2

STEREOTYPE ACCURACY

ONE OF THE LARGEST AND MOST REPLICABLE EFFECTS IN ALL OF SOCIAL PSYCHOLOGY

Lee Jussim
RUTGERS UNIVERSITY

Jarret T. Crawford
THE COLLEGE OF NEW JERSEY

Stephanie M. Anglin
RUTGERS UNIVERSITY

John R. Chambers
SAINT LOUIS UNIVERSITY

Sean T. Stevens
RUTGERS UNIVERSITY

Florette Cohen
CUNY-STATEN ISLAND

Stereotype accuracy is one of the largest and most replicable effects in all of social psychology. It took social psychology nearly a century to recognize that not only had it been *declaring* stereotypes to be inaccurate on the basis of little data, but once the data started to come in, to accept that this data often (though not always) demonstrated moderate to high stereotype accuracy. This resistance to the data has constituted a significant impediment to understanding the existence, causes, and consequences of *both* stereotype accuracy and inaccuracy.

This chapter is divided into three major sections. The first (History of Obstacles to Social Psychology Accepting Its Own Data on Stereotype Accuracy) reviews some of the obstacles social psychology has faced with respect to accepting that stereotype (in)accuracy is an empirical question, and that the empirical data do not justify assumptions,

definitions, or declarations that stereotypes are inaccurate. The second (The Empirical Assessment of Stereotype (In)Accuracy) summarizes what is now an impressive body of literature assessing the (in)accuracy of racial, gender, age, national, ethnic, political, and other stereotypes. The third (Stereotype (In)Accuracy: Knowns, Unknowns, and Emerging Controversies) summarizes broad and emerging patterns in that body of literature, highlighting unresolved controversies, and identifying important directions for future research.

HISTORY OF OBSTACLES TO SOCIAL PSYCHOLOGY ACCEPTING ITS OWN DATA ON STEREOTYPE ACCURACY

Why have claims of stereotype inaccuracy persisted in the absence of much evidence demonstrating such inaccuracy? There are probably multiple answers to this question. A strong contender is that social psychologists have long been concerned with alleviating social problems, especially those that arise from prejudice and discrimination. As such, much foundational and influential research, and many social psychology textbooks, have decried the many ways stereotypes reflect and cause social problems (Allport, 1954/1979; APA, 1991; Fiske & Neuberg, 1990; Fiske & Taylor, 1984, 1991; Jost & Banaji, 1994; Katz & Braly, 1933; LaPiere, 1936; Snyder, 1984; Steele, 1997). Because stereotypes were so obviously (to many psychologists) bad things, *ipso facto*, they must be inaccurate.

Nonetheless, several early reviews pointed out that defining stereotypes as inaccurate was quite common but rarely justified by empirical data (Ashmore & Del Boca, 1981; Brigham, 1971; Mackie, 1973; McCauley, Stitt, & Segal, 1980). Indeed, the upshot of Mackie's (1973) article, titled "Arriving at 'Truth' by Definition: The Case of Stereotype Inaccuracy," was that declarations of stereotype inaccuracy actually required empirical data, which, at the time, was lacking.

The Black Hole at the Bottom of Many Declarations of Stereotype Inaccuracy

In science, the convention is to support empirical claims with evidence, typically via a citation. This should be an obvious point, but far too often, scientific articles have declared stereotypes to be inaccurate either *without a single citation*, or by citing an article that declares stereotype inaccuracy without actually citing empirical evidence. We call this "the black hole at the bottom of declarations of stereotype inaccuracy" and give some examples next.

> . . . 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)

No evidence was cited to support this claim.

> Journalist and political commentator Walter Lippmann, who coined the term, made a distinction between the world "out there" and the stereotype—the little pictures in our heads that help us interpret the world we see. To stereotype is to allow those pictures to dominate our thinking, leading us to assign identical characteristics to any person in a group, regardless of the actual variation among members of that group.
>
> (Aronson, 2008, p. 309)

Assigning *identical* characteristics to *any* person in a group, *regardless* of the actual variation is an extreme claim. Aronson (2008) does not cite anything to support such an extreme claim because he cannot. After nearly 100 years of empirical research on social stereotypes, there is not a single study that has reported a single person who believes all members of any group have identical characteristics.

> 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.

Even the APA, in its official pronouncements, has not avoided the inexorable pull of this conceptual black hole. They first declare:

> Stereotypes 'are not necessarily any more or less inaccurate, biased, or logically faulty than are any other kinds of cognitive generalizations,' Taylor, *supra* note 11, at 84, and they need not inevitably lead to discriminatory conduct (p. 1064).

They then declare:

> "The problem is that stereotypes about groups of people often are *overgeneralizations and are either inaccurate or do not apply to the individual group member in question*." *Sex Bias in Work Settings, supra* note 11, at 271 (emphasis in original).

Evaluating the rationale for first declaring stereotypes to be not necessarily inaccurate, immediately followed by declaring them to be either inaccurate or inapplicable is beyond the scope of this chapter. The APA does include a reference. "Sex bias in work settings" is an article by Heilman (1983), which does *declare* stereotypes to be inaccurate. It also reviews evidence of bias and discrimination. But it neither provides nor reviews empirical evidence of stereotype accuracy (and, as we show later, bias and accuracy are not mutually exclusive, so that demonstrations of bias rarely constitute demonstrations of inaccuracy). Thus, another declaration of stereotype inaccuracy ends in a black hole.

Our point here is not to condemn these particular researchers. Indeed, they were simply following a commonly accepted practice that has been widespread in social psychology for decades. This black hole phenomenon, therefore, both reflected and constituted an obstacle that social psychologists faced in recognizing first, that for many years, their field lacked evidence demonstrating stereotype inaccuracy, and second, over the last 40 years, that it has provided abundant evidence of stereotype accuracy. First, its existence strongly suggests that social psychologists once so firmly *believed* in stereotype inaccuracy, that declaring stereotypes to be inaccurate did not even require a reference! Just as one does not need to cite a reference to declare something as obvious as "the sky is blue," social psychologists evidently needed no reference to declare stereotypes inaccurate.

Second, it modeled inappropriate scientific behavior. Famous psychologists publishing in prestigious outlets declaring stereotypes inaccurate without a citation meant that *anyone* could likewise do so. Third, it also *created* an obstacle by constructing an illusion that pervasive stereotype inaccuracy was "settled science." Many articles and chapters had *stated* that stereotypes were inaccurate, so it was easy for subsequent researchers not only to similarly declare stereotypes inaccurate but also to provide supportive citations. It was only if one examined the empirical research underlying such claims that one discovered that there was nothing there, just a black hole.

Concern with Alleviating Oppression

Many social psychologists have been deeply concerned with combating oppression—anti-Semitism in the immediate aftermath of World War II, racism and sexism following the civil rights and women's movements in the 1960s, respectively, and other types of bigotries and prejudices. Combating oppression is a good thing. However, as we (Jussim et al., 2009, p. 199) wrote for the first edition of this Handbook:

> Sixty years of empirical research has told us much about stereotypes. Stereotypes can arise from, and sustain, intergroup hostility. They are sometimes linked to prejudices based on race, religion, gender, sexual orientation, nationality, and just about any other social category. They can serve to maintain and justify hegemonic and exploitative hierarchies of power and status. They can corrupt interpersonal relations, warp public policy, and can play a role in the worst social abuses, such as mass murder and genocide. For all these reasons, social scientists—and especially social psychologists—have understandably approached stereotypes as a kind of social toxin.
>
> Perhaps equally understandable, but scientifically untenable, is the corresponding belief that because stereotypes contribute to these many malignant outcomes, that they must also be—in the main—inaccurate. The tacit equation is, if stereotypes are associated with social wrongs, they must be factually wrong. However, the accuracy of stereotypes is an empirical question, not an ideological one. And for those of us who care deeply about stereotypes, prejudice, and social harmony, getting to the truth of these collective cognitions should guide inquiry about them.

Nonetheless, it is also now clear that some feared that evidence of stereotype accuracy could be used to undermine the efforts of social psychologists committed to reducing prejudice, discrimination, and oppression. These concerns also constituted an obstacle to widespread acceptance of the stereotype accuracy data. For example, Stangor (1995, pp. 288–289) explains why stereotype accuracy is not worthwhile to study, in part this way: "As scientists concerned with improving the social condition, we must be wary of arguments that can be used to justify the use of stereotypes." And then later in the same paragraph: "we cannot allow a bigot to use his or her stereotypes, even if those beliefs seem to them to be accurate."

In Fiske's 1998 (p. 381) *Handbook* chapter on stereotypes, prejudice, and discrimination, she makes a related point:

> Moreover, they [referring to McCauley, Jussim, & Lee, 1995] differ from the present review in their conclusions, which do not follow from their premises: If two resumés

are otherwise equivalent, it is permissible to use stereotypes associated with group membership as a factor in hiring choice, if group membership has previously predicted success on the job. (In this they evidently disagree with U.S. civil rights law).

These are explicitly *political* rationales for quashing accuracy research. Both quotes refer quite bluntly to political power rather than science ("cannot allow a bigot" and "permissible to use stereotypes"). People in power make decisions about what is allowable and permissible, whereas, presumably, scientific research does not. Fiske (1998) is clearly correct in declaring that *discriminating* on the basis of certain social categories—such as race, ethnicity, religion, and sex—is illegal. Nonetheless, the comment, "they evidently disagree with civil rights law" is curious because, although the McCauley et al. (1995) review did address accuracy, it did not address discrimination, legal issues, or civil rights law. Thus, rather than presenting a cogent critique of the methods, findings, or interpretation of accuracy research, Fiske's (1998) discussion of McCauley et al. (1995) functions as a red flag, warning researchers that if they study stereotype accuracy empirically, they risk being subject to accusations of "disagreeing with civil rights law" or to blunter accusations of bigotry.

Defining Stereotypes as Inaccurate is Logically Incoherent

Defining stereotypes as inaccurate is both common and logically incoherent, for reasons briefly summarized here (see Jussim, 2012; Jussim et al., 2009 for extended presentations). Defining stereotypes as inaccurate can only mean one of two things: 1) All beliefs about all groups are stereotypes and all are inaccurate; or 2) Stereotypes are the subset of beliefs about groups that are inaccurate, whereas accurate beliefs about groups are not stereotypes. Both are incoherent, at least for social psychology.

Only descriptive beliefs can be accurate or inaccurate. "Jews are richer than other Americans" can be evaluated for accuracy; the accuracy of "I dislike Jews," however offensive and psychologically important, cannot be evaluated for accuracy. Stereotypes as prescriptive beliefs, too, cannot be evaluated for their accuracy. Accuracy is irrelevant to notions such as "children should be seen and not heard" or "men should not wear dresses." Therefore, to the extent that stereotypes are defined as something *other* than descriptive or predictive beliefs, one is precluded from making *any* claim about inaccuracy.

The assumption that stereotypes are inaccurate, therefore, can mean only one of two things. First it may mean that all beliefs about all groups are stereotypes, and all are inaccurate. If stereotypes are inaccurate, and all beliefs about all groups are stereotypes, then all beliefs about groups are inaccurate. This is logically incoherent. This means that it is inaccurate to believe two groups differ, and inaccurate to believe they do not differ. Thus, this meaning of "stereotypes are inaccurate" can be dismissed out of hand.

Second, defining stereotypes as inaccurate may instead mean that stereotypes are the subset of beliefs about groups that are inaccurate. Beliefs that are accurate are *not* stereotypes; only inaccurate beliefs about groups are stereotypes. This, however, also needs to be dismissed, unless one is willing, instead, to dismiss the vast body of social psychological research on stereotypes. That is because we are aware of no research—not a single study—that has been framed as follows:

Is THIS belief about THAT group a stereotype? We are going to figure out whether THIS belief about THAT group is a stereotype by assessing whether that belief is

inaccurate. If THIS belief is inaccurate, we will conclude that it is a stereotype. If THIS belief accurately described THAT group, we will conclude that it is not a stereotype.

Absent an a priori demonstration that a belief about a group is inaccurate, the researcher cannot claim to be studying a stereotype. Because no research framed as studying stereotypes includes such an a priori demonstration, one cannot be sure that any research framed as addressing a stereotype actually has done so (if one accepts this definition).

The Criterion "Problem" As Double Standard

Double standards are a close conceptual cousin of logical incoherence, and can reflect morally or politically motivated biased reasoning (Altemeyer, 1996; Crawford, 2012; Haidt, 2012). The double standard here occurs when social psychologists raise questions, criticisms, or concerns about the existence of criteria for assessing accuracy (e.g., Fiske, 1998; Stangor, 1995), but enthusiastically embrace other research requiring identical criteria—which occurs when that other research can be used to construct narratives emphasizing the power and pervasiveness of oppression. Self-fulfilling prophecies are often discussed in ways consistent with narratives of oppression—as a social process by which stereotypes lead to discrimination and inequality (e.g., Darley & Fazio, 1980; Ross, Lepper, & Ward, 2010; Weinstein, Gregory, & Strambler, 2004). Evidence of accuracy seems to contest those narratives, in part, because self-fulfilling prophecies begin with an erroneous expectation (Merton, 1948). If the expectation is not erroneous, self-fulfilling prophecies do not occur. Similarly, accuracy implies that individual or group differences have some objective social reality to them, thereby seeming to undercut narratives blaming oppression, in part, on perceiver biases (see Jussim, 2012 for an elaboration of this analysis). Thus, self-fulfilling prophecies seem to support narratives of oppression, and accuracy may often be perceived as contesting those narratives.

Many researchers have raised the issue of identifying criteria for assessing accuracy as something problematic (e.g., Fiske, 1998; Jones, 1985; Kruglanski, 1989). There is some validity to such a critique because there rarely is a perfect criterion against which to assess the accuracy of lay judgments. Most criteria have advantages and disadvantages, something accuracy researchers have long recognized (for reviews, see Funder, 1987; Judd & Park, 1993; Jussim, 2012; Ryan, 2002).

A double standard arises when researchers (often the same ones) write enthusiastically and uncritically about the power and pervasiveness of self-fulfilling prophecies (e.g., Fiske & Neuberg, 1990; Jones, 1986; Jost & Kruglanski, 2002) without raising similar issues about criteria. For both self-fulfilling prophecies and accuracy, one must establish correspondence between perceiver expectation and targets' outcomes. There is a difference in how the correspondence comes about, but there is no difference in the criteria for establishing whether that correspondence has come about. Criteria for establishing correspondence are inherently identical for self-fulfilling prophecies and accuracy. By routinely raising critical questions about the difficulty in identifying criteria to assess accuracy but not self-fulfilling prophecy, the politically distasteful phenomenon (accuracy) is held to greater critical scrutiny than the politically palatable phenomenon (self-fulfilling prophecy).

Regardless, anyone who believes there are no criteria available for assessing stereotype accuracy has also precluded themselves from declaring stereotypes to be inaccurate. If it

is impossible to assess how (in)accurate stereotypes are, it then becomes impossible to declare them to be either accurate or inaccurate.

THE EMPIRICAL ASSESSMENT OF STEREOTYPE (IN)ACCURACY

Preliminary Caveats

Accuracy is quantitative and probabilistic, not absolute (Jussim, 2005, 2012). As such, accuracy, error, and bias are *not* mutually exclusive (Jussim, 1991; West & Kenny, 2011). Therefore, declaring some stereotype (or stereotypes in general) to be moderately to highly accurate *does not* preclude the possibility that it (or they) also contain errors and biases.

Second, accuracy refers to correspondence between belief and criteria (Funder, 1987, 1995; Jussim, 1991, 2005, 2012). As such, accuracy questions are fundamentally about the content of people's beliefs, *not* the processes of social perception. How people arrive at their stereotyped beliefs is a process question, not an accuracy question. Similarly, how individuals and groups develop their characteristics are interesting and important social process questions, but they are not accuracy questions.

Third, methodological difficulties once plagued accuracy research. Those difficulties, however, have been addressed by a broad range of statistical, methodological, and conceptual advances over the last 25 years (e.g., Funder, 1987, 1995; Jussim, 1991, 2005, 2012; Jussim et al., 2009; Kenny, 1994; Ryan, 2002). Accuracy is now a thriving area of research within social psychology.

A Neutral Definition of Stereotype

Some modern definitions of stereotypes deftly avoid the logical incoherence that results from defining them as inaccurate simply by being agnostic with respect to stereotype (in)accuracy. One of the simplest of these definitions, and the one we will use throughout this chapter, was provided by Ashmore and Del Boca (1981, p. 21) who concluded that there is widespread agreement on this one aspect of stereotypes: "a stereotype is a set of beliefs about the personal attributes of a social group." Stereotypes, by this definition, may or may not be: accurate and rational, widely shared, conscious, rigid, exaggerations of group differences, positive or negative, or based on essentialist or biological rationales. Stereotypes may or may not be the cause or the effect of prejudice, or the cause of biases and self-fulfilling prophecies.

One of the great advantages of the neutral definition is that it does not presume that any time a person holds or uses a stereotype, something inherently bad (or good) is happening. Instead, it opens the door for understanding when stereotypes wreak damage, when they simply reflect social reality, and, possibly, when they actually perform a social good.

Our rejection of defining stereotypes as inaccurate is not equivalent to defining them as accurate. Accuracy is an empirical issue, which naturally raises a question: How (in)accurate are people's beliefs about groups? Before we can review the evidence that bears on the question, however, we consider different aspects of accuracy that can and have been assessed.

Types of Stereotype Accuracy

Accuracy is often a multidimensional construct (e.g., Judd & Park, 1993; Kenny, 1994), as can be readily illustrated with a simple example. Consider Fred, judging the average

height of male Americans, Columbians, and Dutch. Fred estimates the average heights, respectively, as 5'8", 5'5", and 5'10". Let's say the real average heights are, respectively, 5'10", 5'7", and 6'0". In absolute terms, Fred is inaccurate—he consistently underestimates height by two inches. However, in relative terms, Fred is perfectly accurate—his estimates correlate 1.0 with the actual heights. Although Fred has a downward bias in perceiving the absolute heights among men in the different countries, he is superb at perceiving the relative height differences.

Thus, 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. For example, people might be asked, "What is the median household income?" for various racial and ethnic groups. According to the U.S. Census (DeNavas-Walt & Proctor, 2014), for Asians, that value is $67,366. Any estimate other than exactly $67,366 will produce a nonzero discrepancy—thus, 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" to highlight the point that *any* discrepancy, however minute, will be produced by any judgment other than an absolute 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. 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.

Discrepancy scores and correlations have each 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.

What is a Reasonable Standard for Characterizing a Stereotypic Belief as "Accurate"?

There is no objective gold standard with which to answer this question. Indeed, the answer may depend on many issues, such as the context, difficulty of the judgment, and what is considered accurate in other situations. Even when social scientists generate hypotheses that predict differences on some outcome between groups (whether experimental or demographic), they are often quite satisfied with correlations of .2 or less (Richard, Bond, & Stokes-Zoota, 2003). This helps provide some context for making judgments about what should be considered accurate. Because there are two broad types of accuracy, discrepancy from perfection and correspondence with real differences, there needs to be two separate standards.

Discrepancy scores. 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. 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 and 39% and between 61 and 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).

Correspondence with real differences. How much correspondence should be considered "accurate"? Again, this is a judgment call. Nonetheless, we advocate holding people to a high standard—the same standards to which social scientists hold themselves.

Cohen (1988), in his classic statistical treatise imploring social scientists to examine the size of the effects they obtained in their studies and not just the "statistical significance" of the results, suggested that effect sizes above .8 could be considered "large." Such an effect size translates into a correlation of about .4 (in the remainder of this paper, "effect sizes" are discussed exclusively as correlations). By this standard, correlations of .4 and higher could be considered accurate because they represent a "large" correspondence between stereotype and reality.

Furthermore, only 24% of social psychological effects exceed .3 (Richard et al., 2003). Effect sizes of .4 and higher, therefore, constitute a strong standard for accuracy. Last, according to Rosenthal's (1991) binomial effect size display, a correlation of at least .4 roughly translates into people being right at least 70% of the time. This means they are right more than twice as often as they are wrong. That seems like an appropriate cut-off for considering a stereotype to be reasonably accurate.

Moderate correspondence, of course, is less than high correspondence. It reflects a mix of accuracy and inaccuracy. Following the same standards as science (Cohen, 1988; Richard et al., 2003), we will characterize correlations between people's beliefs and reality ranging from .25 to .4 as moderately accurate. Such correlations do not reflect perfect accuracy, but nor do they reflect complete inaccuracy. We consider correlations below .25 to be inaccurate.

Criteria for Inclusion

To be included here, empirical studies assessing the accuracy of stereotypes needed to meet a single criterion. They had to compare perceivers' beliefs about one or more target groups with measures of what that group was actually like. Studies assessing social cognitive processes, even when those processes are widely presumed to be flawed and invalid (illusory correlations, confirmation biases, attributional biases, various other motivated and self-justifying biases, etc.), are not included here, because such studies provide no direct information about accuracy (Funder, 1987; Jussim, 2005).

Our previous reviews of this literature (Jussim, 2012; Jussim et al., 2009) included a second criterion for inclusion: studies needed to use an appropriate target group. Sometimes, researchers have assessed people's beliefs about a group, and used as criteria the characteristics of a haphazard sample of members of the target group. These studies have an important mismatch between the stereotype they are assessing and the criteria they use. Consider, for example, a study in which perceivers provide their beliefs regarding men and women *in general*, and the criterion sample is a convenient but haphazard sample of college students. In this case, even if perceivers' stereotypes corresponded perfectly with men and women *in general*, they may not correspond to the characteristics

of this criterion sample, if the criterion sample's characteristics differ from those of men and women in general. Our view is that population data, representative samples, and large-scale meta-analyses of group differences generally constitute appropriate criterion samples against which to evaluate the accuracy of stereotypes.

However, over the last few years, there has been a great blossoming of research that uses the self-reports or observer reports of small or haphazard samples on the Big Five Personality traits (questionnaires assessing the personality traits of neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness have been administered to thousands, sometimes tens of thousands, of people around the world, e.g., Costa & McCrae, 2008; McCrae & Terracciano, 2005). This research has inspired an outburst of empirical research on the accuracy of national character stereotypes. Although the lack of representative criterion samples is a bona fide limitation to such research, our judgment was that the growth and broad international scope of this research warranted inclusion here. As such, this review includes, and, more important, *distinguishes* between studies that use haphazard or higher quality criterion samples.

Brief Summary of Studies Reviewed in Recent Publications

Detailed descriptions and critical evaluations of four studies of the (in)accuracy of racial and ethnic stereotypes, and of eight papers reporting 14 studies of the (in)accuracy of gender stereotypes can be found in Jussim (2012). A shorter summary of those findings can be found in our chapter for the prior edition of this handbook (Jussim et al., 2009). Jussim (2012) also reviewed the research on the accuracy of several other stereotypes besides race and gender (such as dancers, occupations, college majors, and sororities). Thus, we only summarize those older findings very briefly here.

Racial and ethnic stereotypes. Four studies assessed the racial and ethnic stereotype accuracy for specific judgments. The predominant pattern of consensual discrepancies was accuracy, which occurred for a plurality or majority of all judgments (Ashton & Esses, 1999; Kaplowitz, Fisher, & Broman, 2003; McCauley & Stitt, 1978; Ryan, 1996). Three studies found more evidence of underestimating than of exaggerating real differences (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) reported results that, by our standards, meant that 36 participants' stereotypes were accurate, 33 exaggerated real differences, and 25 underestimated real differences.

The original studies rarely reported consensual correlations. However, they could be computed *if* the original report included: 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 we consider $r \geq .40$ highly accurate and $\underline{.25} < r < .40$ 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, 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 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.

Other stereotypes. Stereotypes of college majors, occupations, sororities, and jazz vs. modern dancers (Cejka & Eagly, 1999; Clabaugh & Morling, 2004; Judd, Ryan, & Park, 1991) were generally consistent with the results summarized above for race and gender— very high consensual stereotype accuracy correlations (.80 to .90, when they could be assessed), high personal stereotype accuracy correlations (.40 to .70), and consensual discrepancies that were usually accurate more than they either exaggerated or underestimated group differences (personal discrepancies were not reported in these studies—see Jussim, 2012 for a detailed review).

Two Older Studies of Gender Stereotypes Not Included in Prior Reviews Because of Criterion Samples Mismatched to the Stereotype

Martin (1987). Martin (1987) assessed the accuracy of gender stereotypes in two studies. The first used the Bem Sex Role Inventory (BSRI, Bem, 1981) to assess stereotypes about

"North American adult males and females" on 30 trait descriptors (such as leadership, gentle, helpful and likeable). Unfortunately, the criterion sample was 94 male and 56 female adult visitors to a campus open-house day. This created a mismatch between the stereotype target group and the criterion sample.

Furthermore, this research highlights some of the difficulties in using self-reports as a criterion. All 30 traits were positive. Not surprisingly, self-reports on these positive traits were extremely high: both men and women described themselves as, overall, having the trait about 80% of the time (e.g., 91–98% of men and women described themselves as tactful, sincere, reliable, and helpful). The point of the study was not to assess the validity of such self-descriptions. Therefore, the extent to which self-descriptions are overstatements is unknown.

Martin (1987) presented the main results in terms of *diagnostic ratios* (DRs, proportion of males/proportion of females, computed separately for the stereotype judgments and criterion self-reports). By this computation, people mostly exaggerated differences.

DRs, however, have some computational quirks. For example, let's say Fred believes 5% of men and 1% of women like science fiction movies, whereas the real percentages are, respectively, 20% and 16%. The perceived difference (4%) exactly equals the real difference. However, the DR will make it appear as if Fred exaggerates the real difference, because his DR is 5, whereas the DR for the actual 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 exactly what Martin, 1987 found), thereby leading to artificially inflated estimates of exaggeration.

We focus, therefore, on the raw percentage estimates, which, fortunately, were reported for Study One (Table 3, p. 493), separately for male and female targets. With respect to consensual discrepancies, people consistently underestimated the traits of both genders, by an average of about 20% each. However, whether these represent true underestimates or inflated criteria, cannot be known from the data. Martin (1987) did not report personal discrepancy score analyses. Nonetheless, even using this likely flawed criteria, five of 60 judgments (30 each for males and females) were accurate, there were 23 near misses, and 32 were inaccurate. Perceptions of male/female differences were accurate 24 times, there were 4 near misses, and two judgments were inaccurate. The extraordinary accuracy here could have occurred because: 1) the 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. A total of 22 judgments overestimated real differences and 8 underestimated them.

Although Martin (1987) did not report consensual stereotype accuracy correlations, they are readily 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 Two 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." Participants were merely described as 106 female and 33 male undergraduate

volunteers. No mention of representative or random sampling appears, so, presumably, this is a haphazard convenience sample of unknown representativeness even of "students at this university" and likely suffers from a stereotype/criterion sample mismatch problem. Further, the male sample is particularly small.

Unfortunately, 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 again showed mostly exaggeration. Given that only DRs were reported, we simply correlated the 40 DRs on the self-report criteria with the 40 stereotype judgments. That correlation (r = .53) indicated considerable accuracy in perceptions of differences.

Allen (1995). 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 Two). Allen (1995) also collected a behavioral assessment as a criterion. However, the behavioral assessment was based on a brief discussion (18 minutes or less) of six social and political issues. Our view is that such a brief and limited behavioral sample is a very poor criterion for such judgments, and it is not discussed further here.

Unfortunately, Allen's (1995) method section did not explicitly articulate who participants were being asked about. However, in Martin's (1987) second study, they were asked about "students at this university." We presume that, as a replication, Allen (1995) did the same. Thus, this study suffers from the same stereotype/criterion mismatch as Martin's (1987) research.

Allen's (1995) main results were reported as diagnostic ratios. Unfortunately, the raw percentage estimates were not reported. For both stereotype judgments and self-reports, the diagnostic ratios ranged from 0.16 to 4.22, with one exception: the diagnostic ratio 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 summary below, 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 2. That correlation was .61—a stunning degree of accuracy for an article titled "Gender Stereotypes Are Not Accurate." Allen (1995) did not report personal stereotype accuracy correlations or discrepancies.

RECENT STEREOTYPE ACCURACY RESEARCH

In this section, we review stereotype accuracy research that did not appear in prior reviews of stereotype accuracy (Jussim, 2012; Jussim et al., 2009; Ryan, 2002). This section reviews research on race, gender, national character, and political stereotypes.

Recent research on race and gender stereotypes. We located no studies of racial stereotype accuracy published after 2003. This strikes us as unfortunate, because so many important questions regarding the nature, degree, and sources of racial stereotype (in)accuracy still remain unanswered.

There has, however, been recent research on gender stereotype (in)accuracy. One study found results for consensual gender stereotype discrepancies consistent with those

reported above (Halpern, Straight, & Stephenson, 2011). Indeed, the title of the article is "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 stereotypes. 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 again self-reports on the Big Five. Three patterns clearly 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.

National character/personality stereotypes. Because Big Five Personality Inventories have now been administered to thousands of people worldwide (e.g., Costa & McCrae, 2008; Costa, Terracciano, & McCrae, 2001; McCrae & Allik, 2002; McCrae & Terracciano, 2005), there are abundant data on personality characteristics in many different countries.

This has sparked an outburst of research assessing the accuracy of what has generally been called "national character (i.e., personality) stereotypes."

The first of these studies found that there is little evidence of accuracy (consensual stereotype accuracy correlations averaged near-zero) in 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). The criterion samples used in this study were haphazard samples of convenience. Terracciano et al. (2005) did not assess personal stereotype accuracy. Consensual stereotype accuracy correlations were near-zero, and consensual discrepancies consistently exaggerated real differences.

However, Heine, Buchtel, and Norenzayan (2008) suggested that Terracciano et al.'s (2005) failure to find more evidence of accuracy could have occurred, in part, because people rely on their local cultural norms when arriving at perceptions of themselves (called "the reference group effect"). They suggested that this can distort the measurement of cultural differences. This analysis led them to re-evaluate the validity of the "no accuracy in national character stereotypes" conclusion by comparing stereotypes to behavior potentially reflecting conscientiousness. When behavior (GDP, longevity, walking speed, clock accuracy, and postal worker speed) rather than self-reports on Big Five personality questionnaires were used as the criteria the correlations between consensual stereotypes and behavior averaged about .60.

Even so, a recent replication (McCrae et al., 2013, N>3000) addressed many of the issues raised by Heine et al. (2008) and, still using self-reports on the Big Five as criteria, showed almost no accuracy in national character stereotypes. These findings are consistent with other recent research showing very little accuracy in the personality stereotypes of one another held by over 2,000 Austrian, Czech, German, Polish, and Slovak college students (Hrebickova & Graf, 2014). As is typical in this area, the criterion samples were large and international, but not sampled in a random or representative manner.

In contrast, a recent study (Allik, Alyamkina, & Mescheryakov, in press) of the consensual stereotype accuracy correlations held by three ethnic groups (Erzians, Mokshans, Russians) regarding one another's personalities (Big Five), in the Republic of Mordovia (Russian Federation) found: 1) at least moderate accuracy in autostereotypes (correlations of .38, .51, .25, respectively); 2) substantial accuracy in heterostereotypes (stereotypes regarding the other group) held by two of the groups (Erzians and Mokshans, correlations ranging from .39 to .51); and 3) the stereotype accuracy correlations for Russian perceivers regarding Erzians and Mokshans were near-zero. The accuracy obtained in this study was particularly striking 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 200–286 participants for each perceiver group (most were college students, but there were subsamples of working adults among the Latvians and Estonians). Haphazard convenience samples constituted perceiver and target groups, whose personality and perceptions of national character were assessed on the Big Five. Realo et al. (2009) found a mixed pattern, wherein consensual autostereotype correlations ranged from .07 to .52 (three were below .20, but the other 6 were all .35 or higher, see their Table 1, correlations under the heading "ICC[1]"). The only heterostereotype examined was regarding Russians, and that was not accurate (of eight consensual stereotype accuracy correlations, only one was above .3 and the median correlation was near zero).

A Finnish study (Lonnqvist, Yijala, Jasinskaja-Lahti, & Verkasalo, 2012) also found a mixed pattern of consensual stereotype accuracy correlations (which was the only type of accuracy they assessed). They had three perceiver samples: Finnish students, Ingrian Finns (ethnic Finns living in Russia), and Russians, Ns = 286, 151, and 78, respectively. They assessed each groups' stereotypes of Finns' and Russians' personal values. This is one of few national stereotype accuracy studies that compared stereotypes to responses (on the same values survey) of representative samples of Finns and Russians. Consensual stereotype accuracy correlations were very high for Ingrian Finn and Russian perceptions of Finns (both over .6), and moderate for Finnish students' perceptions of Finns (.29). All three groups' perceptions of Russians showed, essentially, no consensual stereotype accuracy, and, indeed, were in the wrong direction (−.23 to −.33).

Why there is so much evidence of accuracy in almost all other stereotypes that have been studied, but so little among national stereotypes is certainly a puzzle. In the final section of this chapter (Stereotype (In)Accuracy: Knowns, Unknowns, and Emerging Controversies) we consider some possible explanations and directions for future research.

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 U.S. 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 (Big Five by 20 clusters of states). A total of 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 extroversion, 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 .78; the previously near-zero correlation for agreeableness became .56; the correlation for neuroticism was largely unchanged at .60; the near-zero correlation for extroversion 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 .25, respectively), but near-zero for extroversion, agreeableness, and conscientiousness.

Political stereotypes. Evidence from a wide variety of studies, using different methods, measures, criteria, and target groups, strongly suggests that people's political stereotypes—e.g., in the U.S., stereotypes of liberals vs. conservatives or of Democrats vs. Republicans—typically get the direction of the differences right, but exaggerate them. For example, Graham, Nosek, and Haidt (2012) assessed an Internet sample of over 2,000 people's beliefs about the moral values held by liberals and conservatives, and compared those stereotypes to the actual moral values as indicated by a representative sample of over 1,000 adults.

We computed consensual stereotype accuracy correlations based on the means reported in the lower panel of their Table 1, which reports results for the criterion

representative sample. The consensual stereotype accuracy correlations were: .89, .61, and .98, respectively, for stereotypes of liberals, conservatives, and their differences.

This, however, is also a real-world example of how accuracy and bias are not mutually exclusive. Only three of 18 judgments were accurate, 7 were near misses, and 8 were inaccurate. Furthermore, people consistently exaggerated real differences between liberals and conservatives. Exaggeration was typically more than an entire scale point on a six-point scale (e.g., if the real liberal–conservative difference was two points, the estimated difference was over three points).

Chambers, Baron, and Inman (2006) directly assessed partisans' beliefs about one another's attitudes. One study focused on pro-choice and pro-life advocates' beliefs about one another's positions on issues related to legalized abortion, such as the value of human life and women's reproductive rights. A second study addressed Democrats' and Republicans' beliefs about one another's positions on crime, military spending, public education, and eliminating inequality. In both studies participants were asked about *other people in the same study,* ensuring their perceptions were matched against an appropriate criteria (i.e., the average responses of other participants in that study) to assess (in)accuracy. At the consensual level, there was evidence of high accuracy and exaggeration. Consensual stereotype accuracy correlations (unreported but readily computed from the reported means in Tables 1 and 4) ranged .75 to .99, with a median of .86. As with the Graham et al. (2012) study, people understood the direction of differences between left and right on many issues.

Although Chambers et al. (2006) did not report personal stereotype accuracy correlations, fortunately, those data were still available. For Study One, the average personal stereotype accuracy correlations were .61 and .48, respectively, for pro-choice and pro-life respondents. For Study Two, the average stereotype accuracy correlations were .78, .57, .37, and .42, respectively, for Republicans' perceptions of Democrats, Democrats' perceptions of Republicans, and politically neutral respondents' perceptions of Democrats and of Republicans.

However, there was also clear evidence of exaggeration, particularly for values central to their group's own beliefs, but not for values central to their opponents' beliefs. For example, Chambers et al. (2006) found that pro-choice proponents greatly exaggerated the difference between pro-choicers' and pro-lifers' attitude positions concerning women's reproductive rights (a core value issue for pro-choice proponents), whereas pro-life proponents greatly exaggerated the difference between both groups' attitude positions on the value of human life (a core value issue for pro-life proponents). Chambers et al. (2006) found that these exaggerations were even stronger in an activist sample than in a student sample. Similar patterns occurred in the second study, where Republicans exaggerated the difference between Republicans' and Democrats' attitude positions concerning the prevention of crime and a strong military (central value issues for Republicans), but the Democrats exaggerated the difference between the two party's attitude positions concerning public education funding and eliminating inequalities (central value issues for Democrats). On the core issues, exaggeration ranged from 1 to almost 4 full-scale points (on an 11-point scale). Across the two studies, nearly all consensual discrepancies were more than half a standard deviation discrepant from the criteria—i.e., they were sufficiently large to be considered inaccurate by our standard.

Chambers and Melnyk (2006, Study Two) also assessed Democrats' and Republicans' perceptions of each other. However, in addition to politically relevant issues (e.g., perceived support for military spending, public education, etc.), they included several

issues that are less germane to politics (e.g., perceived support for organ donation, road and highway improvements). Nevertheless, stereotypes were generally accurate here as well. Consensual stereotype accuracy correlations ranged from .66 to .73, and the average personal stereotype accuracy correlation was .47 for Republicans' perceptions of Democrats, and .49 for Democrats' perceptions of Republicans. However, there was again clear evidence of exaggeration on the perceiver's own core value issues, and the consensual discrepancies for these issues (all greater than .5 standard deviations) were large enough to be considered inaccurate by our standards. Interestingly, on the *other party's* core value issues, stereotypes were either accurate or slightly *underestimated* the actual group differences (this same pattern was observed in Chambers et al., 2006; see also Graham et al., 2012). Chambers and Melnyk (2006) also found a highly similar overall pattern in the first study, which focused on stereotypes about people who were pro-life or pro-choice.

Westfall, Van Boven, Chambers, and Judd (2015) have not only replicated the exaggeration pattern with nationally representative American samples (as have Judd & Park, 1993), their results were broadly consistent with those of Chambers et al. (2006). Specifically, whereas Chambers et al. (2006) found the strongest evidence of exaggeration on partisans' core beliefs, Westfall et al. (2015) found that exaggeration was greater among partisans more strongly identified with a major political party, or whose attitudes were more extreme. Together, these studies (Chambers et al., 2006; Westfall et al., 2015) strongly suggest that the people who care most about politics (core values, strongly identified partisans, extremists) are most likely to exaggerate real political differences.

Judd and Park (1993) also reported personal stereotype accuracy correlations (between people's beliefs about Democrats' and Republicans' positions on 10 policy issues, and a nationally representative sample of Democrats' and Republicans' actual positions). The correlations were computed *within* Democratic target positions, and within Republican target positions, so they do not capture perceived differences. Those correlations averaged .25, indicating a moderate level of accuracy in perceptions of differing levels of support for various issues within parties.

This general pattern—where people get the direction of political differences right, but consistently exaggerate them—has been highly robust and replicable, and occurs regardless of whether the beliefs are about morals, values, policy positions or traits, and has been obtained by multiple independent labs conducting research decades apart (Crawford, Modri, & Motyl, 2013; Dawes, Singer, & Lemons, 1972; Judd & Park, 1993; Keltner & Robinson, 1996; Robinson, Keltner, Ward, & Ross, 1995; Westfall, et al., 2015). The pervasiveness of the pattern of partisan exaggeration is striking.

Stereotypes of Other People's Stereotypes!

There is, however, one more domain that shows consistent evidence of stereotype exaggerations. People stereotype other people as holding more extreme stereotypes than they really do. Four studies examined gender, college major, and U.S. regional stereotypes *and* people's beliefs about others' gender, college major, and regional stereotypes (Rettew, Billman, & Davis, 1993). In this study, therefore, the criterion attribute is the stereotype (belief about a group), and stereotype (in)accuracy was assessed with respect to the *stereotype of the stereotype* (i.e., the accuracy of the belief about others' stereotypes). Unfortunately, the "real" stereotype was assessed by a haphazard (rather than representative sample or population) sample of undergraduates, so this study suffers from a mismatch between stereotype and criterion sample.

In all cases, on average, people exaggerated others' stereotypes. For example, people were asked to: 1) estimate the proportion of males and females, respectively, who scored higher on their verbal than math SAT (the actual stereotypes); and then 2) estimate the estimate (stereotype of stereotypes) provided by males and females. On average, across ratings, people estimated a 29% difference between males and females (stereotypes), but estimated that other people would estimate a 45.1% difference—they exaggerated others' stereotypes. Subsequent studies replicated this pattern for gender, college major stereotypes, and for U.S. regional stereotypes.

STEREOTYPE (IN)ACCURACY: KNOWNS, UNKNOWNS, AND EMERGING CONTROVERSIES

In this section, we review broad patterns in the existing data on stereotype accuracy. We also highlight the implications of those data, contradictions and controversies, and directions for future research.

Pervasive Stereotype Accuracy

For the racial, ethnic, gender, political, occupational, and college major and residence stereotypes that have been assessed, stereotype accuracy correlations are among the largest and most replicable effects in all of social psychology (see Table 2.1). They are typically far larger than the effect sizes routinely interpreted as support for far more famous social psychological hypotheses (bystander effect, attribution errors, mere exposure, etc.). We use the term "pervasive stereotype accuracy" here to refer to the widespread evidence of at least some accuracy, and sometimes quite high accuracy, found in nearly every study that has assessed stereotype accuracy, with the exception of studies of national character stereotypes that have used the Big Five as criteria. Nonetheless, there is no evidence that stereotypes are *perfectly* accurate. Furthermore, even within a single study, accuracy levels may vary, not just across judgments or perceivers, but, because accuracy is a multifaceted construct, across the four types of accuracy described previously. Nonetheless, the evidence of pervasive stereotype accuracy is inconsistent with virtually all perspectives defining stereotypes as inaccurate, or emphasizing their inaccuracy.

Several methodological aspects of the empirical research on stereotype accuracy are worth noting because they bear on the generalizability of these results. Although many studies assessed the accuracy of undergraduates' stereotypes, a substantial minority assessed the accuracy of samples of adults, and some have included representative samples of perceivers. Some of the highest levels of accuracy occurred with these adult samples, suggesting that the levels of accuracy obtained do not represent some artifact resulting from the study of undergraduate samples.

Second, the studies used a wide variety of criteria: U.S. Census data, self-reports, Board of Education data, nationally representative surveys, locally representative surveys, U.S. government reports, etc. The consistency of the results across studies, therefore, does not reflect some artifact resulting from use of any particular criteria.

Third, the studies examined a wide range of stereotype content: beliefs about demographic characteristics, academic achievement, personality, attitudes, beliefs, and behavior. The consistency of the results across studies, therefore, does not reflect some artifact resulting from a particular type of stereotype content.

Table 2.1 makes three points vividly clear. The first is that when criterion samples are mismatched to the stereotype (use of haphazard criterion samples), evidence

Table 2.1 Stereotypes are (Usually) More Valid than Most Social Psychological Hypotheses

	Proportion of Social Psychological Effects $r \geq .30, r \geq .50$	Proportion of Consensual Stereotype Accuracy Correlations[2] $r \geq .30, r \geq .50$	Proportion of Personal Stereotype Accuracy Correlations $r \geq .30, r \geq .50$
Studies with criterion samples matched well to the <u>assessed stereotype</u>	24%[1], 5%[1]		
Race		95%, 95%[3]	47%, 18%[4]
Gender		100%, 94%[5]	79%, 58%[6]
Political stereotypes		100%, 100%[7]	89%, 33%[8]
National stereotypes		43%, 43%[9]	None assessed
Other stereotypes[10]		100%, 96%[11]	100%, 63%[12]
Studies with haphazard <u>criterion samples</u>			
Race		None assessed	None assessed
Gender		80%, 80%[13]	None assessed
Political stereotypes		None assessed	None assessed
National character stereotypes using a big five personality measure		17%, 4%[14]	None assessed
Other stereotypes		63%, 50%[15]	64%, 45%[16]

Notes

1 Data obtained from Richard et al.'s (2003) review of meta-analyses including thousands of studies. Effects are in terms of the correlation coefficient, r.

2 When papers reported correlations of stereotypes with multiple criteria (e.g., self-reports and observer reports), we averaged them to produce a single correlation for this table. Even though papers often did not report these, they are included here if consensual stereotype accuracy correlations were computable from their published data.

3 Based on 21 correlations obtained from three studies (Ashton & Esses, 1999; McCauley & Stitt, 1978 (based on raw data reported in McCauley, 1995); Ryan, 1996).

4 Based on 17 correlations obtained from two studies (Ashton & Esses, 1999; Ryan, 1996).

5 Based on 33 correlations obtained from 14 studies (Beyer, 1999; Briton & Hall, 1995; Cejka & Eagly, 1999; Diekman et al., 2002; Hall & Carter, 1999; McCauley et al., 1988, McCauley & Thangavelu, 1991; Swim, 1994).

6 Based on 19 correlations obtained from 10 studies (Beyer, 1999; Diekman et al., 2002; Hall & Carter, 1999; McCauley et al., 1988; McCauley & Thangavelu, 1991).

7 Based on 13 correlations obtained from five studies (Chambers et al., 2006; Chambers & Melnyk, 2006; Graham et al., 2012).

8 Based on 9 correlations. One was reported in Judd & Park (1993). The other 8 were based on data collected for Chambers et al., 2006 and Chambers & Melnyk, 2006, but were not reported in the original papers.

9 Based on 7 correlations reported in two studies (Heine et al., 2008; Lonnqvist et al., 2012).

10 Includes Chan et al. (2012), which had both haphazard and representative samples as criteria, and found little difference in accuracy results between them.

11 Based on 1 correlation reported in Cejka & Eagly (1999) and 25 correlations (each for different samples) reported in Chan et al. (2012, results based on country in their Table 1).

12 Based on 24 correlations between stereotype beliefs about attributes and criteria reported in two studies (Judd et al., 1991; Ryan & Bogart, 2001). Both papers also assessed accuracy in beliefs about variability, but those are not included here.

13 Based on 5 correlations obtained from four studies (Allen, 1995; Löckenhoff et al., 2014; Martin, 1987).

14 Based on 141 correlations reported in five papers, each including multiple separate samples from different countries and cultures (Allik et al., 2014; Hrebickova & Graf, 2014; McCrae et al., 2013; Realo et al., 2009; Terracciano et al., 2005).

15 Based on 8 correlations reported in two studies (Clabaugh & Morling, 2004; Rogers & Wood, 2010). Rogers & Wood (2010) include Alaska and Hawaii, which the text discusses as outliers (reporting the results with the outliers removed would have increased the number of correlations greater than .3 and .5).

16 Based on 11 correlations reported in two studies (Clabaugh & Morling, 2004; Rogers & Wood, 2010).

of stereotype accuracy is not as strong as when criterion samples are well matched to the stereotype. Or, put differently, mismatched criterion samples make stereotypes *appear* to be less accurate than they really are. Second, however, the difference is not particularly large, and *even among studies with mismatched criterion samples, stereotype accuracy is one of the largest effects in social psychology, with one major exception.* The third point clearly apparent in Table 2.1 is that studies using the Big Five as criteria indicate stereotypes of national character are almost completely inaccurate.

Pervasive Inaccuracy has been Found for National Character Stereotypes when Assessed against Measures of the Big Five

What explains the differences between the high levels of accuracy in many other stereotypes and the inaccuracy of these other stereotypes? Although we are currently in no position to answer this question, we consider several possibilities.

Sampling explanations. Most of the national character studies finding inaccuracy have compared stereotypes to haphazard criterion samples. Perhaps the use of convenience samples accounts for some of the inaccuracy, i.e., it is possible that the observed stereotypes were more accurate than indicated in these studies.

Despite the appeal of this explanation, we doubt it goes very far for two reasons. First, for all other stereotypes, use of haphazard samples reduces stereotype accuracy, but not very much. We have no reason to believe this would be different for national character stereotypes. If it is not different, sampling explanations would likely account for only a small portion of the difference in accuracy obtained in studies of national character stereotypes as compared to other stereotypes.

Second, one study (Chan et al., 2012) directly examined this issue for age stereotypes of Big Five personality traits, and found high consensual stereotype accuracy correlations regardless of whether the criterion samples were haphazard or representative. On the other hand, Chan et al. (2012) examined age stereotypes so that its relevance to research on national character stereotypes is unclear.

Measurement explanations. Another methodological consideration is the criterion measure itself. Two studies of national stereotype accuracy that use some measure *other than the Big Five* show high (Heine et al., 2008) or moderate (Lonnqvist et al., 2012) levels of accuracy. This raises the possibility that there is something problematic with the Big Five measures. This, however, seems unlikely for two reasons: 1) The Big Five measures are among the most highly validated questionnaires in all of psychology (e.g., Costa & McCrae, 2008); and 2) two studies have found considerable accuracy in age and gender stereotypes using the Big Five as criteria (Chan et al., 2012; Löckenhoff et al., 2014). Thus, there is no evidence of pervasive failure or invalidity of the measure.

Nonetheless, it is still possible that people's stereotypes are generally more accurate when target group behavior, rather than target group self-reports, are used as criteria. Only one study investigated this question (Heine et al., 2008), and for only one personality characteristic (conscientiousness). Future research should examine whether this pattern replicates in other samples, with other behavioral criteria, and on other personality characteristics. However, the problem does not seem to be purely associated with reliance on self-reports, as some studies showing pervasive inaccuracy have used both self-reports and observer reports about specific individuals in a culture as criteria (McCrae et al., 2013; Terracciano et al., 2005).

The reference group effect. Heine et al. (2008) suggested that members of different cultures implicitly use different standards of evaluation from one another because they implicitly arrive at ratings by comparison with *the typical person in their own culture* (i.e., the *reference group effect*). Thus, if a culture is generally high in conscientiousness, an individual member who is similarly high may merely rate themselves as average. This is why they recommend behavioral criteria. Nonetheless, McCrae et al. (2013) point out that if this problem was pervasive, nearly all cultures would, on average, rate themselves as average on personality traits—something which much data demonstrates is clearly not the case. Furthermore, Heine et al. (2008) is just a single study focusing on a single trait (conscientiousness). Thus, the extent to which the reference group effect explains the lack of apparent accuracy in studies of national character stereotypes is unclear.

Contact. Contact has long been recommended as one way to reduce prejudice, in part because it can potentially correct unjustified negative stereotypes (Allport, 1954/ 1979). Although the contact hypothesis has a long and complex history and has rarely been unequivocally supported without the inclusion of many other conditions (e.g., Pettigrew, 1998), it is indeed possible that contact increases stereotype accuracy (see also *Understudied Process: The "Stereotypes As Knowledge" Hypothesis*, later in this chapter). Consistent with this idea, Lonnqvist et al. (2012) found that the more contact people had with the outgroup, the more accurate their stereotypes.

Of course, people generally have much contact with individuals within their own national culture, whereas these, too, are consistently found to be inaccurate (see Table 2.1). Thus, the role of contact in increasing accuracy is unclear and remains an important question for future research.

National character stereotypes as false social constructions. This is a common interpretation of national character stereotypes that produce little evidence of accuracy (e.g., Allik et al., 2014; Terracciano et al., 2005). Some indirect support for this notion is provided by national character stereotype studies showing that people in different countries are often more likely to agree with one another about the national character of a target country than they are to be accurate (Allik et al., 2014; Realo et al., 2009; Terracciano et al., 2005).

However, evidence of inaccuracy supports *no particular explanation* by default. Thus, our view is that it behooves those arguing that national character stereotypes are false social constructions to provide specific positive evidence demonstrating how specific social processes lead national character stereotypes astray.

What is needed to help resolve these issues. Empirically based explanations for the lack of evidence of accuracy in many studies of national character stereotypes could be advanced by:

1. Research using criteria other than Big Five personality inventories. At minimum, such work will indicate whether inaccuracy is limited to, or simply lower when compared to such inventories, rather than when other measures, such as behavior, values, and socioeconomic characteristics are used as criteria.
2. Greater use of more representative criterion samples. This would indicate whether any of the inaccuracy found is some artifact resulting from nonrepresentative criterion samples.
3. Research directly assessing the role of the reference group effect as a source of national stereotypes.

4. Additional research on factors increasing (contact? consensually shared standards?) and reducing (conflict, propaganda, belief extremity, etc.) stereotype accuracy.
5. Empirical research demonstrating how social constructivist processes specifically lead to particular inaccurate stereotypes under naturalistic, real world (as opposed to artificial laboratory) conditions.

Consensual and Personal Stereotype Accuracy Correlations and the Wisdom of Crowds

The average judgment of a group is often more accurate than the judgment of most or all of the individual members of that group (Surowiecki, 2004). This has been true when estimating the weight of an ox, the number of beans in a jar, or the location of a sunken submarine. And it is generally true when estimating the characteristics of social groups. Consensual stereotypes—which average over all perceivers in a sample—are typically more accurate than personal stereotypes.

But this only works 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. One way or another, social reality appears to be a major influence on stereotypes.

The wisdom of crowds may help explain why consensual stereotype correlations are often .80 and higher. 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 primarily *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 social psychology.

The published stereotype accuracy literature provides far less information about personal stereotype accuracy than about consensual stereotype accuracy. Personal stereotype accuracy correlations have also been quite high (see Table 2.1). Thus, with respect to understanding variations in the traits, attitudes, or characteristics of target groups, people are generally at least moderately accurate, and frequently very accurate.

Stereotype Accuracy as Discrepancies from Criteria

Consensual stereotype discrepancies were within 10% or .25 SD of criteria for either a majority or plurality of estimates in most studies of race, gender, occupational, or college major stereotypes. These estimates benefit from the same "wisdom of crowd" effects as do consensual stereotype correlations and they, too, fail to support declarations of "stereotype inaccuracy" or any perspective suggesting that stereotypes are primarily false cultural myths.

We are aware of only a single study that clearly assessed personal stereotype discrepancies (Ashton & Esses, 1999) and another that did so summing over all judgments (Diekman et al., 2002). In the absence of more evidence, we are in no position to reach general conclusions about degrees of (in)accuracy in personal stereotype discrepancies. Clearly, this is an area that warrants more empirical research.

Stereotypes Do Not Necessarily—Or Even Frequently—Exaggerate Real Differences

The assumption that stereotypes exaggerate real differences has long and deep roots in social psychology. Allport (1954/1979) and Campbell (1967) defined stereotypes as exaggerations. Also known as the "kernel of truth" hypothesis (e.g., Schneider, 2004), this was long the only perspective that permitted researchers to acknowledge that people were not always completely out of touch with social reality, while simultaneously allowing researchers to position themselves well within the longstanding traditions emphasizing stereotype error and bias.

Frequently, when social psychologists make the "stereotypes exaggerate real differences" claim, they cite Tajfel and Wilkes (1963), which found that when lines were organized into two categories, people exaggerated the differences between the two categories and underestimated the within-category differences among the lines. Although this was not a study of stereotypes, the interpretation is often that this study captured basic processes of perception, such that categorization (including social categorization such as stereotypes) will exaggerate real differences. Regardless, whether *stereotypes* exaggerate real differences cannot be tested against studies of perceptions of lines. And most studies of race and gender stereotypes show either a greater tendency to underestimate than exaggerate real differences, or no systematic tendency to exaggerate or underestimate.

Does Exaggeration Ever Occur?

Exaggeration can and should live on, however, not as a defining feature of stereotypes, but as a *hypothesis* to be tested independently for specific stereotypes. As such, there is some evidence of exaggeration in stereotypes.

Scattershot exaggeration in demographic stereotypes. Although most studies of racial and ethnic stereotypes yield more evidence of *underestimating* real differences, two provided more evidence of exaggeration (Ashton & Esses, 1999; Ryan, 1996). More research is needed to determine whether this pattern is just a random blip in an otherwise pervasive pattern of underestimation, or reflects some sort of currently poorly understood systematic variation.

There has been only a single paper examining the accuracy of age stereotypes, which were exaggerations (Chan et al., 2012). Although it was only a single paper, it included samples of adults from 26 different countries, totaling over 3,000 respondents. Of course, whether people also hold exaggerated age stereotypes regarding characteristics other than the Big Five personality traits is unknown.

Consistent exaggeration of political stereotypes. The best case for pervasive exaggeration, however, is in political stereotypes. People exaggerate the attitudes, beliefs, values, morals, and traits of various political groups. They exaggerate differences between liberals and conservatives, between Democrats and Republicans, between those who oppose and favor legalized abortion, between opponents and proponents of affirmative action, between those who supported and opposed the Vietnam War, between partisans on opposite sides of a racial violence incident – even between the literature preferences of both sides involved in the "Western Cannon debate" (Robinson et al., 1995; Robinson & Keltner, 1996; Sherman, Nelson, & Ross, 2003). This pattern has been found among representative samples (Graham et al., 2012; Judd & Park 1993; Westfall et al., 2015), college student samples (e.g., Chambers & Melnyk, 2006; Dawes et al., 1972; Robinson et al., 1995; Sherman et al., 2003), college instructors (Robinson & Keltner, 1996), and members of political advocacy groups (e.g., Chambers et al., 2006; Robinson et al., 1995).

Westfall et al. (2015) argue that such exaggeration derives from three sources: categorization, identification, and extremity of own attitude. However, the pure categorization hypothesis is best tested by assessment of whether independents exaggerated (because they are not identified with a party, and typically also have less extreme attitudes). Westfall et al. (2015) found that independents perceived considerably smaller differences between Democrats and Republicans than did Democrats and Republicans. Although it is unclear from the reported data whether independents engaged in any exaggeration at all, if they did, it was quite modest. Thus, there was, at best, weak support for the hypothesis that pure categorization leads people to exaggerate real differences.

Westfall et al. (2015) provided stronger support for the *identification* and *attitude extremity* hypotheses. The more strongly people were identified with either political party, and the more extreme their own attitudes, the more they exaggerated real differences (these were independent effects). This raises the possibility that exaggeration does occur more broadly. Perhaps exaggeration is not restricted to political perceptions, but occurs frequently *when groups are in conflict*, at least among those strongly identified with their group, regarding core values (Chambers et al., 2006) and extremists. This raises the possibility, for example, that strongly identified and extremist members of groups with recent histories of conflict (e.g., Israelis and Palestinians, Indians and Pakistanis, environmentalists and fossil fuel producers, etc.), may tend to hold exaggerated stereotypes of one another.

Exaggerated stereotypes of other people's stereotypes. Last, there is also evidence of exaggeration of *other people's stereotypes* (Rettew et al., 1993). Although this was only a single paper, and focused exclusively on undergraduate samples, it reported four separate studies, and replicated the exaggeration pattern across several stereotypes of stereotypes (gender, college major, U.S. regional stereotypes). This study may be particularly important because it could help explain why, for nearly 100 years, social psychology has clung to the assumptions that stereotypes are inaccurate and exaggerate real differences.

Perhaps many social scientists' beliefs about stereotypes can be understood as reflecting the *psychological phenomenon* discovered in Rettew et al.'s (1993) research showing that people exaggerate other's stereotypes. Social scientific claims of stereotype inaccuracy or exaggeration bear little or no connection to the accumulating (and by now overwhelming) *scientific evidence* of pervasive stereotype accuracy and only spotty evidence of exaggeration (limited to political stereotypes and a few other studies). That is, many social science perspectives emphasizing stereotype inaccuracy appear to be exaggerated, inaccurate, rigidly resistant to change in the face of relentless disconfirming evidence, and maintain their conclusions by virtue of a very selective focus on studies and findings that confirm the a priori belief in the irrationality and badness of stereotypes. Defining stereotypes as inaccurate is logically incoherent (which some might suggest means that it is irrational). In other words, Rettew et al. (1993) raises the possibility that stereotypes have been stereotyped!

Un(der)studied Stereotypes

Although there has been a great blossoming of research on stereotype accuracy, many stereotypes remain unstudied or understudied. We are aware of no research on the accuracy of social class, religion, or sexual orientation stereotypes. Perceivers have, so far, been primarily American, Canadian, European, and, to a lesser extent, Asian samples. There is little data on the (in)accuracy of Middle Eastern, African, and South American

perceiver groups. Given the potential for such stereotypes to play a role in certain major internal or international conflicts, such research could be invaluable.

Understudied Process: The "Stereotypes As Knowledge" Hypothesis

The sources and causes of (in)accuracy in social stereotypes have yet to receive much empirical attention. The clear evidence of accuracy strongly suggests that a major source of many 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 people learn quite a lot about what many groups are like. That is, stereotypes are much like other generalizations (e.g., Florida is warmer than Montana; gazelles run faster than gorillas).

Beyond the mere demonstrations of accuracy, several lines of research are converging around a *stereotypes as knowledge* hypothesis. If stereotypes are (at least in part) a form of everyday knowledge, then more informed people should hold more accurate stereotypes. Consistent with this hypothesis, more highly educated people held more accurate racial stereotypes (Kaplowitz et al., 2003).

Another study consistent with the *stereotypes as knowledge hypothesis* found that the stereotypes held by multicultural migrants (ethnic Finns who had lived in Russia but who had migrated back to Finland; Russians emigrants to Finland) were considerably more accurate than many other national stereotypes (Lonnqvist et al., 2012). This increased accuracy might have occurred because the migrants had more direct experience with—and therefore, increased knowledge of—both cultures. Being encouraged to recognize, as opposed to ignore, group differences also increased racial stereotype accuracy (Wolsko et al., 2000). All of these findings are consistent with the *stereotypes as knowledge* hypothesis.

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 other nations, and rarely have reason to study the academic literature on national character. 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).

Understudied Process: The Egalitarian Denial Hypothesis

The *egalitarian denial hypothesis* is that, in their attempt to be good, decent, unbigoted egalitarians, many people are motivated to deny real group differences. Because groups often do really differ on many attributes, such people should often inaccurately see groups as differing *less* than they really do. This may help explain the widespread evidence of *underestimation* of real differences.

The *egalitarian denial hypothesis* is not intended to describe people in general. Instead, it predicts who will be more or less accurate: lower accuracy among people who deny group differences, and higher accuracy among those who do not deny group differences. Who is likely to deny group differences? The *egalitarian denial hypothesis* predicts two groups: People on the far left of the political spectrum (consider, e.g., the Marxist emphasis on equality at the expense of freedom; e.g., Rokeach, 1968), and people

(regardless of ideology) highly motivated to be or appear egalitarian. These hypotheses should be tested in future research. Nonetheless, some existing research supports the *egalitarian denial hypothesis*:

- Ashton and Esses (1999) show that people very low in right-wing authoritarianism—i.e., people likely to be politically left-wing—inaccurately underestimated real differences. And finding that intelligence did not matter for this group strongly hints at the possibility that this denial of differences is motivated, rather than merely a reflection of ignorance. This might be a case where motivation overrides knowledge to produce inaccurate stereotypes.
- Wolsko et al. (2000) show that when people are instructed to adopt a color-blind mindset, their stereotypes more strongly underestimate real differences than when instructed to adopt a multicultural mindset. Being directly instructed to discount group differences caused greater underestimation inaccuracy—a result consistent with the *egalitarian denial hypothesis.*

Of course, even motivated egalitarianism will not *always* lead to underestimation of real differences. Motivated egalitarianism likely leads to hostility toward those perceived as opposing egalitarianism—and, as such, likely contributes to exaggeration of differences between their own and opposing political groups.

Underreported Data in Existing Research

Personal stereotype accuracy. Few studies report data regarding all four types of stereotype (in)accuracy identified here. Our view is that, absent compelling reasons not to, the default should be to do so. Far more studies have reported information on consensual stereotype accuracy than on personal stereotype accuracy (this has been true for both discrepancies and correlations). Personal stereotype accuracy analyses require computing discrepancies and correlations for each individual participant, separately. Although requiring greater effort, it is only by performing such analyses that we will be able to reach broad and general conclusions about the (in)accuracy of individual perceivers' stereotypes. Although the research conducted thus far suggests that, in general, even personal stereotypes are often at least moderately accurate—and more accurate than most social psychological hypotheses—there is less data that bears on personal stereotypes than on consensual stereotypes.

Advantages and disadvantages of componential approaches. Because accuracy has many different aspects, several models have been proposed for separating accuracy into separate *components* (e.g., ingroup bias, exaggeration—Cronbach, 1955; Judd & Park, 1993; Kenny, 1994). Although its advocates often declare that accuracy can *only* be assessed with components, our view is that componential approaches provide many valuable insights into the processes affecting (in)accuracy, but so do many non-componential approaches (see Jussim, 2012 for a review).

One disadvantage of most componential approaches, at least as they have been used in stereotype accuracy research, is they often *obscure* many aspects of accuracy. Typically, scores across many different perceptions are summed or averaged, and then ANOVA is performed on the total scores (e.g., Judd & Park, 1993; Wolsko et al., 2000). This is invaluable for determining the main effects and interactions for different aspects or components of stereotypes. But summing or averaging overall attributes produces a very

high level of aggregation, which may, through the wisdom of crowd effects, obscure inaccuracies with respect to specific perceptions, and especially personal discrepancies. For example, if Fred overestimates the likelihood of Whites having a college degree by 30%, but *underestimates* the likelihood of them earning over $50,000 by 30%, an average will conclude that Fred is accurate. By our standards, however, Fred would be considered inaccurate on both judgments. Although this type of aggregation is common to componential approaches, it is not restricted to them (e.g., Diekman et al., 2002). However useful such aggregation may be, it also obscures finer grained analyses of accuracy. Our view is that if researchers wish to perform componential analyses, they do so *in addition to*, *not instead of*, reporting results for each stereotype attribute studied.

Consequences of Accurate and Inaccurate Stereotypes

Although a vast body of literature has assessed the effects of social stereotypes (see Schneider, 2004 for a review), research has only just begun to address the specific consequences of holding accurate versus inaccurate stereotypes. Westfall et al. (2015) found, for example, that those who most greatly exaggerated partisan political differences—i.e., those who were most *inaccurate*—were *most* likely to be actively engaged in political action (such as donating money or working for a campaign). They concluded that it seemed likely that those who perceived greater polarization feel more of a need to support and defend their political group from the (perceived) threat posed by the opposing political party. This raises the possibility that inaccurate stereotypes are at least one source of the gridlock that has plagued American national politics for some time—if it is primarily extremists who are involved in politics, and if they demonize the opposition's views as so extreme as to be "beyond the pale" (e.g., Republicans who liken mainstream Democrats to communists, or Democrats who liken mainstream Republicans to fascists and religious zealots), there may be little interest or perceived ability to reach common ground for moving forward.

Another consequence of increasing stereotype accuracy may be to increase multicultural sensitivity. Although different proponents of multiculturalism have different rationales, themes typically emphasize understanding and respecting the beliefs, values, and practices of people from different groups and background than one's own. Respecting others' values, beliefs, and practices makes sense primarily if one has a reasonably clear (accurate) sense of what those values, beliefs, and practices are. Or, put differently, part of multiculturalism can be viewed as a call for increasing *stereotypes as knowledge* relative to *stereotype inaccuracy.*

Limitations to this Review

Many aspects of stereotypes cannot be assessed for their accuracy (prescriptive stereotypes, stereotypes as attitudes), and were therefore not reviewed here. It is possible that biases and motivated distortions operate more strongly for beliefs with no objective anchor in reality. Similarly, because the social sciences have produced little clear, unambiguous understanding of the origins of group differences, it is currently not possible to assess the accuracy of people's beliefs about the *sources or explanations* for group differences.

This review also did not address the role of stereotypes in person perception (see our chapter in the prior edition of this Handbook for a more detailed review; Jussim et al., 2009). Although the dominant view in social psychology was once that stereotype biases

powerfully and pervasively distort person perception (e.g., Darley & Fazio, 1980; Fiske & Neuberg, 1990), multiple meta-analyses (see Jussim, 2012 for a review of those meta-analyses) show that stereotype biases in person perception tend to be quite small whereas reliance on individuating information is another one of the largest effects in social psychology. Considerable evidence has accumulated that people usually (though not always) apply their stereotypes in a largely rational manner, relying heavily on stereotypes in the absence of clear individuating information, but relying heavily on relevant individuating information when it is available (e.g., Baron, Albright, & Malloy, 1995; Crawford, Jussim, Madon, Cain, & Stevens, 2011; Locksley, Borgida, Brekke, & Hepburn, 1980; see reviews by Jussim, 2012; Jussim et al., 2009; Kunda & Thagard, 1996).

CONCLUSION

It ain't what you don't know that gets you into trouble.
It's what you know for sure that just ain't so.

(Mark Twain)

For decades, researchers declared stereotypes to be inaccurate on the basis of little or no evidence. This claim was repeated so frequently that it was presumed to be true in the absence of evidence. Blanket, unqualified declarations of stereotype inaccuracy are unjustified. To be sure, evidence of stereotype inaccuracy clearly exists. But it exists amidst a body of literature that also includes abundant evidence of stereotype accuracy. In 2014, when this chapter was written, it is only possible to maintain blunt declarations of stereotype inaccuracy by cherry-picking studies demonstrating inaccuracy, systematically ignoring studies that demonstrate accuracy, and employing double standards that permit the derogation of accuracy studies that are at least as methodologically sound as inaccuracy studies.

Thus, our final recommendation is that researchers simply stop declaring stereotypes, *in general*, to be inaccurate (obviously, declaring any particular perceiver or perceiver group's stereotypes to be inaccurate in a particular study may be completely justified). Instead, progress in the social psychology of intergroup relations will be elevated by scholarship that recognizes that stereotype (in)accuracy is an empirical question, and that degrees of (in)accuracy may vary across perceivers, perceiver groups, target groups, and target attributes. As that notion begins to take hold, it will open up a wealth of socially and theoretically important questions about the causes and consequences of stereotype accuracy and inaccuracy. We hope that more social scientists accept the challenge of trying to obtain *empirical* answers to those questions.

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.

Allik, J., Alyamkina, E., & Mescheryakov, B. (2014). The personality stereotypes of three cohabiting ethnic groups: Erzians, Mokshans, and Russians. *Cross-Cultural Research*, *49*, 111–134.

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

Altemeyer, B. (1996). *The authoritarian specter*. Cambridge, MA: Harvard University Press.

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

Aronson, E. (2008). *The social animal* (10th 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*, 225–236.

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

Baron, R. M., Albright, L., & Malloy, T. E. (1995). The effects of behavioral and social class information on social judgment. *Personality and Social Psychology Bulletin, 21*, 308–315.

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

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.

Campbell, D. T. (1967). Stereotypes and the perception of group differences. *American Psychologist, 22*, 817–829.

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*, 413–423.

Chambers, J. R., & Melnyk, D. (2006). Why do I hate thee? Conflict misperceptions and intergroup mistrust. *Personality and Social Psychology Bulletin, 32*, 1295–1311.

Chambers, J. R., Baron, R. S., & Inman, M. L. (2006). Misperceptions in intergroup conflict: Disagreeing about what we disagree about. *Psychological Science, 17*, 38–45.

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.

Clabaugh, A., & Morling, B. (2004). Stereotype accuracy of ballet and modern dancers. *Journal of Social Psychology, 144*, 31–48.

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

Costa, P. T., Jr., & McCrae, R. R. (2008). The revised NEO Personality Inventory (NEO-PI-R). In G. J. Boyle, G. Matthews, & D. H. Saklofske (Eds.), *The Sage handbook of personality theory, and assessment: Personality measurement and testing, Volume 2.* London: Sage Publications.

Costa, P. Jr., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology, 81*, 322–331.

Crawford, J. T. (2012). The ideologically objectionable premise model: Predicting biased political judgments on the left and right. *Journal of Experimental Psychology, 48,* 138–151.

Crawford, J. T., Modri, S. A., & Motyl, M. (2013). Bleeding-heart liberals and hard-hearted conservatives: Subtle political dehumanization through differential attributions of human nature and human uniqueness traits. *Journal of Social and Political Psychology, 1*, 86–104. doi:10.5964/jspp.v1i1.184

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,* 529–542.

Cronbach, L. J. (1955). Processes affecting scores on "understanding of others" and "assumed similarity." *Psychological Bulletin, 52*, 177–193.

Darley, J. M., & Fazio, R. H. (1980). Expectancy-confirmation processes arising in the social interaction sequence. *American Psychologist, 35*, 867–881.

Dawes, R., Singer, D., & Lemons, F. (1972). An experimental analysis of the contrast effect and its implications for intergroup communication and the indirect assessment of attitude. *Journal of Personality and Social Psychology, 21*, 281–295.

DeNavas-Walt, C., & Proctor, B. D. (2014). *Income and poverty in the United States: 2013. Current Population Reports.* Retrieved June 29, 2015, from: https://www.census.gov/content/dam/Census/library/publications/2014/demo/p60-249.pdf.

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.

Duarte, J. L., Crawford, J. T., Stern, C., Haidt, J., Jussim, L., & Tetlock, P. E. (2014). Political diversity will improve social psychological science. Manuscript submitted for publication.

Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. In D. Gilbert, S. T. Fiske, & G. Lindzey (Eds). *The handbook of social psychology* (Vol. 2, 4th ed., pp. 357–411). New York: 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. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 1–74). New York: Academic Press.

Fiske, S. T., & Taylor, S. E. (1984). *Social cognition.* New York: Random House.

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition* (2nd ed.). New York: McGraw-Hill.

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*, 652–670.

Graham, J., Nosek, B. A., & Haidt, J. (2012). The moral stereotypes of liberals and conservatives: Exaggeration of differences across the political spectrum. *PLoS One, 7*, e50092. doi:10.1371/journal.pone.0050092

Haidt, J. (2012). *The righteous mind: Why good people are divided by religion and politics.* Pantheon Books: New York.

Hall, J. A., & Carter, J. D. (1999). Gender-stereotype accuracy as an individual difference. *Journal of Personality and Social Psychology, 77*, 350–359.

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.

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

Heine, S. J., Buchtel, E. E., & Norenzayan, A. (2008). What do cross-national comparisons of personality traits tell us? The case of conscientiousness. *Psychological Science, 19,* 309–313.

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.

Jones, E. E. (1985). Major developments in social psychology during the past five decades. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (Vol 1, 3rd ed., pp. 47–107). New York: Random House.

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., & 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*, 168–187.

Judd, C. M., & Park, B. (1993). Definition and assessment of accuracy in social stereotypes. *Psychological Review, 100*, 109–128.

Judd, C. M., Ryan, C. S., & Park, B. (1991). Accuracy in the judgment of in-group and out-group variability. *Journal of Personality and Social Psychology, 61*, 366–379.

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., & Stevens, S. T. (2014). Ideological bias in social psychological research. In J. Forgas, W. Crano, & K. Fiedler (Eds.), *Social psychology and politics* (pp. 91–109). New York: Psychology Press.

Jussim, L. Crawford, J. T., Stevens, S. T., & Anglin, S. M. (2014). Political distortions in the social psychology of intergroup relations. In P. Valdesolo & J. Graham (Eds.), *Claremont Symposium on Applied Social Psychology: Bridging ideological divides.*

Kaplowitz, S. A., Fisher, B. J., & Broman, C. L. (2003). How accurate are perceptions of social statistics about Blacks and Whites? *Public Opinion Quarterly, 67,* 237–243.

Katz, D., & Braly, K. (1933). Racial stereotypes of one hundred college students. *Journal of Abnormal and Social Psychology, 28*, 280–290.

Keltner, D., & Robinson, R. J. (1996). Extremism, power, and the imagined basis of social conflict. *Current Directions in Psychological Science, 5*, 101–105.

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

Kruglanski, A. (1989). The psychology of being "right": The problem of accuracy in social perception and cognition. *Psychological Bulletin, 106,* 395–409.

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*, 232–254.

Löckenhoff, C. E., Chan, W., McCrae, R. R., De Fruyt, F., Jussim, L., De Bolle, M., & Pramila, V. S. (2014). Gender stereotypes of personality: Universal and accurate? *Journal of Cross-Cultural Psychology, 45*, 675–694.

Locksley, A., Borgida, E., Brekke, N., & Hepburn, C. (1980). Sex stereotypes and social judgment. *Journal of Personality and Social Psychology, 39*, 821–831.

Lonnqvist, J.-E., Yijala, 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.

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*, 929–940. McCauley, C., & Thangavelu, K. (1991). Individual differences in sex stereotyping of occupations and personality traits. *Social Psychology Quarterly, 54*, 267–279.

McCauley, C., Jussim, L., & Lee, Y. T. (1995). Stereotype accuracy: Toward appreciating group differences. In Y. T. Lee, L. Jussim, & C. R. McCauley (Eds.), *Stereotype accuracy: Toward appreciating group differences* (pp. 293–312). Washington, DC: American Psychological Association.

McCauley, C., Stitt, C. L., & Segal, M. (1980). Stereotyping: From prejudice to prediction. *Psychological Bulletin, 87*, 195–208.

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.

McCrae, R. R., & Allik, J. (Eds.). (2002). *The five-factor model of personality across cultures*. New York: Kluwer Academic/Plenum Publishers.

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.

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.

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

Martin, C. L. (1987). A ratio measure of sex stereotyping. *Journal of Personality and Social Psychology, 52*, 489–499.

Merton, R. K. (1948). The self-fulfilling prophecy. *The Antioch Review, 8*, 193–210.

Miller, D. T. & Turnbull, W. (1986). Expectancies and interpersonal processes. *Annual Review of Psychology, 37*, 233–256.

Pettigrew, T. F. (1998). Intergroup contact theory. *Annual Review of Psychology, 49*, 65–85.

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.

Rettew, D. C., Billman, D., & Davis, R. A. (1993). Inaccurate perceptions of the amount others stereotype: Estimates about stereotypes of one's own group and other groups. *Basic and Applied Social Psychology, 14*, 121–142.

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.

Robinson, R. J., & Keltner, D. (1996). Much ado about nothing? Revisionists and traditionalists choose an introductory English syllabus. *Psychological Science, 7*, 18–24.

Robinson, R. J., Keltner, D., Ward, A., & Ross, L. (1995). Actual versus assumed differences in construal: "Naïve realism" in intergroup perception and conflict. *Journal of Personality and Social Psychology, 68*(3), 404–417.

Rogers, K. H., & Wood, D. (2010). Accuracy of United States regional personality stereotypes. *Journal of Research in Personality, 44,* 704–713.

Rokeach, M. (1968). *Beliefs, attitudes, and values*. San Francisco, CA: Jossey-Bass.

Rosenthal, R. (1991). Effect sizes: Pearson's correlation, its display via the BESD, and alternative indices. *American Psychologist, 46*, 1086–1087.

Ross, L., 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* (Vol. 1, 5th ed., pp. 3–50). Hoboken, NJ: Wiley.

Ryan, C. S. (1996). Accuracy of Black and White college students' in-group and out-group stereotypes. *Personality and Social Psychology Bulletin, 22*, 1114–1127.

Ryan, C. S. (2002). Stereotype accuracy. *European Review of Social Psychology, 13*, 75–109.

Ryan, C. S., & Bogart, L. M. (2001). Longitudinal changes in the accuracy of new group members' in-group and out-group stereotypes. *Journal of Experimental Social Psychology,* 118–133.

Schneider, D. J. (2004). *The psychology of stereotyping*. New York: Guilford Press.

Sherman, D. K., Nelson, L. D., & Ross, L. D. (2003). Naïve realism and affirmative action: Adversaries are more similar than they think. *Basic and Applied Social Psychology, 25*, 275–289.

Snyder, M. (1984). When belief creates reality. *Advances in experimental social psychology, 18,* 247–305.

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.

Stangor, C. (1995). Content and application inaccuracy in social stereotyping. In Y. T. Lee, L. Jussim, & C. R. McCauley (Eds.), *Stereotype accuracy* (pp. 275–292). Washington, DC: American Psychological Association.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52,* 613–629.

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: 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.

Tajfel, H., & Wilkes, A. L. (1963). Classification and quantitative judgment. *British Journal of Psychology,* 54, 101–114.

Terracciano, A., Abdel-Khalek, A. M., Adam, N., Adamovova, 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.

Weinstein, R. S., Gregory, A., & Strambler, M. J. (2004). Intractable self-fulfilling prophecies: Fifty years after Brown v. Board of Education. *American Psychologist, 59*, 511–520.

West, T. V., & Kenny, D. A. (2011). The truth and bias model of judgment. *Psychological Review*, *118,* 357–378.

Westfall, J., Van Boven, L., Chambers, J., & Judd, C. M. (2015). Perceiving political polarization in the United States: Party identity strength and attitude extremity exacerbate the perceived partisan divide. *Perspectives on Psychological Science, 10*, 145–158.

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.