletin,* *101*, 75–90.

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review,* *102*, 652–70.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*, 4–27.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. Journal of *Personality and Social Psychology*, *74*, 1464–1480.

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., & Goh, J. X. (2017). Studying Stereotype Accuracy from an Integrative Social‐Personality Perspective. *Social and Personality Psychology Compass*, *11*(11).

Halpern, D.F., Straight, C.A., & Stephenson, C.L. (2011). Beliefs about cognitive gender differences: Accurate for direction, underestimated for size. *Sex Roles*, *64*, 336-347.

Hamilton, D. L., Sherman, S. J., & Ruvolo, C. M. (1990). Stereotype-based expectancies: Effects on information processing and social behavior. *Journal of Social Issues,* *46*, 35–60.

Heilman, M.E. (1983). Sex bias in work settings. *Research in Organizational Behavior*, *5*, 269-298.

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, *60*, 581–592.

Jones, E. E. (1986). Interpreting interpersonal behavior: The effects of expectancies. *Science*, *234*, 41–46.

Jost, J. T., & Banaji, M. R. (1994). The role of stereotyping in system-justification and the production of false consciousness*. British Journal of Social Psychology,* *33*, 1– 27.

Jost, J.T., & Kay, A.C. (2010). Social justice: History, theory, and research. In S.T. Fiske, D. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (5th edition, Vol. 2, pp. 1122-1165). Hoboken, NJ: Wiley.

Jost, J. T., & Kruglanski, A. W. (2002). The estrangement of social constructionism and experimental social psychology: History of the rift and prospects for reconciliation. *Personality and Social Psychology Review,* *6*, 168–87.

Judd, C. M., & Park, B. (1993). Definition and assessment of accuracy in social stereotypes. *Psychological Review,* *100*, 109–28.

Judd, C. M., Park, B., Ryan, C. S., Brauer, M., & Kraus, S. (1995). Stereotypes and ethnocentrism: Diverging interethnic perceptions of African American and White American youth. *Journal of Personality and Social Psychology,* *69*, 460–81.

Jussim, L. (1991). Social perception and social reality: A reflection-construction model. *Psychological Review,* *98*, 54–73.

Jussim, L. (2005). Accuracy in social perception: Criticisms, controversies, criteria, components, and cognitive processes. *Advances in Experimental Social Psychology*, *37*, 1-93.

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

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

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: *Handbook of prejudice, stereotyping, and discrimination*, 2nd edition, ed. T. Nelson, pp. 31–63. Psychology Press.

Jussim, L., Crawford, J.T., Anglin, S.M., & Stevens, S.T. (2015a). The politics of social psychological science: Distortions in the psychology of intergroup relations. In P. Valdeso & J. Grahame (Eds.), The Claremont Symposium on Applied Social Psychology Bridging Ideological Divides.

Jussim, L., Crawford, J. T., & Rubinstein, R. S. (2015b). Stereotype (in)accuracy in perceptions of groups and individuals. *Current Directions in Psychological Science,* *24*, 490–97.

Jussim, L., Eccles, J., & Madon, S. (1996). Social perception, social stereotypes, and teacher expectations: Accuracy and the quest for the powerful self-fulfilling prophecy. *Advances in Experimental Social Psychology,* *28*, 281–388.

Kang, J., & Banaji, M. R. (2006). Fair measures: A behavioral realist revision of affirmative action. *California Law Review*, *94*, 1063–1118.

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.

Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis*. Guilford Press.

Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychological Review,* *103*, 284–308.

LaPiere, R. T. (1936). Type-rationalizations of group antipathy. *Social Forces*, *15*, 232– 37.

Lick, D. J., Alter, A. L., & Freeman, J. B. (2017). Superior pattern detectors efficiently learn, activate, apply, and update social stereotypes. *Journal of Experimental Psychology: General.* Advance online publication. http://dx.doi.org/10.1037/xge0000349

Lippmann, W. (1922/1991). *Public opinion*. Transaction Publishers.

Löckenhoff, C. E., Chan, W., McCrae, R. R., De Fruyt, F., Jussim, L., De Bolle, M., Costa, P. T., Jr., Sutin, A. R., Realo, A., Allik, J., Nakazato, K., Shimonaka, Y., Hrěbíčková,M.,Graf,S.,Yik,M.,Ficková,E.,Brunner-Sciarra,M.,Leibovich de Figueora, N., Schmidt, V., Ahn, C.-K., Ahn, H.-N., Aguilar Vafaie, M. E., Siuta, J., Szmigielska, B., Cain, T. R., Crawford, J. T., Anwar Mastor, K., Rolland, J.-P., Nansubuga, F., Miramontez, D. R., Benet-Martínez, V., Rossier, J., Bratko, D., Marušic,́ I., Halberstadt, J., Yamaguchi, M., Kneževic,́ G., Martin, T. A., Gheorghiu, M., Smith, P. B., Barbaranelli, C., Wang, L., Shakespeare- Finch, J., Lima, M. P., Klinkosz, W., Sekowski, A., Alcalay, L., Simonetti, F., Avdeyeva, T. V., Pramila, V. S., & Terracciano, A. (2014). Gender stereotypes of personality: Universal and accurate*? Journal of Cross-Cultural Psychology,* *45*, 675–94.

Madon, S. J., Jussim, L., Keiper, S., Eccles, J., Smith, A., & Palumbo, P. (1998). The accuracy and power of sex, social class and ethnic stereotypes: Naturalistic studies in person perception. *Personality and Social Psychology Bulletin,* *24*, 1304–18.

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.

McCauley, C., & Stitt, C. L. (1978). An individual and quantitative measure of stereotypes. *Journal of Personality and Social Psychology,* *36*, 929–40.

McCauley, C., & Thangavelu, K. (1991). Individual differences in sex stereotyping of occupations and personality traits. *Social Psychology Quarterly,* *54*, 267–79.

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.

Mendelberg, T. (2001). *The race card: Campaign strategy, implicit messages, and the norm of equality*. Princeton, NJ: Princeton University Press.

Miller, D.T., & Turnbull, W. (1986). Expectancies and interpersonal processes. *Annual Review of Psychology,* *37*, 233–56.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, N. J: Prentice-Hall.

Ottati, V., & Lee, Y. T. (1995). Accuracy: A neglected component of stereotype research. In Y. T. Lee, L. Jussim, & C. R. McCauley (Eds.), *Stereotype accuracy: Toward appreciating group differences* (pp. 29-59). Washington, DC: American Psychological Association.

Payne, B. K., Vuletich, H. A., & Lundberg, K. B. (2017). The bias of crowds: How implicit bias bridges personal and systemic prejudice. *Psychological Inquiry*, *28*(4), 233-248.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118*, 219–235.

Richard, F.D., Bond Jr., C.F., & Stokes-Zoota, J.J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology, 7,* 331-363.

Rosenthal, R., & Rubin, D.B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology, 74,* 166-169.

Rubinstein, R.S., Jussim L., & Stevens, S.T. (in press). Reliance on individuating information and stereotypes in implicit and explicit person perception. *Journal of Experimental Social Psychology,* DOI: https://doi.org/10.1016/j.jesp.2017.11.009.

Ryan, C. (1996). Accuracy of Black and White college students’ in-group and out- group stereotypes. *Personality and Social Psychology Bulletin,* *22*, 1114–27.

Ryan, C.L., & Bauman, K. (2016). Educational attainment in the United States: 2015. *Current Population Reports*. U.S. Census Bureau. Retrieved from: https://www.census.gov/content/dam/Census/library/publications/2016/demo/p20-578.pdf

Stangor, C. (1995). Content and application inaccuracy in social stereotyping. In: *Stereotype accuracy: Towards appreciating group differences*, ed. Y. T. Lee, L. Jussim & C. R. McCauley, pp. 275–92. American Psychological Association.

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*. Doubleday.

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.

Tetlock, P. E. (1994). Political psychology or politicized psychology: Is the road to scientific hell paved with good moral intentions? *Political Psychology 15*, 509–29.

Tetlock, P.E., & Gardner, D. (2015). *Superforecasting: The art and science of prediction*. New York, NY: Crown Publishers.

Tetlock, P.E., Kristel, O.V., Elson, S.B., Green, M.C., & Lerner, J.S. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology, 78,*853-870.

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133,* 859-883.

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*, 635–54.