**Abstract:**

In my précis of Social Perception and Social Reality (SPSR), I argued that the social science scholarship on social perception and interpersonal expectancies was characterized by a tripartite pattern: 1. Errors, biases, and self-fulfilling prophecies in person perception were generally weak, fragile, and fleeting; 2. Social perceptions were often quite accurate; and 3. Conclusions appearing throughout the social psychology scientific literature routinely overstated the power and pervasiveness of expectancy effects, and ignored evidence of accuracy. Most commentators concurring with the validity of these conclusions. Two, however, strongly disagreed with the conclusion that the evidence consistently has shown that stereotypes are moderately to highly accurate. Several others, while agreeing with most of the specifics, also suggested that those arguments did not necessarily apply to contexts outside of those covered in SPSR. In this response, I consider both the limitations to the tripartite pattern, the role of politics and confirmation biases in distorting scientific conclusions, common obstructions to effective scientific self-correction, and how to limit them.
Dear Paul,

I have pasted in my cover email from 2/7 at the bottom of this email, for your records. In the next part of this cover letter, I only address the issues raised in your email to me from 2/8.

The Word version I sent did not accept the “track changes” so you can see the many revisions. Obviously, you can read the pdf for the clean version.

**Processistic Fallacy**

You have been reasonable and thoughtful throughout. It is distressing to me that, somehow, on the issue of the processistic fallacy, we seem to be talking past each other. I am not sure if we will see eye to eye on this. Nonetheless, I think it is worth the effort to make the attempt.

I have thoroughly reworked that section to integrate your reservations. In this revision I acknowledge the existence of conditions under which inferring widespread inaccuracy from laboratory revelations of flawed processes could be justified. Those conditions reflect the reservations you raised, as well as other related points you did not raise. Such inferences can be justified when:

- The identified process is known to definitively produce pervasively inaccurate judgments (ala your dice examples)
  AND
- It is empirically shown that
  - people are incapable of overcoming the bias by relying on superior processes
    OR
  - across a widely representative array of situations, people both rely on the flawed process and rarely enlist superior processes.

These latter points go beyond the ones you raised. I have genuinely attempted to grapple with the substance of your resistance to the idea that inferring error in real life from lab studies of process is usually unjustified.

This section, which begins on page 11, makes these two points:
1. if the conditions described above are not met, inferring widespread error under naturalistic conditions from laboratory demonstrations of processes the researchers believe to be flawed is not justified.
2. I could find no research in social perception that has met those conditions.
3. The work described by Terbeck and by Bian&Cimpian did not meet those conditions.

There is also a brief discussion of perspectives by Cohen and by Gigerenzer who have long shown that laboratory demonstrations of error and bias rarely, if ever, actually mean that people will be inaccurate in real life.

I still keep the term processistic fallacy to refer to situations in which researchers essentially leap to “inaccuracy” conclusions on the basis of studies that fail to meet the standards necessary to
justify such generalizations. Under those circumstances, characterizing those studies as “supposedly” capturing flawed processes duly reflects the ideas of folks like Cohen, Gigerenzer and Funder who have shown that time and time again, processes declared flawed by some, and sometimes a great many, researchers, have ultimately been shown to often serve people well. However, it should now also be clear that the term “processistic fallacy” only applies to conclusions of inaccuracy on the basis of process studies that fail to meet the standards necessary to infer widely inaccurate judgments from studies of process, rather than to any conceivable study demonstrating any conceivably massively flawed process.

Am I using a double standard? Your point 3 does not quite put it this way but it comes pretty close. I do think political biases are a problem, but not because of process per se. The key problem are the errors, not the processes.

I think you have my inference processes backwards. My starting points have been the many things social psychologists, psychologists, and other social scientists get demonstrably wrong. The priorities are, first, to figure out what is wrong and what would be right, and then, only secondarily, to figure out why science has gone wrong.

In SPSR itself, I spent more space arguing that it was theoretically-driven confirmation biases (infatuation with bias and error) than I did on political biases (see Chapters 2, 6-9, 10-12; I touch on political biases in Chapters 15, 19, and 21). Kahan’s commentary suggested that certain types of biases, especially political ones, might be more powerful than those that I reviewed in SPSR, and he suggested that I consider that evidence more deeply. I did so, but with a twist: I argued that it probably applied about as much to psychological science as to laypeople. Thus, the response probably has a bit more on political distortion than either the target article or the book.

When one is consistently reaching wrong conclusions, especially conclusions that consistently lean in a particular direction (bias), one is probably doing something systematically wrong. It then becomes reasonable to consider changing how one is going about one’s business. In contrast to, e.g., Terbeck and Bian & Cimpian who concluded that the processes they discussed mean that stereotypes are inaccurate, the reply does not declare that political motivations do explain errors and distortions in the literature. It is, however, one possibility, and the section on Kahan ends with raising this as an empirical question.

I see no double standard here, though I remain open to revising text to either clarify or soften certain points.

Muslims are terrorists but not female?
I was perplexed by your argument here. Google hits as evidence of what people are reading about? Sure. I get that people are reading more about Islamic terrorists than Islamic women. But Bian&Cimpian’s (& Leslie’s) claim was more people agree that Muslims are terrorists than Muslims are female. This might be true, but Google hits are not evidence of laypeople’s stereotypes. I do not see how, e.g., there being more private detective novels than cab driver novels could be taken to mean people believe there are more detectives than cabbies. There are more Google hits for “Republican candidates” than for “teenagers.” I doubt this means people believe there are more Republican candidates than teenagers.
Bitterly? Demands?
I did not see this piece as particularly bitter or outraged and was taken aback to discover that you saw it that way. Perhaps that is my own blind spot; if there are specific places or points that you see as over the top in some way, please do point them out to me. I am not sure how to see Jost & Kruglanski’s characterization of social constructionism as having “cultivated a leftist revolutionary spirit…” as anything other than a blatant call for leftist political influences on science. Is rejecting that call bitter? Is integrating Madison et al’s point about the “clever sillies” with Kahan’s point about signaling bitter? I do not see it, but I am open to softening sections that you see as harsh or shrill.

Last, after documenting a slew of scientific errors and speculating on some of their sources, it seemed to me constructive to end with recommendations on how to avoid them. I considered these to be suggestions, not demands. Today, after reviewing your emailed comments, I softened some of the language leading up to those suggestions in the hope of mitigating any demanding tone.

I think the paper is better with recommendations than without them. So this version has the shortened and softened version. However, I am not wedded to keeping them in. I will defer to your judgment regarding whether to:
1. Keep the recommendations as is
2. Soften them further
3. Delete them entirely. In this case, I would just keep the first paragraph under Conclusion, and end with the very last paragraph of the text, on pages 27-8.

If there are things you think I can do to soften the tone further, please let me know.

Sincerely,

Lee

---------------------------------------

Email of 2/7/16:

Dear Paul,

I am sending these to you directly, rather than through the BBS portal in case we have some negotiating to do. This seemed more efficient, especially since I am sending you two versions of the interim revised draft response, a clean one (the first attachment), and one with all track changes so you can see exactly what has been done (the second attachment).

I address each of your points, in turn here:

End on page 18. This was the most significant revision you called for. I only partially agreed, and am hoping we can meet about half way. Put differently, I cut this section in half, removing
all the stereotype threat stuff, the entire discussion around Table 2 (and Table 2 entirely), and lots of other tangential discussion. That was a tangential distraction, and I agreed with your call to remove those sections.

As a result, this version is 9 pages shorter.

However, what is left is, in my view, cogently focused on the commentaries, my response, and the main points of the book. The final conclusion seems to me to be just the right place to review the most important points that have remained intact, and the implications for the field, which is what those final pages do.

The weird "surprise" business. Removed. (left over from earlier drafts, I thought I had removed all of them, glad you caught them).

P. 4 "Two" --> "Two of the commentators". Done.

P.4, "Given the generally positive... commentators as representative sample"? Actually, I was trying to dissuade the reader from believing the commentators were at all representative. I mean, much of the point of the section on self-correction is how little self-correction there has actually been. Indeed, the original text, emphasized here, was:

Given the generally positive nature of most of the commentaries, one might suspect that the answer is “yes indeed.” One would, however, be wrong.

On the other hand, if you misread this, then so might others. I was not sure how to head this misinterpretation off, but I did remove the word "generally" from the sentence, and hope that will help avoid that misunderstanding.

P 4. "Response to Commentaries." Removed

p.5, “Agreement, but ..”, as a main heading.
p.6, “Seek and ..” as a main heading (I won’t keep mentioning this …)
I was not completely sure what you were suggesting here, but I did revise both headings, and hope they are now fine.

p.6, You need to remind the reader what political goals have to do with perceptually accuracy more generally. Done.


p.9, later: “Few in psychology …”. I’m not sure what you’re doing here — are you attacking Jost, citing him approvingly, what?
My intention here was to present this quote as a great example of social constructionist endorsement of politicizing research. I basically moved one sentence earlier in the paragraph, which, I think, makes the passage clearer.
P. 11-12, your doubts about the processistic fallacy. I disagree. The fundamental problem is that just because someone believes a process is not truth sensitive does not mean, in the real world, that it is not truth sensitive or, at least, not adequately mitigated by other truth sensitive processes. I will give you three examples, and then address your coffee example head on.

1. About 15 years ago, I was driving my kids home around 8pm. And it started to snow fairly heavily. I have three kids, two of whom are quite bright, and one who has neurological impairment and cognitive deficits. The snow got heavier as we were driving, and I asked if they thought they would have school tomorrow. The first said, "Yes, there really is not that much snow, it is not sticking, so there will be school." The second said, "It is early enough that even if it does stick, they will have plenty of time to clear it, so, yes, we will have school tomorrow." My cognitively impaired daughter just said, "No." I said, "Why?" She said, "Because I do not want to go to school tomorrow."

The first two had thoughtful, logically well-reasoned predictions. My third one did not. Guess who was right?

2. Ty Cobb. Did you ever play baseball or softball? Everyone knows the "process" for hitting the ball includes holding your hands together on the bat. Everyone knows that holding your hands apart is completely dysfunctional, except when bunting.

And, for most of us, it is. But it wasn't for Ty Cobb. He held the bad with his hands about 2-3 inches apart. If you go to Google Images, and search for "Ty Cobb holding the bat" you can easily see this for yourself.

If one judged Ty by his process of hitting, one would have been compelled to conclude that he was a terrible hitter. I do not know how well you know baseball, but he was one of the greatest hitters of all time, probably second only to Babe Ruth.

3. The conjunction fallacy. It is definitely a fallacy. It is completely impossible for a conjunction to be more likely than any of its components. The classic Kahneman & Tversky example is believing in the probability of two particular disasters.

Consider Barbara. She estimates that, historically there has been a 2% chance of an earthquake triggering a flood in California that kills 1000 people each year, and that there has been a 1% chance of a natural disaster occurring in the U.S. that kills 1000 people each year.

Barbara has definitely committed the conjunction fallacy. She is wrong for believing that a 1000 deaths by Cali/earthquake/flood is more likely than 1000 deaths anywhere from any natural cause.

Now consider Fred. Fred does not commit the conjunction fallacy. He estimates that, historically there has been a 95% chance of an earthquake triggering a flood in California that kills 1000 people each year, and that there has been a 100% chance of a natural disaster occurring in the U.S. that kills 1000 people each year.
Barb commits the fallacy, Fred does not. Whose judgments are more accurate?

4. Your coffee example. Coffee intake can influence accuracy. It can increase energy, alertness, and focus. In some ways, it is an intelligence-booster, at least temporarily. Could it influence the content of some stereotype? Of course it could. Perhaps it primes "warmth" (ala Bargh), thereby affecting judgments of warmth. Whether the subsequent judgments are correct or not hinges on many things, of which how much warmth one sees is only one. Maybe the perceiver is high on neuroticism/dysphoria, and sees people as nastier than they are most of the time. Thus, by increasing how warm this perceiver views others, coffee increases the accuracy of his judgments.

I do not know if it is impossible to come up with some process that must lead to inaccurate judgments, but it is at least close. And, most important, even if it was true in some narrow context that does not mean people's judgments are ultimately generally inaccurate. Let's say you do identify some process that must cause inaccuracy. Fine. Does that mean the ultimate judgments are inaccurate? Not at all. In the real world, there are almost always multiple influences on judgments, many of which might enhance accuracy, so that even if your process undermined accuracy, the ultimate judgment would still be mostly accurate. On top of that, people are often responsive to corrective feedback, so that, even if their judgments start out wrong, they may be revised to be more correct. Thus, the proposed bad process, even if undeniably bad, may not ultimately have much influence on real world judgments.

I am sticking to my guns here, Paul. Psychologists have a long and somewhat sordid history of leaping to what turns out to be unjustified and invalid conclusions that people are inaccurate on the basis of processes that supposedly cause inaccuracy but which, in fact, do not.

p.12: Your discussion of B&C is at times ungenerous... You write:
In short, this is no evidence whatsoever that bears on the claim that “more people believe Muslims are terrorists than Muslims are female.”
But there is a lot of evidence that more people will assent to “Muslims are terrorists” and not to “Muslims are female”. (Do a google search).
I reviewed the section and removed some text that seemed to border on "snarl-ish." That should help the tone. On the merits, though, I took your advice, did several google searches, in both regular Google, and Google Scholar, and found almost nothing. There is lots on generic beliefs, but, except for their piece and the Leslie article, I found neither research nor even news stories about more people believing Muslims are terrorists than Muslims are female. I realize there may have been some deficiency in my search. If you ultimately think this is important, I am going to need you to send me some specific articles or, at least, references.

However, my point was not that "there is no evidence" about this. The science world is filled with lots of evidence on lots of topics about which I will readily acknowledge my own ignorance. My point was that: 1. B&C only cited Leslie in support of this; and 2. Leslie neither presented nor reviewed evidence for it. That seemed cogent.

So, the original sentence read:
In short, what this presents is no evidence whatsoever that bears on the claim that “more people
believe Muslims are terrorists than Muslims are female.”

It now reads:
In short, Leslie (in press) presents no evidence whatsoever that bears on the claim that “more people believe Muslims are terrorists than Muslims are female.”

Furthermore, the immediately following text takes seriously that the claim may have some validity, regardless of Leslie’s lack of evidence for it. I contested the conclusion that such evidence would mean the stereotype is necessarily inaccurate, but I did not dismiss the possibility that people might actually hold such beliefs, despite the lack of evidence for it in the commentary. See the section that starts ” Of course, even if this were valid, how it would bear on stereotype accuracy is unclear...” and which ends with the following, "Whether any particular generic beliefs is, in fact, absolutist requires evidence. In the absence of such evidence, researchers are welcome to present their predictions as hypotheses or speculations about stereotypes' supposed absolute or inaccurate content, but they should not be presenting their own presumptions as facts."

For the record, before I submitted this to BBS, I sent this section on the processistic fallacy to both Cimpian and Terbeck to be sure that they agreed that, at least, I summarized their critiques correctly, and both signed off on that (I am not saying they signed off on my pushback). Tangentially, I liked both of them, I think they both felt similarly, and have ongoing correspondences with them now.

I think that is everything. If I missed anything, I know you will let me know. If you have strong disagreements with any of this, I am open to discussing them. And, of course, if you want me to send these directly through the BBS portal before you make any such decisions, please do let me know.

Best,

Lee
Accuracy, Bias, Self-Fulfilling Prophecies and Scientific Self-Correction

Lee Jussim
Rutgers University
Abstract

In my précis of Social Perception and Social Reality (SPSR), I argued that the social science scholarship on social perception and interpersonal expectancies was characterized by a tripartite pattern: 1. Errors, biases, and self-fulfilling prophecies in person perception were generally weak, fragile, and fleeting; 2. Social perceptions were often quite accurate; and 3. Conclusions appearing throughout the social psychology scientific literature routinely overstated the power and pervasiveness of expectancy effects, and ignored evidence of accuracy. Most commentators concurring with the validity of these conclusions. Two, however, strongly disagreed with the conclusion that the evidence consistently has shown that stereotypes are moderately to highly accurate. Several others, while agreeing with most of the specifics, also suggested that those arguments did not necessarily apply to contexts outside of those covered in SPSR. In this response, I consider both the limitations to the tripartite pattern, the role of politics and confirmation biases in distorting scientific conclusions, common obstructions to effective scientific self-correction, and how to limit them.
When I began writing *Social Perception and Social Reality* (hence abbreviated as *SPSR*), my goal was to offer a corrective to a slew of manifestly false claims in common conclusions about social perception. The *Precis* summarized this as the *tripartite pattern*: 1. Errors, biases, and self-fulfilling prophecies in person perception were occur, but are generally weak, fragile, and fleeting; 2. Social perceptions were often quite accurate; and 3. Conclusions appearing throughout the psychological literature (with educational psychology as a notable exception) were often unhinged from these data by virtue of routinely declaring expectancy effects powerful and pervasive, and consistently ignoring evidence of accuracy. I also argued that defining stereotypes as inaccurate is logically incoherent, and that, in sharp contrast to 100 years of claims to the contrary, the evidence is that stereotypes are often (not always) quite accurate. All of this was “news,” not because of new dramatic data, but because the old data never justified the strong claims about the power of expectancies to create social reality. To paraphrase Winston Churchill’s characterization of the early architects of appeasement, social psychology periodically stumbled on the truth but simply picked itself up and hurried along as if nothing had happened.

The evidence supporting an emphasis on error, bias, and irrationality in social perception was maintained by virtue of overstatements of what research actually found (e.g., by ignoring effect sizes), selective citation of a small number of dramatic “wow effect” studies (Jussim, Crawford, Anglin, Stevens, & Duarte, in press), many of which have proven difficult to replicate. The error and bias (“stupidism” as per Kihlstrom) perspective was also maintained by reliance on a “toolbox” of double standards, blind spots, and researcher confirmation biases that served to elevate the supposed importance of “bias” results and studies while denigrating or dismissing evidence of accuracy. Indeed, *SPSR* pointed out that even some of the most classic demonstrations of “bias” in social perception (e.g., Hastorf & Cantril, 1955; C. E. Cohen, 1981; Rosenhan, 1973) actually provided more evidence of accuracy than of bias – as did many of the follow-ups to classic studies such as Rosenthal & Jacobson (1968), Snyder & Swann, 1978, and Darley & Gross (1983).

To my great surprise, nearly all of the commentators accepted the fundamental conclusions of *SPSR*. Of the seventeen commentaries, only two took issue with the major conclusions (Bian & Cimpian; Terbeck); and two disagreements occurred, not over conclusions, but over definitions (Andrews, Springer). Two others
similarly did not take issue with the tripartite conclusion, but also suggested the phenomena outside the scope of SPSR might have yielded different conclusions (Kahan, Wilson).

There was abundant embracing of the main conclusions. Trafimow’s consideration of statistical effect sizes concludes that the weight of the evidence supports the tripartite conclusions. Two of the commentators (Marczyk, Little) pointed out that it would be bizarre if evolution did not lead us to be in touch with reality much of the time. Several commentators pointed to evidence outside the scope of SPSR that also often yields evidence of accuracy and rationality (Kihlstrom, Mousavi & Funder, Bonnefon et al, Wagoner). And one commentary (Madison et al) endorsed and expanded on one of the undercurrents of SPSR – that politics and dogmas have probably distorted the scientific conclusions in this and many other areas of social science.

Does this mean psychology has finally turned a corner and is self-correcting towards fully recognizing that social perception is often largely rational and accurate and generally only modestly subject to error and bias? Given the generally positive nature of most of the commentaries, one might suspect that the answer is “yes indeed.” One would, however, be wrong. Before returning to that question, however, I respond to the specific issues raised in the commentaries.

Response To Commentaries

Definitional Differences

Sensory perception vs. social perception. Church correctly pointed out that the research reviewed in SPSR does not make a hard distinction between sensory perception and more global cognitive representations, such as interpretations, beliefs, judgments, etc. In fact, to social psychologists, social perception rarely deals with sensory perception per se, and generally deals with “perception” in this more cognitive and molar sense. When we seek answers to questions like, “How do voters perceive President Obama?” we are almost always interested in beliefs, attitudes, and opinions about Obama, not their sensory perceptions. I am glad to have the opportunity to clarify that SPSR’s focus is on molar social perceptions, and not sensory perceptions. However, the idea that beliefs and expectations influence sensory perceptions has even less support than it does with respect to molar social perceptions (Firestone & Scholl, in press).

If every concept is a really a “stereotype” the term loses all meaning. Andrews’ commentary
advocates abandoning the distinction between stereotypes and individuating information. I disagree. If anything one person believes about another’s behaviors and characteristics is a “stereotype” then the term loses all meaning and usefulness as a theoretical construct.

Social psychologists (and I suspect many other people—see, e.g., the 14th Amendment to the U.S. Constitution, and the Civil Rights acts of the 1960s) usually consider it important to understand where beliefs (at least, if those beliefs influence behavior) about, e.g., Alfonso, originate. Do they result from: 1. his being a male or Latino or 2. The fact that he acts in his local theater group, earned a law degree from Harvard, likes skinny jeans, and vacations in the Bahamas? It does not matter much, I suspect, whether we call them stereotypes vs. individuating information, or “concepts” about groups and concepts about individuals. What is important is that beliefs about Latinos are not the same thing as beliefs about Alfonso. They may have some superordinate similarities (both are beliefs), in the same way that beliefs about apples (they are usually red and good to eat) have some superficial similarity to this apple (which is brown and rotting and not good to eat). Nonetheless, both apples and this apple are concepts and, in that way, have some similarity. The distinction is also important, however, because it highlights important differences between apples and this apple, and between Alfonso in particular, and Latinos in general.

Agreement, But Filling in Gaps and Expanding the Scope

**Evolutionary explanations for (in)accuracy.** Two commentators brought up the important theoretical point that evolutionary theories could explain both why there is relatively high levels of accuracy in social perception and also conditions likely to produce low accuracy because adaptations to advance goals other than accuracy might take precedence (Little, Marczyk). These commentators: 1) Are correct in pointing out that *SPSR* did not draw on evolutionary perspectives; 2) point out that doing so is likely to provide important theoretical advances to understanding conditions under which accuracy is likely to be higher or lower. At its most basic level, it is hard to imagine any successful organism that has evolved to have completely invalid reactions to its environment.

On the other hand, evolution emphasizes adaptations that enhance an organism’s ability to produce viable offspring. Thus, as both Marczyk and Little correctly point out, there are many goals that might accomplish
that have little relevance to, or which might even conflict with, accuracy (e.g., identifying and attracting fertile mates, attracting resources to support offspring, etc.). Deception is prominent in the animal world (and even occurs in the plant world; consider carnivorous plants posing as nectar-rich flowers) because there are so many ways that it could have adaptive advantages. Thus, it also seems implausible that evolution would yield social perceptions that were perfectly accurate (von Hippel & Trivers, 2011). In psychology, error management theory (Haselton & Buss, 2000) was, as pointed out by Little, an early and constructive attempt to identify when evolutionary pressures were more likely to lead to accuracy versus certain specific types or patterns of errors. Indeed social psychological theorizing will likely be enhanced and sharpened by further efforts to exploit evolutionary ideas to understand when people are likely to be accurate and when they are likely to be systematically inaccurate.

**Seek and (often) ye shall find: Accuracy in other contexts.** Several commentators pointed out that, in other contexts, there is often: 1. Surprising evidence of accuracy; and 2. Similar patterns of political or theoretical double standards in evaluations of the evidence. Such double standards occur when people hold research that advances their theoretical perspectives or political values to lower standards than they hold research that opposes their theoretical perspectives or political values. For example, research demonstrating biases produced by stereotypes or prejudice to be “powerful and important” (see many examples in SPSR), while, at the same time, downplaying, ignoring, or dismissing the typically far more powerful effects of individuating information.

Bonnefon, Hopfensitz, and De Neys, for example, point out these issues in the study of the accuracy of judgments of trust at zero acquaintance. Although the levels of accuracy are much lower than among expectancies, they point out (correctly, in my view) that any accuracy on the basis of a mere photograph is quite striking. Furthermore, their analysis strongly suggests that political/advocacy goals have led to a set of logically contradictory conclusions about accuracy in perceptions of trust, in a way quite reminiscent of the double standards and logical incoherence I identified with respect to self-fulfilling prophecies and stereotypes. In short, perceptions of trustworthiness have been declared both inaccurate and self-fulfilling, and these are mutually exclusive conclusions. A belief that a target is untrustworthy can be, at one moment in time,
inaccurate, and the next, self-fulfilling such that the target becomes untrustworthy. After the target becomes untrustworthy, however, subsequent perceivers are not wrong for believing the target to be untrustworthy. They are accurate. Consequently, Bonnefon et al correctly point out that perceivers’ beliefs about trustworthiness cannot be generally inaccurate and self-fulfilling.

Mousavi and Funder similarly point out that judgments are often “ecologically rational,” meaning that they are adapted well to their environments. Fast and frugal heuristics, though technically constituting “biases,” especially in laboratory studies, often lead to moderately to highly accurate judgments in much of the rest of daily life. They, too, echo the political implications of accuracy, pointing out that an overweening emphasis on error and bias misses a great deal of evidence of accuracy. This is important, they argue (and I agree), in part, because in such situations, efforts to solve social problems by changing supposedly erroneous beliefs are doomed to failure because the beliefs are not particularly erroneous in the first place.

Wagoner pointed out that distortions of the scientific record similar to those described in Social Perception and Social Reality have long characterized perspectives on memory. The schema concept, which is hypothetically at least neutral with respect to accuracy (as are interpersonal expectancies) has become virtually synonymous with error and bias (as have interpersonal expectancies). That many modern perspectives have just as blithely ignored Bartlett’s (1932) balanced view of accuracy/error in memory as F. H. Allport’s (1955) balanced views on perception is a testament to the long reach of the distorting power of theoretical perspectives emphasizing distortion.

**Constructivism (Both Cognitive and Social) Lives!**

**Cognitive constructivism.** Kihlstrom agrees with the general thrust of SPSR but also urges not to throw out the baby (cognitive constructivism) with the bathwater (the excessive emphasis on error and bias). And nor did I intend to do so. Kihlstrom’s commentary presents a very thoughtful and balanced view of realism and constructivism, and is a great primer on how social psychology can be enriched by not dismissing ideas from any of those broad perspectives writ large.

The cognitive constructionist processes highlighted by Kihlstrom undoubtedly can and do influence memory and the types of social perceptual processes addressed in SPSR. And, surely, sometimes those effects
constitute do indeed errors and biases. However, constructivism and error/bias are not synonymous.

Although Kihlstrom did not argue that they were, because cognitive constructivist is sometimes presumed to mean something like “perceivers making stuff up that supports their pre-existing beliefs, expectations, and values” (Kihlstrom’s own emphasis on the over-reach of “stupidism” perspectives; see also commentaries by Mousavi & Funder; Madison, Dutton, and Stern; and Wagoner), it is, perhaps, worth walking through why bias and constructivism are not synonymous.

“Constructive accuracy” refers to the process by which expectancy-induced “biases” can increase accuracy (Jussim, 1991). Figure 1A presents the Reflection-Construction Model (Jussim, 1991), which depicts relations among the key variables involved in accuracy, bias, and self-fulfilling prophecy. Figure 1B depicts constructive accuracy. It shows that impression accuracy (correspondence between perceivers’ judgments of targets and those targets’ behavior or attributes) can be quite high, even when perceivers base their judgments of individual targets exclusively on their own expectations, and are oblivious to (ignore, overlook, do not have access to) targets’ actual behavior or attributes. If all three paths shown are high enough, perceiver judgments will correspond to (correlate with) target behavior or attributes, even though perceiver judgments are heavily based on their own expectations and not at all based on target behaviors or attributes. This is because, in Figure 1B, the correlation between perceiver judgments and target behavior or attributes is the multiplicative product of the three paths. For example, if all three equal .8, then impression accuracy equals .8^3 = .51. In psychological, rather than mathematical terms, this means that, if perceivers’ expectations are strongly based on highly valid information, the more they rely on those expectations when judging targets, the more accurate they will be.

This is a constructive phenomenon, because, in this example, the judgment is based entirely on perceiver expectations, with no direct input from targets’ actual behavior or attributes. Furthermore, even if perceivers do partially base their judgments directly on targets’ behavior or attributes, relying on accurate expectations can still increase accuracy further (see Jussim, 1991, for details). Thus, I concur with Kihlstrom that constructive processes can and do play an important role in person perception; however, I would emphasize that, even so, such processes may, at least when those expectations are themselves based on valid information,
increase rather than reduce accuracy.

Social constructionism. Two of the commentaries (Tappin, McKay, & Abrams; Wilson & Huang) make measured appeals to not completely throw out the social constructionist baby with (what I would call) the excessively politically tainted bathwater. Neither explicitly uses the term, social constructionism or its kindred (post-modern, critical theory, etc.), very much (Tappin, McKay, & Abrams, not at all) to develop their own perspectives, perhaps because in many scientific circles, the efforts never made serious inroads, in large part, because they have been largely dismissed as unscientific. However, both commentaries, in somewhat different ways, present defenses of social constructionist processes. Tappin et al do so by arguing for the importance of collective action as a major influence on social reality; Wilson & Huang do so by emphasizing the role of institutions in creating social reality. I see these arguments as mutually reinforcing so I address them both here.

This reply is not the place for a comprehensive review or critique of social constructionism, which is a single term that refers to quite a variety of perspectives. I would, however, divide social constructionism into two main veins (which are not necessarily mutually exclusive). One is primarily a political liberation perspective, with the goals of combating oppressive and hegemonic practices and discourses, in part, by revealing them, thereby advancing the interests of certain groups that the users of such terms deem unfairly victimized or exploited in some way. Few in psychology capture the politicized nature of the efforts better than Jost & Kruglanski (2003. P. 175) who approvingly declared: “From this perspective, we have a professional obligation to weigh in on ideological issues, policies, and decisions” and later on the same page, “The social constructionist movement emerged on the social science scene as a force for change and cultivated a leftist revolutionary spirit that posed a distinctive challenge to established scientific authority.” As a movement with primarily political goals, my view is that this sort of politicization has little place in scientific psychological theorizing. Few in psychology capture the politicized nature of the efforts better than Jost & Kruglanski (2003. P. 175) who approvingly declared: “From this perspective, we have a professional obligation to weigh in on ideological issues, policies, and decisions” and later on the same page, “The social constructionist movement
emerged on the social science scene as a force for change and cultivated a leftist revolutionary spirit that posed social constructionism are blatantly political or liberationist. Some, instead, seeks to understand social relationships, including but not restricted to relationships of power and status, and the reciprocal influences among individuals and institutions (regardless of whose interests or advocacy agendas these understandings might advance). In this spirit, Tappin et al are surely correct in arguing that, under some conditions, collective actions can alter the nature of intergroup relations.

Nonetheless, it is also, perhaps, worth pointing out that little, if any, of the research they cited links laypeople’s actual stereotypes to collective action. They review abundant evidence that target groups are motivated by perceived slights to and injustices inflicted on their groups; but none of it links laypeople’s actual stereotypes to collective action. I suspect that this failure stems from: 1. Social psychologists taking for granted that laypeople hold unjustified and pernicious stereotypes; so it was 2. not even necessary to assess actual stereotypes, the derogatory nature of stereotypes could simply be taken for granted.

Space does not permit a citation by citation critical analysis of the work they presented in support of their perspective, so perhaps one example will suffice. They cite a series of studies by Ellemers and Barreto (2009) three times in their short commentary, so they seem to consider it important to their perspective. Ellemers and Barreto (2009) showed that believing others had an insult view of one’s group (e.g., for women, someone believing that women are unintelligent), motivated collective action. But what if lay people do not routinely believe women are unintelligent? There is evidence that people believe boys are better at math than girls, but the same studies show that people also believed girls have higher verbal skills than boys (Swim, 1994). Regardless, Ellemers and Barreto (2009) did not assess anyone’s sex stereotypes regarding intelligence, nor do they review research that has done so. Their findings are still interesting, because they say something about how perceived intergroup insults motivate collective action, but not because it says anything about the role of the actual stereotypes held by any actual people in leading to collective action. This is not meant to dismiss the perspective entirely. In fairness, it probably can be interpreted as showing that when people hold derogatory stereotypes, if targets become aware of those stereotypes, collective action may result. Of course, history is filled with counter-examples, cases where people did hold derogatory stereotypes and collective
action did not result over vast periods of time (consider, e.g., the inferior status ascribed women, three hundred years of slavery in the United States, and the Hindu caste system). I suspect, therefore, that the “holding insulting stereotypes – collective action” link is tenuous at best, and subject to many conditions not articulated in either Tappin et al’s commentary or much of the underlying research. Indeed, I do not doubt the effect exists, but I would suspect that, in the real world, across many situations and contexts, it fits the pattern described in SPSR: occasionally strong, but usually quite weak, fragile, and fleeting.

Wilson & Huang are correct in pointing out that my conceptual analysis “freezes” institutions at a given point in time, and then examines the rationality vs. biased nature of social perception. As they point out, institutions are not actually frozen, and are subject to both slow-moving and, occasionally, dramatic and sudden changes. My argument was never intended to be “there are no conditions under which stereotypes or social beliefs construct reality.” Indeed, SPSR is peppered with both real world examples and scientific studies showing that, sometimes, such effects can be quite powerful.

SPSR, however, did not have as its purpose identifying the nature of collective movements or the inter-relationships between institutions, demographic groups, and individuals. Instead, the purpose of the book was to review evidence regarding the extent to which individuals’ beliefs about groups or other individuals were accurate, biased, or self-fulfilling. This is, as these commentaries suggest, a limitation of its scope. It is certainly an important and appropriate social science endeavor to address issues such as collective action and institutions. SPSR made no claims about such issues. It was, however, an attempted corrective to longstanding and unjustified social science claims about how individuals’ beliefs relate to social reality – and, on this issue, it is, perhaps, worth noting that both Tappin et al and Wilson & Huang presented neither argument nor evidence against the central claim that such corrective is justified and past due.

Victims of the Processistic Fallacy

Two commentaries aspire to refute the conclusion reached in SPSR that stereotypes have been widely found to be at least moderately accurate. Both Terbeck and Bian & Cimpian propose processes that they believe cause inaccuracy in stereotypes. Both critiques fall victim to the processistic fallacy, which was addressed in SPSR. Thus, my response to these critiques begins by quoting that text (p. 394):
To address accuracy, research must somehow assess how well people’s stereotypes (or the researchers believe to be flawed. This is a fallacy for several reasons: (1) The process may not be as flawed as the researchers believe, and its degree of “flaw” cannot be assessed without assessing the validity or success of the judgments and decisions by people who do versus do not rely on this process (something social scientists rarely do); (2) even if the process is indeed flawed, in real life, people may rely on many other less flawed processes when making judgments and decisions; and (3) in real life, social reality often intrudes upon people’s erroneous beliefs — that is, it provides feedback that permits people to recognize their initial beliefs were wrong and to alter them accordingly. So, again, we cannot know how flawed the outcome is — the judgment or decision — unless we evaluate its success, accuracy, validity, etc. (which is another thing social scientists emphasizing error and bias do not often do).

The processistic fallacy is a form of overgeneralization. It occurs whenever researchers who demonstrate some error or bias under a very small set of (typically artificial laboratory) conditions unjustifiably assume or conclude that their findings mean that there is widespread human error under naturally occurring conditions (e.g., Cohen, 1981; Funder, 1987). Lab studies are often well-designed to test basic processes but not to generalize results to naturally occurring conditions (Mook, 1983). It is hypothetically possible to appropriately generalize widespread error under naturalistic conditions on the basis of studies revealing flawed processes in the laboratory, but only under conditions that are almost never met. One way to do justify such generalizations is to discover a process so flawed that it must be definitively known to produce pervasive inaccuracy in situations that go well beyond those studied in the lab. For example, the human visual system cannot detect radio waves, so that it is safe to conclude people will be universally inaccurate in their visual assessment of the presence/absence of such waves.

Such demonstrations are few and far between in psychology. A wide range of judgmental and perceptual errors and biases found in laboratory studies have turned out to be functional outside those studies. For example, Gigerenzer & Brighton (2009, p. 107) reviewed evidence showing that: “In contrast to the widely held view that less processing reduces accuracy, the study of heuristics shows that less information, computation, and time can in fact improve accuracy. We review the major progress made so far: (a) the discovery of less-is-more effects; (b) the study of the ecological rationality of heuristics, which examines in which environments a given strategy succeeds or fails, and why…” Such findings should give deep pause to modern researchers who, upon discovering some laboratory bias, leap to the assumption that, therefore, that process undermines accuracy in naturalistic conditions. Regardless, I am not aware of any research that has
documented a social perceptual process so flawed that it can be definitively known to produce inaccuracy on purely logical grounds comparable to the radio wave example above.

However, even if such a process were discovered, additional conditions must also be met to generalize from lab studies of biased processes to a conclusion of widespread inaccuracy in life. It must be shown either that people are incapable of overcoming the bias or error by relying on alternative, superior or corrective processes, or, empirically, that, across most of a widely representative array of situations, people both rely on the flawed process and rarely enlist superior or corrective processes. If a program of research engages in a sufficiently large representative sampling of situations (Brunswick, 1957; Monin & Oppenheimer, 2014; Westfall, Judd, & Kenny, 2015), and shows that, in most such situations, a flawed process is heavily relied upon and other more appropriate processes are rarely enlisted, inferring widespread error in real life can be justified. Few programs of research, however, meet these standards, whether in social perception or other areas of psychology (e.g., Cohen, 1981; Westfall et al., 2015). This is probably because, as Wells & Windschitl (1999, p. 1115) found, among psychology faculty, there was an “... insensitivity to the need for stimulus sampling except when the problem is made rather obvious.”

Common flaws in the critiques. Both commentaries identified potentially flawed processes, and both perspectives are capable of generating testable (and falsifiable) hypotheses about potential patterns and sources of stereotype inaccuracy. As such, their perspectives are potentially constructive and generative of new research directions and potentially valuable insights into sources and conditions of stereotype inaccuracy.

Nonetheless, neither Terbeck, nor Bian and Cimpian discussed any research that meet the standards articulated above for concluding that stereotypes must be inaccurate on the basis of the supposedly flawed processes that were identified. Neither presented justification for assuming that those supposedly flawed processes inherently produce inaccuracy in other judgments (comparable to showing that vision cannot detect radio waves). Thus, we do not know that those processes produce inaccuracy.

Neither discussed any program of research that has shown that those processes produce inaccuracy most of the time in a representative sampling of situations. Therefore, it is not knowable from that research whether people commonly rely on those processes, even if they are truly flawed. Last, even if those processes are truly
flawed and widely relied upon, neither commentary reviewed any research demonstrating that people rely exclusively on supposedly flawed process across situations (even unrepresentative ones). They have not eliminated the possibility that there are other, more appropriate processes that people rely, when arriving at stereotypes. Both critiques therefore, declare stereotypes to be inaccurate on the basis of research incapable of justifying such a conclusion. Both committed the processistic fallacy: over-inferring pervasive (“stereotypes must be inaccurate”) error in real life from laboratory studies of processes that were incapable of generalizing to much of real life. It is of course possible that such studies do generalize widely; but that cannot be known without empirical demonstrations that they actually do generalize widely. To make these issues more concrete, the specific evidence each discussed is reviewed next.

Terbeck, categorization, implicit prejudice, and the brain. Terbeck referred to research showing that:
1. infants and primates categorize, 2. Specific brain areas are associated with face recognition; and 3. Drugs alter scores on the race implicit association test (IAT). This is all fine as far as it goes. Categorization is ubiquitous, thus, this passes the test for a justified generalization to real life. However, categorization is not inherently universally invalid, in the same way that visual detection of radio waves is. People are not wrong for believing that chairs usually have four legs, that Alaska is colder than Arizona, or that men are, on average, taller than women. Thus, the claim that any particular category is wrong requires evidence, which Terbeck did not provide.

Similarly, specific brain areas may well be associated with face recognition, but the very term “recognition” implies that, at least some and perhaps most of the time, people are correctly distinguishing faces from other features of the stimulus array. It certainly provides no evidence that facial recognition is wrong. Finally, I have no doubt that drugs can alter IAT scores. Racial prejudice IAT scores are attitudes, and individuals and societies may deem certain attitudes morally good or bad, but attitudes cannot be factually correct or incorrect. It is possible that one’s reasons for disliking diet soda, the Yankees, and Fred are factually incorrect, but the attitude itself cannot be accurate or inaccurate. Thus, all three phenomena identified by Terbeck may lead to falsifiable hypotheses about sources of stereotype inaccuracy; but, absent direct data on stereotype accuracy, they do not justify concluding that stereotypes are inaccurate.
This succinctly captures the central flaws in both commentary’s attempt to resuscitate the conclusion that necessary to infer widespread naturally occurring error from studies of supposedly flawed processes. Prototypical cases of supposedly inherently erroneous generic beliefs are those such as “mosquitos carry the West Nile virus” and “ducks lay eggs” (which was the example highlighted in the title of one of the articles they cite in support of their view: Leslie, Khemlani, & Glucksberg, 2011, and which I use here). They cite evidence that people judge such statements to be true. They argue that this renders people inaccurate because few mosquitos carry West Nile virus and not all ducks lay eggs.

Does agreeing that “mosquitos carry West Nile” mean that we can now assume that people’s beliefs about mosquitos and West Nile are pervasively inaccurate? If these are absolutist beliefs (“all mosquitos carry West Nile”) then they are clearly wrong and no further evidence is needed. SPSR made exactly this point when discussion absolutist stereotypes, which, because of widespread human variation, are almost always invalid. But there is no evidence that generic beliefs are always, necessarily, or widely absolutist.

Perhaps, instead, they capture the phenomenology of distinctive or salient differences between categories. I can only get West Nile from mosquitos, not from moths, mice, or musk ox. Perhaps people agree that “mosquitos carry West Nile” not because they believe “all mosquitos carry West Nile,” but because they believe that “only mosquitos carry West Nile.” Because generic beliefs, as studied, are not inherently inaccurate, the research does not meet the first standard necessary to avoid the processistic fallacy. We cannot assume all generic beliefs are necessarily inaccurate.

It also fails the second standard (even if not inherently inaccurate, is the process empirically found to be generally invalid?). One of the articles cited by Bia and Cimpian (Leslie et al, 2011) found that participants rephrased only 18 of 100 experimenter-provided generic statements as absolutist (“universals” in Leslie et al’s, 2011, terminology). Furthermore, overwhelming majorities (over 90%) most recognized that, in fact, male sheep do not produce milk, male snakes and male ducks do not lay eggs, and so on for nearly all absolute beliefs studied. Thus, one cannot interpret agreement with the generic beliefs as evidence of widespread reliance on an invalid process. The Leslie et al (2011) research did include a wide range of generic beliefs, so it is reasonable to conclude that their results are broadly generalizable to generic beliefs. What is
generalizable, however, is that most generic beliefs do not equate to absolutist or inherently inaccurate beliefs. Of course, it is still possible that when stereotypes are generic beliefs, they are widely inaccurate. That is another falsifiable hypothesis about which there is currently no data. Again, inferring that stereotypes are inaccurate from such data is unjustified.

Even if people’s beliefs about ducks’ egg laying were generic and wrong, we would still have no direct information about the accuracy of their beliefs about other people. So, how does this translate to stereotypes? Bian and Cimpian cite another paper by Leslie (in press) in support of the claim that “more people hold the generic belief that Muslims are terrorists than hold the generic belief that Muslims are female.” What was Leslie’s (in press) “evidence”? But Leslie (in press) provides no data whatsoever that bears on the frequency with which people hold such beliefs. Instead, she quotes from headline-seeking politicians and cited a rise in hate crimes post-9/11. In short, this is no evidence whatsoever that bears on the claim that “more people believe Muslims are terrorists than Muslims are female.” Such information may be interesting, but it does not address the frequency of lay beliefs about anything whatsoever.

Of course, even if the claim that more people agree that “Muslims are terrorists” than that “Muslims are women” is valid, it would not constitute evidence stereotypes in general, or the Muslim stereotype in particular, must be inaccurate. How it would bear on stereotype accuracy. Its status as such evidence does not hinge is entirely unclear, because that would depend, not on researcher assumptions about what people mean when they agree with statements like, “Muslims are terrorists” but on evidence assessing what people actually mean. Because research on generics fails the first two tests necessary to avoid the processistic fallacy (they do not inherently produce inaccuracy and they have not been empirically demonstrated to usually produce inaccuracies). one could not conclude that greater agreement that “Muslims are terrorists” than that “Muslims are women” necessarily means people believe there are more Muslim terrorists than Muslim women. It may simply mean “some Muslims are terrorists” or “Muslim terrorism is more widespread than other forms of terrorism” and that “being female is not an important distinguishing characteristic of Muslims.” Absent data, we just do not know. The bias literature writ large (L. J. Cohen, 1981; Gigerenzer & Brighton, 2009; see also Mousavi & Funder’s commentary and SPSR and the stereotyping literature in particular) is so strongly riddled
with invalid researcher *presumptions* about lay people’s beliefs, that, absent hard empirical evidence about what people actually believe, researcher assumptions of inaccuracy that are not backed up by empirical evidence demonstrating widespread inaccuracy rarely do act warrant credibility.

If, as seems to be widely assumed in discussions such as Bian and Cimpian’s, agreeing that “Muslims are terrorists” means “all Muslims are terrorists” then such stereotypes are clearly inaccurate (indeed, *SPSR* specifically points out that all or nearly all absolute stereotypes of the form ALL of THEM are X are inherently inaccurate, because human variability is typically sufficient to invalidate almost any such absolutist claim that most stereotypes are not statistical beliefs, or, at least, generically based stereotypes are more potent influences on social perceptions. They present no assessment, however, of the relative frequencies with which people’s beliefs about groups are generic versus statistical, and, given Leslie et al.’s (2011) evidence that people do not usually translate generics into absolutes, it may well be that agreement with generics such as “ducks lay eggs” and “Muslims are terrorists” does not preclude the statistical understanding that fewer than half of all ducks are even capable of laying eggs or that the proportion of Muslims who are terrorists is tiny. Again, there is an assumption without evidence.

We already know that the predictions generated from the “most stereotypes are generic and are therefore statistically clueless” are disconfirmed by the data summarized in *SPSR* (see also Jussim, Crawford, & Rubinstein, 2015, for an updated review of stereotype accuracy that includes additional studies). Bian and Cimpian have developed compelling descriptions of the processes that they believe *should* lead people to be inaccurate. In point of empirical fact, however, people have mostly been found to be relatively accurate. Disconfirmation of such predictions can occur for any of several reasons: 1. The processes identified as “causing” inaccuracy do not occur with the frequency that those offering them assume (maybe most stereotypes are not generic); 2. The processes are quite common and do cause inaccuracy, but are mitigated by other countervailing processes that increase accuracy (e.g., adjusting beliefs in response to corrective information); or 3. The processes are common, but, in real life, lead to much higher levels of accuracy than those emphasizing inaccuracy presume (see Mousavi & Funder’s commentary for exactly such a point).

Regardless, making declarations about levels of stereotype inaccuracy on the basis of a speculative prediction
that some process causes stereotype inaccuracy, rather than on the basis of evidence that directly bears on accuracy, is a classic demonstration of the processistic fallacy.

**Questionable Interpretive Practices and the Politics of Social Psychological Research**

Kahan did not disagree with a single claim in *SPSR*; he did, however, urge me to consider the issues of bias and accuracy more broadly, and I do so here. Kahan correctly pointed out that there is an extensive literature on confirmation bias especially in politicized judgments that *SPSR* largely ignores. My goal was to evaluate the literature on social perception – how people view other people, especially individuals and groups; and especially with respect to judgments that could conceivably assessed for their accuracy. To compare bias, self-fulfilling prophecy, and accuracy, it was necessary to focus on judgments could be biased, self-fulfilling or accurate. *SPSR* purposely excluded people’s beliefs about scientific or social science facts or evidence because I do not consider them social perception in the classic sense of “how people understand specific other people or groups.” *SPSR* also excluded moral and political beliefs because they often have no criteria for assessing accuracy. I concur with Kahan’s view that confirmation biases can be quite powerful with respect to many of these excluded judgments.

Indeed, the very validity of Kahan’s commentary highlights an interesting, delicious irony. Exactly the types of confirmation biases in perceptions of science highlighted by Kahan’s commentary seem to may characterize social psychological science. There is ample evidence that scientists’ confirmation biases about research conclusions are demonstrably powerful in at least many cases. Social psychological perspectives that emphasize the power of lay confirmation biases in person perception do so on the basis of a highly selective review of the evidence. Any review reaching the conclusion that the evidence shows that person perception is powerfully characterized by confirmation biases must be based on researcher confirmation bias because the evidence so overwhelmingly shows that lay person perception is mostly motivated by the desire to be accurate (e.g., Devine, Hirt, & Gehrke, 1990; Trope & Bassok, 1983). Chapter’s 5 and 8 addressed this issue at length. With respect to seeking information that bears on their interpersonal expectations, in general, the evidence shows that people overwhelmingly seek and prefer diagnostic, not confirmatory, information.
Kahan’s perspective, however, which focuses a great deal on the role of confirmation biases in how people evaluate science, exquisitely describes the production of social psychological theories and conclusions about person perception, and many other topics. Other examples consistent with Kahan’s confirmation bias perspective applied to how psychologists reach conclusions, from *Social Perception and Social Reality* include:

- Overstated claims about the power of self-fulfilling prophecies
- Overstated claims about expectancy- or stereotype-induced perceptual biases
- Underestimations of the power of accuracy, especially though not exclusively stereotype accuracy, and/or dismissals of its “importance”
- Decades of misinterpretations of studies such as Hastorf & Cantril, 1955 and Rosenhan, 1973 as demonstrating the power of bias, when, in fact, they demonstrated overwhelmingly the power of accuracy

That science *sometimes* goes wrong is a normal part of science. But when science goes off the rails and fails to self-correct for decades, especially when the evidence is sitting in plain daylight from within the original published reports, something other than “pure science” may be going on. Kahan’s work points to some likely possibilities, helps understand exactly what. Kahan’s work helps explain the prevalence of questionable interpretive practices (QIPs) – narrative, conceptual, and interpretive means by which scientists can and do reach unjustified conclusions, even in the completely absence of statistical or methodological errors and flaws, and even when findings are replicable, (Jussim Crawford, Anglin, Stevens & Duarte, in press; Jussim, Crawford, Anglin, & Stevens, 2015; Jussim, Crawford, Stevens, & Anglin, in press; Jussim, Crawford, Stevens, Anglin, & Duarte, in press). QIPs captured in *SPSR* include:

**Logical incoherence:** Reaching opposite or contradictory conclusions, as long as both advance one’s preferred narratives, values, theory, or ideology. Simple example: Claiming there are no good criteria for assessing the accuracy of stereotypes yet accepting “known groups validity” as a reasonable way to validate new measures.

**Phantom facts:** Declaring something to be a fact without evidence. Simple example from *SPSR*: Declaring stereotypes to be inaccurate without evidence.
Blind spots: Overlooking or ignoring research that contests one’s preferred perspective. Simple example from SPSR: Citing Darley & Gross’s (1983) single study that they interpreted as showing that stereotypes lead to their own confirmation, and ignoring Baron, Malloy & Albright’s (1995) two failed replications.

Double standards: Subjecting the research producing conclusions one dislikes to withering criticisms, and extolling the virtues and value of research producing conclusions one likes, even when the research one dislikes is of equal or higher methodological quality. Simple example: the common claim that there are no “good” criteria for assessing accuracy, while, at the same time, extolling the power of self-fulfilling prophecies. This is a double standard because both accuracy and self-fulfilling prophecies require showing correspondence between belief and reality, so that the criteria for doing so must be identical.

Psychology and the social sciences are riddled with more conclusions that have gone off the rails (if “the rails” are defined as “scientific evidence”) than can be reviewed here. Exposés of major disconnects between accumulated data and common conclusions have been recently published regarding broad areas within cognitive psychology (Firestone & Scholl, 2015), social psychology (Jussim, Crawford, Anglin et al, in press), social neuroscience (Vul, Harris, Winkielman, & Pashler, 2009), and sociology (Martin, 2015). Over a decade ago Pinker (2002) exposed how political motivations led to invalid claims about education, parenting, crime, personality, evolution, and more.

What is going on here? Is it really possible that trained social psychologists, people with PhDs and years of experience, routinely engage in substantial confirmation bias in interpreting scientific research? Many scholarly perspectives answer this question with a clear, “yes indeed” (for general reviews of scientific susceptibility to confirmation bias, see: Greenwald, Pratkanis, Leippe, & Baumgardner, 1986; Ioannidis, 2012; Lilienfeld, 2010). For a review of how confirmation biases have led social psychology to specific unjustified conclusions in areas such as discrimination, stereotype threat, unconscious influences on sensory perception and more, see Jussim, Crawford, Anglin, et al (in press). “Successful” motivated reasoning driven by the goal of reaching some particular conclusion requires information, experience, and skill with logic and argumentation. People with PhDs and extensive training – especially those with training in telling “compelling narratives” (Bem, 2002; Jordan & Zanna, 2007) – are more able to dismiss findings they do not
like and defend findings they do like in the face of challenges than are less intelligent and less well-trained laypeople. Indeed, Kahan himself (Kahan et al, 2012) has found that views about climate change become more polarized as people’s science knowledge increases (see also Haidt, 2012).

An even stronger view is presented by Indeed, I am grateful to Madison, Dutton, & Stern when for their commentary highlighting scholarship on the clever sillies – which presents a perspective suggesting just how extremely distorted “scholarly” conclusions can get. Much of that research suggests that social scientists who are obviously very intelligent and have extraordinary levels of knowledge and expertise manifestly silly claims primarily to signal their intelligence (Charlton, 2009; Dutton & van der Linden, 2015). Because manifestly silly ideas are often presented in high-falutin and sophisticated-sounding language, they can appear rigorous and (to paraphrase Stephen Colbert) high in “scientificiness” and, therefore, can create an illusion of plausibility and validity. In the social sciences, such ideas often include the denial of evolutionary or biological bases of human psychology and behavior (see, e.g., Pinker, 2002 for a broad review), the denial of stereotype accuracy, and, I would argue, attempts to stigmatize and ostracize those who point out that the data does not always advance social scientific narratives that are presumed to advance the interests of the oppressed (Gottfredson, 2010; Pinker, 2002).

I do not doubt that desire to signal one’s brilliance may indeed be one motivation underlying the clever sillies, but I do not think it is the only one, and, perhaps, not even the most important one in the social sciences. In addition to signaling intelligence, staking out positions that are logically incoherent or disconnected from scientific evidence can signal not just intelligence, but one’s political allegiances, one’s moral positions, and that one is on the “side” of one’s colleagues fighting “the good fight” (e.g., Kahan, in press). The extent to which scientific distortions, such as the denial of stereotype accuracy or evolutionary influences on psychology, result from motivation to signal one’s egalitarian bona fides to one’s colleagues, the desire to advance one’s politics, values, and morals, or other less politicized sources is an important empirical question for the burgeoning area of meta-science and scientific integrity (e.g., Ioannidis, 2012; Jussim, Crawford, Anglin, et al, in press; Simmons et al, 2011).
Was Planck Really—Right?

“A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it.” Max Planck.

Self-correction is often taken to be a hallmark of science. Whereas religious, political, or moral beliefs may not be subject to change based on evidence, scientific beliefs, presumably, should be subject to change when sufficient new evidence contradicts existing conclusions. For example, Jost (2011) wrote: “This is because we, as a research community, take seriously the institutionalization of methodological safeguards against experimenter effects and other forms of bias. Any research program that is driven more by ideological axe-grinding than valid insight is doomed to obscurity, because it will not stand up to empirical replication and its flaws will be obvious to scientific peers.”

If only it were so. SPSR presented numerous cases where: 1. An initial high impact “wow!” study yielding some dramatic result was published; 2. Many follow-up studies revealed that the conclusions based on that “wow” study were mostly not justified; and 3. The “wow” conclusions continued to march on for decades as if the correctives were never published. SPSR documented case after case of just this pattern with respect to self-fulfilling prophecies, biases, stereotypes, and accuracy.

Many of the commentaries (Bonnefon, Hopfensitz & De Neys, Kihlstrom, Little, Madison et al, Martin, Mousavi & Funder, Trafimow, Wagoner) seemed to welcome SPSR as a much-needed corrective to the “stupidism” (Kihlstrom) emphasized by much social psychology and the “clever silly” (Madison et al) perspectives that back up such claims. Many of the rest acknowledged the validity of its main points but raised issues beyond the scope of the book (Church, Kahan, Wilson & Huang). If one is to believe the consensus of the commentaries on SPSR, one might believe that the field’s emphasis on “stupidism” is in decline. Although I hope that is true, based on too much evidence from outside these commentaries, such a conclusion is premature, and not only because two of the commentaries committed the processistic fallacy when attempting to defend claims emphasizing lay “stupidism” (Bian & Cimpian, Terbeck) regarding
My collaborators and I have recently updated the review of stereotype accuracy work that appears in *SPSR* (Jussim, Crawford, Anglin, Chambers, Stevens, & F. Cohen, 2016; Jussim, Crawford, & Rubinstein, 2015). Over 50 studies have been identified, almost double the number reviewed in *SPSR* because there has been an explosion of research on the accuracy of stereotypes about national character and political groups. The main conclusions of *SPSR* were reconfirmed, especially regarding the demographic stereotypes that social scientists generally seem most concerned about. Stereotype accuracy is one of the largest effects in all of social psychology. It has been replicated in multiple independent labs. Given social psychology’s current crisis of replicability, and widespread concerns about questionable research practices (e.g., OSF, 2015; Simmons et al, 2011), one might expect that social psychologists would be shouting to the world that we have actually found a valid, independently replicable, powerful phenomena.

But if one did think that, one could not possibly be more wrong. Testaments to the inaccuracy of stereotypes still dominate textbooks and broad reviews of the stereotyping literature that appear in scholarly books (see Table 1). The new generation of scholars is still being brought up to believe that “stereotypes are inaccurate,” a claim many will undoubtedly take for granted as true, and then promote in their own scholarship. Sometimes, these manifest as definitions of stereotypes as inaccurate; and even when stereotypes are not defined as inaccurate, they manifest as declarations that stereotypes are inaccurate, exaggerated, or overgeneralized.

*Perhaps we should not expect ready self-correction on a hot-button topic like stereotype accuracy. It seems to offend many social scientists’ egalitarian sensibilities, so what about other topics? Surely we readily self-correct on other topics, don’t we?*

**Facilitating Self-Correction in Psychological Science Regarding Accuracy, Bias, and Self-Fulfilling**

Psychology is abuzz with an internal discussion of how it can do better. Greater transparency, pre-registration, replication, and more have all come to the fore. *...I view these as likely positive developments that offer considerable hope that psychological science will produce fewer false conclusions. I hope these changes also mean the field is becoming more ready, willing, and able to correct false conclusions when they do get...*
stereotype biases, and most of the attempts to dismiss the power or importance of accuracy did not result primarily from failed replications or questionable statistical or methodological practices, or even lack of transparency. Instead, they are problems of interpretation and (exactly as Kahan’s commentary and perspective might predict) researcher confirmation biases. Even when failed replications did get published, they were generally ignored. Effect sizes were largely ignored. Simple contextual factors (such as the number of plays in a football game, or the total number of judgments made by staff at psychiatric institutions) that could have reigned in wildly overstated claims of bias were often simply ignored, not just by the original researchers, but by decades of subsequent scientists perpetuating the erroneous testaments to bias. Attention to contextual, statistical, and methodological details was seemingly short-circuited by the ability or desire to tell compelling “wow!” stories about the power and pervasiveness of expectancy effects.

What, then, can researchers who are not grinding axes and who want to present valid and nuanced descriptions of the findings do to limit their vulnerability to perpetuating false claims that routinely appear in scientific literatures? Unfortunately, psychology does not have a consensus on the answers to this question, and is currently in the process of searching for those answers. This issue has been addressed more generally elsewhere (e.g., Jussim, Crawford, Anglin et al, in press). Here, so here, I focus specifically on: 1. Identifying general principles that may be broadly applicable; and then 2. Give examples of how they could be applied to the literatures addressed by SPSR:

1. Resist the urge to tell compelling narratives by glossing over or ignoring contradictory findings and conclusions.
   - Stop citing Rosenthal & Jacobson (1968) as showing that teacher expectation effects are powerful or pervasive.
   - Do not assume that “story studies” — famous classics around which compelling narratives can be told — are necessarily true or replicable. Review the entire relevant literature before making claims regarding expectancies and stereotypes.
   - Avoid cherry-picking a biased sample of studies about expectancies or stereotypes (or any
other topic) to make an argument.

2. **Focus on the actual results of studies, rather than researcher claims about those results.**
   - One can often find evidence of substantial accuracy and rationality in studies that emphasize or only reported error and bias.
   - Biases and self-fulfilling prophecies may be quite modest, or even contingent on moderators, even when the discussion touts their power and pervasiveness.

3. **Search for skeptical reviews and meta-analyses, and do not depend exclusively on reviews or meta-analyses that appear to have as an agenda persuading the reader.** Avoid repeating conclusions based on famous reviews, without either critically examining the basis for those conclusions, or, at least, searching the literature to find out whether other, perhaps less famous but more persuasive, skeptical or critical reviews or meta-analyses have reached different conclusions.

   Abide by the Mertonian (1942/1973) norm of universalism, that evaluation of scientific claims hinges *not at all* on the status or prestige of the scientist making them, but on the quality of the evidence, logic, and argument being put forth.
   - For every review testifying to the power of expectancies, there are now others casting doubt on such conclusions. If one must make a point about expectancies, at minimum, one can reflect the state of the literature with statements such as:
     - “Whereas some reviews have concluded that expectancy effects are powerful and pervasive, others have concluded that such effects are weak, fragile, and fleeting.”
     - “Although stereotypes have long been presumed to be inaccurate, several reviews have concluded that, in general, stereotypes are often at least moderately accurate.”
     - “Although social constructionist phenomena undoubtedly occur and can sometimes be powerful and important, at the level of individuals interacting with other individuals, such effects are usually quite modest.”
     - “Although people undoubtedly cognitively construct their social perceptual worlds to a considerable degree, and, sometimes such constructions can be quite biased, this
does not mean their constructions are always or even mostly inaccurate.”

4. **In new original studies, be excessively transparent about methods and results.** Provide means, standard deviations, and correlations for all variables. When available and relevant, provide frequency distributions. When reporting regression and SEM results, report standardized and unstandardized coefficients and also the t- and F-values associated with each test of significance. If all this cannot make it into the main report, then at least provide it in supplementary materials. Report effect sizes and confidence intervals. This should be done when reporting new empirical studies; and it should be routine when reviewing empirical literatures.

- This is especially important when making claims about the relative power of bias versus accuracy. Distorted claims about bias could have been detected decades earlier, if, e.g., effect sizes had been routinely reported, and if contextual data (e.g., total number of judgments) had been reported.

5. **When reviewing literatures, make frequent use of the new range of skeptical forensic techniques, such as p-curves, funnel plots, forest plots, replication indices, and tests for publication biases to evaluate whether the published literature can be taken at face value.** Meta-analysis is no longer enough, because revelations of publication biases and p-hacking mean that there can be ample published literatures “demonstrating” non-existent effects, or greatly overestimating effect sizes.

- If one defines stereotypes as inaccurate or as exaggerations, then one must be willing to accept that only beliefs about groups that have been demonstrated to be inaccurate and exaggerations among the sample one is studying can be known to be stereotypes.

- One can avoid this problem by defining stereotypes in ways that permit them to be accurate, avoiding presumptions of inaccuracy, exaggeration, or overgeneralization.

6. **Base empirical claims about the state of the world on actual empirical evidence.**

- This may seem obvious, but researchers have been making claims about stereotype inaccuracy without evidence for decades. See Pinker (2002) for similar claims without evidence.
regarding a range of issues, such as human malleability and the role of social factors in everything from intelligence to aggression to sex differences.

- Avoid the processistic fallacy. Do not make claims about error, bias, or the inaccuracy of stereotypes on the basis of process studies, even ones that identify faulty processes in the lab that one speculatively presumes will cause inaccuracy in people’s naturally-occurring judgments. Such processes might have theoretical import (Mook, 1983), and they might generate predictions regarding patterns or sources of inaccuracy. But they do not constitute evidence of inaccuracy.

- Reach conclusions about stereotype accuracy on the basis of studies reporting empirical data rather than sources (even “authoritative” ones such as G. W. Allport, 1954/1979; see also Table 1) declaring stereotypes to be inaccurate (or exaggerations) without data.

- Do not claim that characterizing stereotypes as possessing a “kernel of truth” constitutes some sort of acknowledgement that stereotypes are often substantially accurate. This functions as a disingenuous attempt to maintain the emphasis on inaccuracy, which can readily be seen with a “turnabout test” (Duarte, Crawford, Stern, Haidt, Jussim, & Tetlock, 2015; Tetlock, 1994): Would declaring, “Psychological research has a kernel of truth” be a great testament to the validity of psychological science?

- If stereotypes do influence judgments regarding an individual target do not assume that increases inaccuracy without testing for accuracy.

8.7 Build rational judgment processes into theoretical perspectives on social perception.

- Because of social psychology’s infatuation with error and bias, almost any result, no matter how reasonable and rational, has been framed as flawed. However, such conclusions regarding lay judgments require showing that some particular perceptual result deviates from some normative model. In social psychology, this is rarely done, thereby liberating researchers to cast almost any result as irrational.

- Social psychologists should stop casting results as irrational absent development of normative
model of rational judgment and an assessment of the extent to which lay judgments both correspond to and deviate from that model.

- Social psychologists studying social perception should start developing models of rational judgment processes if they wish to continue reaching judgments about irrationality.

9. **Be clear and consistent with respect to levels of analysis.**

- If one is discussing perceptions of groups, then accuracy refers to correspondence between beliefs about groups and what those groups are like.
- If one is discussing perceptions of individuals, then accuracy refers to correspondence between beliefs about an individual and what that individual is like.
- Cease confounding levels of analysis by declaring that stereotypes are inaccurate because they do not apply to every individual.

10. **Read some of the rapidly growing literature on how researcher confirmation biases—personal, theoretical, and political—can and have distorted conclusions in psychology, sociology, and political science (e.g., Duarte et al., 2015; Jussim, Crawford, Anglin, et al., in press; Lilienfeld, 2010; MacCoun, 1995; Martin, 2015; Pinker, 2002; Tetlock, 1994).**
References


Jussim, L., Crawford, J. T., Anglin, S. M., Stevens, S. M., & Duarte, J. L. (In press). Interpretations and
methods: Towards a more effectively self-correcting social psychology. *Journal of Experimental Social Psychology.*


Figure 1: The Reflection-Construction Model (Jussim, 1991)

**Figure 1A: The Full Model**

Predictive validity

- Background Information (anything on which perceivers may base their expectations)
- Sources of expectations
- Interpersonal Expectations
- Self-fulfilling Prophecy
- Targets’ Behavior or Attributes
- Perceivers’ Judgments of Targets

**Figure 1B: Constructive Accuracy**

Predictive validity

- Background Information
- Sources of expectations
- Interpersonal Expectations
- Bias
- Targets’ Behavior or Attributes
- Perceivers’ Judgments of Targets

Constructive accuracy: Even when perceivers are completely oblivious to targets’ behavior or attributes, their judgments of targets’ will still correspond to (correlate with) targets behavior or attributes if 1. expectations are based on background information that 2. predicts targets behavior or attributes; and if 3. expectations influence (bias) perceiver judgments.
### TABLE 1
Modern Claims about Stereotype (In)Accuracy

<table>
<thead>
<tr>
<th>Scholarly Books</th>
<th>Expects strong evidence of stereotype accuracy</th>
<th>Reviews little or no evidence of accuracy and either dismisses accuracy as unimportant or emphasizes stereotype inaccuracy and bias</th>
<th>Defines/declares stereotypes to be inaccurate</th>
<th>Representative Quotes:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Banaji &amp; Greenwald, 2013, <em>Blindspot: The hidden biases of good people</em></td>
<td>✓</td>
<td></td>
<td></td>
<td>P. 74: Because all stereotypes are partly true and partly false, it may seem pointless to debate their accuracy. P. 89: … stereotyping is an unfortunate by-product of the otherwise immensely useful human ability to conceive the world in terms of categories.</td>
</tr>
<tr>
<td>Brown, 2011, <em>Prejudice: Its Social Psychology</em></td>
<td>✓</td>
<td></td>
<td></td>
<td>p. 71: … the question of whether stereotypes are ‘objectively’ (in)accurate is only of marginal interest to most students of prejudice.</td>
</tr>
<tr>
<td>Fiske &amp; Taylor, 2008, <em>Social cognition: From brains to culture</em></td>
<td>✓</td>
<td></td>
<td></td>
<td>P. 282: Stereotyping is the cognitive aspect of bias … and it comes in both blatant and subtle forms.</td>
</tr>
<tr>
<td>Whitley &amp; Kite, 2009, <em>The psychology of prejudice and discrimination</em></td>
<td>✓</td>
<td></td>
<td></td>
<td>P. 100: At the group level, then, stereotypes may have a kernel of truth, but relying on them at the individual level may lead to serious judgment errors</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Textbooks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aronson, 2011, <em>The social animal</em></td>
</tr>
<tr>
<td>Baumeister &amp; Bushman, 2014, <em>Social psychology and human nature</em></td>
</tr>
</tbody>
</table>
Crisp & Turner, 2014, *Essential social psychology*

**P. 57:** Once a category is activated we tend to see members as possessing all the traits associated with the stereotype.

Greenberg, Schmader, Arndt, & Landau, 2015, *Social psychology: The science of everyday life*

**p. 352:** Even though this kernel [of truth] might be quite small, with much more overlap between groups than there are differences, as perceivers we tend to exaggerate any differences that might exist and apply them to all members of the group.

Grison, Heatherton, & Gazzaniga, 2015. *Psychology in your life*

**P. 385:** Indeed, some stereotypes are based in truth: Men tend to be more violent than women, and women tend to be more nurturing than men. However, these statements are true on average.

King, 2013, *Experiencing psychology*

**p. 402:** A stereotype is a generalization about a group’s characteristics that does not consider any variations from one individual to another.

Schachter, Gilbert, Wegner, & Nock, 2015, *Introducing psychology*

**P. 403:** … stereotyping is a useful process that often produces harmful results, and it does so because stereotypes have four properties: They can be (1) inaccurate, (2) overused, (3) self-perpetuating, and (4) unconscious and automatic.

Table 1 reprinted from Jussim et al, 2015.
### Table 2: Social Psychological Self-Correction?

<table>
<thead>
<tr>
<th>Publication</th>
<th>Narrative</th>
<th>Key Aspects of Methods</th>
<th>Citations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Darley &amp; Gross, 1983</td>
<td>Stereotypes lead to their own confirmation; stereotype bias in the presence but not absence of individuating information</td>
<td>People judge targets vs. without relevant individuating information. Single experiment. N=50-68, depending on analysis.</td>
<td>Total: 1054</td>
</tr>
<tr>
<td>Van Hiel-Onraet, and De Pauw (2010)</td>
<td>Liberal/conservative psychological differences in cognitive style were modest to nonexistent.</td>
<td>Meta-analysis of 124 studies, including five unpublished studies. Clear articulation of study selection criteria.</td>
<td></td>
</tr>
<tr>
<td>Doyen, Klein, Pichon, &amp; Cleermans (2012)</td>
<td>Failed replication of Bargh et al., 1996. No effects of stereotypes on behavior except when experimenters were not blind to condition</td>
<td>Two close replication and extension experiments. Total N=170.</td>
<td></td>
</tr>
</tbody>
</table>

continues on the next page...
Footnote

1 Strictly speaking, the 14th amendment and Civil Rights acts focus on behaviors (discrimination), rather than beliefs. However, failure to provide Alfonso service because the provider believes Latinos are mostly criminals is a violation of those acts; however, failure to provide Alfonso service because the provider believes Alfonso is a criminal is not. Whether the behavior is based on a stereotype or on individualizing information is taken as extremely important, thereby highlighting the perceived value of the distinction in legal contexts.
Accuracy, Bias, Self-Fulfilling Prophecies and Scientific Self-Correction

Lee Jussim

Rutgers University
Abstract

In my précis of *Social Perception and Social Reality (SPSR)*, I argued that the social science scholarship on social perception and interpersonal expectancies was characterized by a tripartite pattern: 1. Errors, biases, and self-fulfilling prophecies in person perception were generally weak, fragile, and fleeting; 2. Social perceptions were often quite accurate; and 3. Conclusions appearing throughout the social psychology scientific literature routinely overstated the power and pervasiveness of expectancy effects, and ignored evidence of accuracy. Most commentators concurring with the validity of these conclusions. Two, however, strongly disagreed with the conclusion that the evidence consistently has shown that stereotypes are moderately to highly accurate. Several others, while agreeing with most of the specifics, also suggested that those arguments did not necessarily apply to contexts outside of those covered in *SPSR*. In this response, I consider both the limitations to the tripartite pattern, the role of politics and confirmation biases in distorting scientific conclusions, common obstructions to effective scientific self-correction, and how to limit them.
When I began writing *Social Perception and Social Reality* (hence abbreviated as *SPSR*), my goal was to offer a corrective to a slew of manifestly false claims in common conclusions about social perception. The *Precis* summarized this as the *tripartite pattern*: 1. Errors, biases, and self-fulfilling prophecies in person perception were occur, but are generally weak, fragile, and fleeting; 2. Social perceptions were often quite accurate; and 3. Conclusions appearing throughout the psychological literature (with educational psychology as a notable exception) were often unhinged from these data by virtue of routinely declaring expectancy effects powerful and pervasive, and consistently ignoring evidence of accuracy. I also argued that defining stereotypes as inaccurate is logically incoherent, and that, in sharp contrast to 100 years of claims to the contrary, the evidence is that stereotypes are often (not always) quite accurate. All of this was “news,” not because of new dramatic data, but because the old data never justified the strong claims about the power of expectancies to create social reality. To paraphrase Winston Churchill’s characterization of the early architects of appeasement, social psychology periodically stumbled on the truth but simply picked itself up and hurried along as if nothing had happened.

The evidence supporting an emphasis on error, bias, and irrationality in social perception was maintained by virtue of overstatements of what research actually found (e.g., by ignoring effect sizes), selective citation of a small number of dramatic “wow effect” studies (Jussim, Crawford, Anglin, Stevens, & Duarte, in press), many of which have proven difficult to replicate. The error and bias (“stupidism” as per Kihlstrom) perspective was also maintained by reliance on a “toolbox” of double standards, blind spots, and researcher confirmation biases that served to elevate the supposed importance of “bias” results and studies while denigrating or dismissing evidence of accuracy. Indeed, *SPSR* pointed out that even some of the most classic demonstrations of “bias” in social perception (e.g., Hastorf & Cantril, 1955; C. E. Cohen, 1981; Rosenhan, 1973) actually provided more evidence of accuracy than of bias – as did many of the follow-ups to classic studies such as Rosenthal & Jacobson (1968), Snyder & Swann, 1978, and Darley & Gross (1983).

Nearly all of the commentators accepted the fundamental conclusions of *SPSR*. Of the seventeen commentaries, only two took issue with the major conclusions (Bian & Cimpian; Terbeck); and two disagreements occurred, not over conclusions, but over definitions (Andrews, Springer). Two others similarly
did not take issue with the tripartite conclusion, but also suggested the phenomena outside the scope of *SPSR* might have yielded different conclusions (Kahan, Wilson).

There was abundant embracing of the main conclusions. Trafimow’s consideration of statistical effect sizes concludes that the weight of the evidence supports the tripartite conclusions. Two of the commentators (Marczyk, Little) pointed out that it would be bizarre if evolution did not lead us to be in touch with reality much of the time. Several commentators pointed to evidence outside the scope of *SPSR* that also often yields evidence of accuracy and rationality (Kihlstrom, Mousavi & Funder, Bonnefon et al, Wagoner). And one commentary (Madison et al) endorsed and expanded on one of the undercurrents of *SPSR* – that politics and dogmas have probably distorted the scientific conclusions in this and many other areas of social science.

Does this mean psychology has finally turned a corner and is self-correcting towards fully recognizing that social perception is often largely rational and accurate and generally only modestly subject to error and bias? Given the positive nature of most of the commentaries, one might suspect that the answer is “yes indeed.” One would, however, be wrong. Before returning to that question, however, I respond to the specific issues raised in the commentaries.

Despite largely supporting the general perspective in the book, even the vast majority who did not contest the general conclusions often did point out gaps in the theorizing and that other phenomena not addressed in the book often yielded greater evidence of error, bias, and constructionist processes. A smaller number did take issue with some of the book’s key conclusions. Each of these are discussed next.

**Definitional Differences**

**Sensory perception vs. social perception.** Church correctly pointed out that the research reviewed in *SPSR* does not make a hard distinction between sensory perception and more global cognitive representations, such as interpretations, beliefs, judgments, etc. In fact, to social psychologists, *social perception* rarely deals with sensory perception per se, and generally deals with “perception” in this more cognitive and molar sense. When we seek answers to questions like, “How do voters perceive President Obama?” we are almost always interested in beliefs, attitudes, and opinions about Obama, not their sensory perceptions. I am glad to have the opportunity to clarify that *SPSR*’s focus is on molar social perceptions, and not sensory perceptions. However,
the idea that beliefs and expectations influence sensory perceptions has even less support than it does with respect to molar social perceptions (Firestone & Scholl, in press).

If every concept is a really a “stereotype” the term loses all meaning. Andrews’ commentary advocates abandoning the distinction between stereotypes and individuating information. I disagree. If anything one person believes about another’s behaviors and characteristics is a “stereotype” then the term loses all meaning and usefulness as a theoretical construct.

Social psychologists (and I suspect many other people – see, e.g., the 14th Amendment to the U.S. Constitution, and the Civil Rights acts of the 1960s) usually consider it important to understand where beliefs (at least, if those beliefs influence behavior) about, e.g., Alfonso, originate. Do they result from: 1. his being a male or Latino or 2. The fact that he acts in his local theater group, earned a law degree from Harvard, likes skinny jeans, and vacations in the Bahama"s? It does not matter much, I suspect, whether we call them stereotypes vs. individuating information, or “concepts” about groups and concepts about individuals. What is important is that beliefs about Latinos are not the same thing as beliefs about Alfonso. They may have some superordinate similarities (both are beliefs), in the same way that beliefs about apples (they are usually red and good to eat) have some superficial similarity to this apple (which is brown and rotting and not good to eat).

Nonetheless, both apples and this apple are concepts and, in that way, have some similarity. The distinction is also important, however, because it highlights important differences between apples and this apple, and between Alfonso in particular, and Latinos in general.

Filling in Gaps and Expanding the Scope

Evolutionary explanations for (in)accuracy. Two commentators brought up the important theoretical point that evolutionary theories could explain both why there is relatively high levels of accuracy in social perception and also conditions likely to produce low accuracy because adaptations to advance goals other than accuracy might take precedence (Little, Marczyk). These commentators: 1) Are correct in pointing out that SPSR did not draw on evolutionary perspectives; 2) point out that doing so is likely to provide important theoretical advances to understanding conditions under which accuracy is likely to be higher or lower. At its most basic level, it is hard to imagine any successful organism that has evolved to have completely invalid
reactions to its environment.

On the other hand, evolution emphasizes adaptations that enhance an organism’s ability to produce viable offspring. Thus, as both Marczyk and Little correctly point out, there are many goals that might accomplish this that have little relevance to, or which might even conflict with, accuracy (e.g., identifying and attracting fertile mates, attracting resources to support offspring, etc.). Deception is prominent in the animal world (and even occurs in the plant world; consider carnivorous plants posing as nectar-rich flowers) because there are so many ways that it could have adaptive advantages. Thus, it also seems implausible that evolution would yield social perceptions that were perfectly accurate (von Hippel & Trivers, 2011). In psychology, error management theory (Haselton & Buss, 2000) was, as pointed out by Little, an early and constructive attempt to identify when evolutionary pressures were more likely to lead to accuracy versus certain specific types or patterns of errors. Indeed social psychological theorizing will likely be enhanced and sharpened by further efforts to exploit evolutionary ideas to understand when people are likely to be accurate and when they are likely to be systematically inaccurate.

Accuracy in other contexts. Several commentators pointed out that, in other contexts, there is often: 1. Surprising evidence of accuracy; and 2. Similar patterns of political or theoretical double standards in evaluations of the evidence. Such double standards occur when people hold research that advances their theoretical perspectives or political values to lower standards than they hold research that opposes their theoretical perspectives or political values. For example, research demonstrating biases produced by stereotypes or prejudice to be “powerful and important” (see many examples in SPSR), while, at the same time, downplaying, ignoring, or dismissing the typically far more powerful effects of individuating information.

Bonnefon, Hopfensitz, and De Neys, for example, point out these issues in the study of the accuracy of judgments of trust at zero acquaintance. Although the levels of accuracy are much lower than among expectancies, they point out (correctly, in my view) that any accuracy on the basis of a mere photograph is quite striking. Furthermore, their analysis strongly suggests that political/advocacy goals have led to a set of logically contradictory conclusions about accuracy in perceptions of trust, in a way quite reminiscent of the
double standards and logical incoherence I identified with respect to self-fulfilling prophecies and stereotypes. In short, perceptions of trustworthiness have been declared both inaccurate and self-fulfilling, and these are mutually exclusive conclusions. A belief that a target is untrustworthy can be, at one moment in time, inaccurate, and the next, self-fulfilling such that the target becomes untrustworthy. After the target becomes untrustworthy, however, subsequent perceivers are not wrong for believing the target to be untrustworthy. They are accurate. Consequently, Bonnefon et al correctly point out that perceivers’ beliefs about trustworthiness cannot be generally inaccurate and self-fulfilling.

Mousavi and Funder similarly point out that judgments are often “ecologically rational,” meaning that they are adapted well to their environments. Fast and frugal heuristics, though technically constituting “biases,” especially in laboratory studies, often lead to moderately to highly accurate judgments in much of the rest of daily life. They, too, echo the political implications of accuracy, pointing out that an overweening emphasis on error and bias misses a great deal of evidence of accuracy. This is important, they argue (and I agree), in part, because in such situations, efforts to solve social problems by changing supposedly erroneous beliefs are doomed to failure because the beliefs are not particularly erroneous in the first place.

Wagoner pointed out that distortions of the scientific record similar to those described in Social Perception and Social Reality have long characterized perspectives on memory. The schema concept, which is hypothetically at least neutral with respect to accuracy (as are interpersonal expectancies) has become virtually synonymous with error and bias (as have interpersonal expectancies). That many modern perspectives have just as blithely ignored Bartlett’s (1932) balanced view of accuracy/error in memory as F. H. Allport’s (1955) balanced views on perception is a testament to the long reach of the distorting power of theoretical perspectives emphasizing distortion.

Constructivism (Both Cognitive and Social) Lives!

Cognitive constructivism. Kihlstrom agrees with the general thrust of SPSR but also urges not to throw out the baby (cognitive constructivism) with the bathwater (the excessive emphasis on error and bias). And nor did I intend to do so. Kihlstrom’s commentary presents a very thoughtful and balanced view of realism and constructivism, and is a great primer on how social psychology can be enriched by not dismissing ideas
from any of those broad perspectives writ large.

The cognitive constructionist processes highlighted by Kihlstrom undoubtedly can and do influence memory and the types of social perceptual processes addressed in SPSR. And, surely, sometimes those effects constitute do indeed errors and biases. However, constructivism and error/bias are not synonymous. Although Kihlstrom did not argue that they were, because cognitive constructivist is sometimes presumed to mean something like “perceivers making stuff up that supports their pre-existing beliefs, expectations, and values” (Kihlstrom’s own emphasis on the over-reach of “stupidism” perspectives; see also commentaries by Mousavi & Funder; Madison, Dutton, and Stern; and Wagoner), it is, perhaps, worth walking through why bias and constructivism are not synonymous.

“Constructive accuracy” refers to the process by which expectancy-induced “biases” can increase accuracy (Jussim, 1991). Figure 1A presents the Reflection-Construction Model (Jussim, 1991), which depicts relations among the key variables involved in accuracy, bias, and self-fulfilling prophecy. Figure 1B depicts constructive accuracy. It shows that impression accuracy (correspondence between perceivers’ judgments of targets and those targets’ behavior or attributes) can be quite high, even when perceivers base their judgments of individual targets exclusively on their own expectations, and are oblivious to (ignore, overlook, do not have access to) targets’ actual behavior or attributes. If all three paths shown are high enough, perceiver judgments will correspond to (correlate with) target behavior or attributes, even though perceiver judgments are heavily based on their own expectations and not at all based on target behaviors or attributes. This is because, in Figure 1B, the correlation between perceiver judgments and target behavior or attributes is the multiplicative product of the three paths. For example, if all three equal .8, then impression accuracy equals .8³ = .51. In psychological, rather than mathematical terms, this means that, if perceivers’ expectations are strongly based on highly valid information, the more they rely on those expectations when judging targets, the more accurate they will be.

This is a constructive phenomenon, because, in this example, the judgment is based entirely on perceiver expectations, with no direct input from targets’ actual behavior or attributes. Furthermore, even if perceivers do partially base their judgments directly on targets’ behavior or attributes, relying on accurate expectations
can still increase accuracy further (see Jussim, 1991, for details). Thus, I concur with Kihlstrom that constructive processes can and do play an important role in person perception; however, I would emphasize that, even so, such processes may, at least when those expectations are themselves based on valid information, increase rather than reduce accuracy.

**Social constructionism.** Two of the commentaries (Tappin, McKay, & Abrams; Wilson & Huang) make measured appeals to not completely throw out the social constructionist baby with (what I would call) the excessively political bathwater. However, both commentaries, in somewhat different ways, present defenses of social constructionist processes. Tappin et al do so by arguing for the importance of collective action as a major influence on social reality; Wilson & Huang do so by emphasizing the role of institutions in creating social reality. I see these arguments as mutually reinforcing so I address them both here.

This reply is not the place for a comprehensive review or critique of social constructionism, which is a single term that refers to quite a variety of perspectives. I would, however, divide social constructionism into two main veins (which are not necessarily mutually exclusive). One is primarily a political liberation perspective, with the goals of combating oppressive and hegemonic practices and discourses, in part, by revealing them, thereby advancing the interests of certain groups that the users of such terms deem unfairly victimized or exploited in some way. Few in psychology capture the politicized nature of the efforts better than Jost & Kruglanski (2003. P. 175) who approvingly declared: “From this perspective, we have a professional obligation to weigh in on ideological issues, policies, and decisions” and later on the same page, “The social constructionist movement emerged on the social science scene as a force for change and cultivated a leftist revolutionary spirit that posed a distinctive challenge to established scientific authority.” As a movement with primarily political goals, my view is that this sort of politicization has little place in scientific psychological theorizing.

However, a separate vein of social constructionism aspires to be bona fide social science. Not all forms of social constructionism are blatantly political or liberationist. Some, instead, seeks to understand social relationships, including but not restricted to relationships of power and status, and the reciprocal influences among individuals and institutions (regardless of whose interests or advocacy agendas these understandings
might advance). In this spirit, Tappin et al are surely correct in arguing that, under some conditions, collective actions can alter the nature of intergroup relations.

Nonetheless, it is also, perhaps, worth pointing out that little, if any, of the research they cited links laypeople’s actual stereotypes to collective action. They review abundant evidence that target groups are motivated by perceived slights to and injustices inflicted on their groups; but none of it links laypeople’s actual stereotypes to collective action. I suspect that this failure stems from: 1. Social psychologists taking for granted that laypeople hold unjustified and pernicious stereotypes; so it was not even necessary to assess actual stereotypes, the derogatory nature of stereotypes could simply be taken for granted.

Space does not permit a citation by citation critical analysis of the work they presented in support of their perspective, so perhaps one example will suffice. They cite a series of studies by Ellemers and Barreto (2009) three times in their short commentary, so they seem to consider it important to their perspective. Ellemers and Barreto (2009) showed that believing others had an insulting view of one’s group (e.g., for women, someone believing that women are unintelligent), motivated collective action. But what if lay people do not routinely believe women are unintelligent? There is evidence that people believe boys are better at math than girls, but the same studies show that people also believed girls have higher verbal skills than boys (Swim, 1994).

Regardless, Ellemers and Barreto (2009) did not assess anyone’s sex stereotypes regarding intelligence, nor do they review research that has done so. Their findings are still interesting, because they say something about how perceived intergroup insults motivate collective action, but not because it says anything about the role of the actual stereotypes held by any actual people in leading to collective action. This is not meant to dismiss the perspective entirely. In fairness, it probably can be interpreted as showing that when people hold derogatory stereotypes, if targets become aware of those stereotypes, collective action may result. Of course, history is filled with counter-examples, cases where people did hold derogatory stereotypes and collective action did not result over vast periods of time (consider, e.g., the inferior status ascribed women, three hundred years of slavery in the United States, and the Hindu caste system). I suspect, therefore, that the “holding insulting stereotypes – collective action” link is tenuous at best, and subject to many conditions not articulated in either Tappin et al’s commentary or much of the underlying research. Indeed, I do not doubt the effect
exists, but I would suspect that, in the real world, across many situations and contexts, it fits the pattern described in SPSR: occasionally strong, but usually quite weak, fragile, and fleeting.

Wilson & Huang are correct in pointing out that my conceptual analysis “freezes” institutions at a given point in time, and then examines the rationality vs. biased nature of social perception. As they point out, institutions are not actually frozen, and are subject to both slow-moving and, occasionally, dramatic and sudden changes. My argument was never intended to be “there are no conditions under which stereotypes or social beliefs construct reality.” Indeed, SPSR is peppered with both real world examples and scientific studies showing that, sometimes, such effects can be quite powerful.

SPSR, however, did not have as its purpose identifying the nature of collective movements or the inter-relationships between institutions, demographic groups, and individuals. Instead, the purpose of the book was to review evidence regarding the extent to which individuals’ beliefs about groups or other individuals were accurate, biased, or self-fulfilling. This is, as these commentaries suggest, a limitation of its scope. It is certainly an important and appropriate social science endeavor to address issues such as collective action and institutions. SPSR made no claims about such issues. It was, however, an attempted corrective to longstanding and unjustified social science claims about how individuals’ beliefs relate to social reality – and, on this issue, it is, perhaps, worth noting that both Tappin et al and Wilson & Huang presented neither argument nor evidence against the central claim that such corrective is justified and past due.

Victims of the Processistic Fallacy

Two commentaries aspire to refute the conclusion reached in SPSR that stereotypes have been widely found to be at least moderately accurate. Both Terbeck and Bian & Cimpian propose processes that they believe cause inaccuracy in stereotypes. Both critiques fall victim to the processistic fallacy, which was addressed in SPSR. Thus, my response to these critiques begins by quoting that text (p. 394):

The processistic fallacy involves concluding that laypeople’s beliefs must be inaccurate because researchers have discovered cognitive processes that the researchers believe to be flawed. This is a fallacy for several reasons: (1) The process may not be as flawed as the researchers believe, and its degree of “flaw” cannot be assessed without assessing the validity or success of the judgments and decisions by people who do versus do not rely on this process (something social scientists rarely do); (2) even if the process is indeed flawed, in real life, people may rely on many other less flawed processes when making judgments and decisions; and (3) in real life, social reality often intrudes upon people’s erroneous beliefs — that is, it provides feedback that permits people to recognize their initial beliefs were wrong and to alter
them accordingly. So, again, we cannot know how flawed the outcome is — the judgment or decision — unless we evaluate its success, accuracy, validity, etc. (which is another thing social scientists emphasizing error and bias do not often do).

The processistic fallacy is a form of overgeneralization. It occurs whenever researchers who demonstrate some error or bias under a very small set of (typically artificial laboratory) conditions unjustifiably assume or conclude that their findings mean that there is widespread human error under naturally occurring conditions (e.g., Cohen, 1981; Funder, 1987). Lab studies are often well-designed to test basic processes but not to generalize results to naturally occurring conditions (Mook, 1983). It is hypothetically possible to appropriately generalize widespread error under naturalistic conditions on the basis of studies revealing flawed processes in the laboratory, but only under conditions that are almost never met. One way to do justify such generalizations is to discover a process so flawed that it must be definitively known to produce pervasive inaccuracy in situations that go well beyond those studied in the lab. For example, the human visual system cannot detect radio waves, so that it is safe to conclude people will be universally inaccurate in their visual assessment of the presence/absence of such waves.

Such demonstrations are few and far between in psychology. A wide range of judgmental and perceptual errors and biases found in laboratory studies have turned out to be functional outside those studies. For example, Gigerenzer & Brighton (2009, p. 107) reviewed evidence showing that: “In contrast to the widely held view that less processing reduces accuracy, the study of heuristics shows that less information, computation, and time can in fact improve accuracy. We review the major progress made so far: (a) the discovery of less-is-more effects; (b) the study of the ecological rationality of heuristics, which examines in which environments a given strategy succeeds or fails, and why…” Such findings should give deep pause to modern researchers who, upon discovering some laboratory bias, leap to the assumption that, therefore, that process undermines accuracy in naturalistic conditions. Regardless, I am not aware of any research that has documented a social perceptual process so flawed that it can be definitively known to produce inaccuracy on purely logical grounds comparable to the radio wave example above.

However, even if such a process were discovered, additional conditions must also be met to generalize from lab studies of biased processes to a conclusion of widespread inaccuracy in life. It must be shown either
that people are *incapable* of overcoming the bias or error by relying on alternative, superior or corrective processes, or, empirically, that, across most of a widely representative array of situations, people both rely on the flawed process and rarely enlist superior or corrective processes. If a program of research engages in a sufficiently large *representative sampling of situations* (Brunswick, 1957; Monin & Oppenheimer, 2014; Westfall, Judd, & Kenny, 2015), and shows that, in most such situations, a flawed process is heavily relied upon and other more appropriate processes are rarely enlisted, inferring widespread error in real life can be justified. Few programs of research, however, meet these standards, whether in social perception or other areas of psychology (e.g., Cohen, 1981; Westfall et al., 2015). This is probably because, as Wells & Windschitl (1999, p. 1115) found, among psychology faculty, there was an “... insensitivity to the need for stimulus sampling except when the problem is made rather obvious.”

**Common flaws in the critiques.** Both commentaries identified potentially flawed processes, and both perspectives are capable of generating *testable (and falsifiable) hypotheses* about potential patterns and sources of stereotype inaccuracy. As such, their perspectives are potentially constructive and generative of new research directions and potentially valuable insights into sources and conditions of stereotype inaccuracy.

Nonetheless, neither Terbeck, nor Bian and Cimpian discussed any research that meet the standards articulated above for concluding that stereotypes *must be* inaccurate on the basis of the supposedly flawed processes that were identified. Neither presented justification for assuming that those supposedly flawed processes *inherently* produce inaccuracy in other judgments (comparable to showing that vision cannot detect radio waves). Thus, we do not *know* that those processes produce inaccuracy.

Neither discussed any program of research that has shown that those processes produce inaccuracy most of the time in a representative sampling of situations. Therefore, it is not knowable from that research whether people commonly rely on those processes, even if they are truly flawed. Last, even if those processes are truly flawed and widely relied upon, neither commentary reviewed any research demonstrating that people rely *exclusively* on supposedly flawed process across situations (even unrepresentative ones). They have not eliminated the possibility that there are other, more appropriate processes that people rely, when arriving at stereotypes. Both critiques therefore, declare stereotypes to be inaccurate on the basis of research incapable of
justifying such a conclusion. Both committed the *processistic fallacy*: over-inferring pervasive (“stereotypes must be inaccurate”) error in real life from laboratory studies of processes that were incapable of generalizing to much of real life. It is of course *possible* that such studies do generalize widely; but that cannot be known without empirical demonstrations that they actually *do* generalize widely. To make these issues more concrete, the specific evidence each discussed is reviewed next.

**Terbeck, categorization, implicit prejudice, and the brain.** Terbeck referred to research showing that: 1. infants and primates categorize, 2. Specific brain areas are associated with face recognition; and 3. Drugs alter scores on the race implicit association test (IAT). This is all fine as far as it goes. Categorization is ubiquitous, thus, this passes the test for a justified generalization to real life. However, categorization is not inherently universally invalid, in the same way that visual detection of radio waves is. People are not wrong for believing that chairs usually have four legs, that Alaska is colder than Arizona, or that men are, on average, taller than women. Thus, the claim that *any particular category* is wrong requires evidence, which Terbeck did not provide.

Similarly, specific brain areas may well be associated with face recognition, but the very term “recognition” implies that, at least some and perhaps most of the time, people are *correctly* distinguishing faces from other features of the stimulus array. It certainly provides no evidence that facial recognition is *wrong*. Finally, I have no doubt that drugs can alter IAT scores. Racial prejudice IAT scores are attitudes, and individuals and societies may deem certain attitudes *morally* good or bad, but attitudes cannot be factually correct or incorrect. It is *possible* that one’s *reasons* for disliking diet soda, the Yankees, and Fred are factually incorrect, but the *attitude* itself cannot be accurate or inaccurate. Thus, all three phenomena identified by Terbeck may lead to *falsifiable hypotheses* about sources of stereotype inaccuracy; but, absent direct data on stereotype accuracy, they do not justify concluding that stereotypes *are* inaccurate.

**Bian and Cimpian and generic beliefs.** Bian and Cimpian’s critique similarly fails to meet the standards necessary to infer widespread naturally occurring error from studies of supposedly flawed processes. Their prototypical cases of supposedly inherently erroneous generic beliefs are those such as “mosquitos carry the West Nile virus” and “ducks lay eggs” (which was the example highlighted in the title of one of the articles
they cite in support of their view: Leslie, Khemlani, & Glucksberg, 2011). They cite evidence that people judge such statements to be true. They argue that this renders people inaccurate because few mosquitos carry West Nile virus and not all ducks lay eggs.

Does agreeing that “mosquitos carry West Nile” mean that we can now assume that people’s beliefs about mosquitos and West Nile are pervasively inaccurate? If these are absolutist beliefs (“all mosquitos carry West Nile”) then they are clearly wrong and no further evidence is needed. SPSR made exactly this point when discussion absolutist stereotypes, which, because of widespread human variation, are almost always invalid. But there is no evidence that generic beliefs are always, necessarily, or widely absolutist.

Perhaps, instead, they capture the phenomenology of distinctive or salient differences between categories. I can only get West Nile from mosquitos, not from moths, mice, or musk ox. Perhaps people agree that “mosquitos carry West Nile” not because they believe “all mosquitos carry West Nile,” but because they believe that “only mosquitos carry West Nile.” Because generic beliefs, as studied, are not inherently inaccurate, the research does not meet the first standard necessary to avoid the processistic fallacy. We cannot assume all generic beliefs are necessarily inaccurate.

It also fails the second standard (even if not inherently inaccurate, is the process empirically found to be generally invalid?). One of the articles cited by Bian and Cimpian (Leslie et al, 2011) found that participants rephrased only 18 of 100 experimenter-provided generic statements as absolutist (“universals” in Leslie et al’s, 2011, terminology). Furthermore, overwhelming majorities (over 90%) most recognized that, in fact, male sheep do not produce milk, male snakes and male ducks do not lay eggs, and so on for nearly all absolute beliefs studied. Thus, one cannot interpret agreement with the generic beliefs as evidence of widespread reliance on an invalid process. The Leslie et al (2011) research did include a wide range of generic beliefs, so it is reasonable to conclude that their results are broadly generalizable to generic beliefs. What is generalizable, however, is that most generic beliefs do not equate to absolutist or inherently inaccurate beliefs.

Of course, it is still possible that when stereotypes are generic beliefs, they are widely inaccurate. That is another falsifiable hypothesis about which there is currently no data. Again, inferring that stereotypes are inaccurate from such data is unjustified.
Bian and Cimpian cite another paper by Leslie (in press) in support of the claim that “more people hold the generic belief that Muslims are terrorists than hold the generic belief that Muslims are female.” But Leslie (in press) provides no data whatsoever that bears on the frequency with which people hold such beliefs. Instead, she quoted headline-seeking politicians and cited a rise in hate crimes post-9/11. Such information may be interesting, but it does not address the frequency of lay beliefs about anything whatsoever.

Of course, even if the claim that more people agree that “Muslims are terrorists” than that “Muslims are women” was valid, it would not constitute evidence stereotypes in general, or the Muslim stereotype in particular, must be inaccurate. Its status as such evidence does not hinge on researcher assumptions about what people mean when they agree with statements like, “Muslims are terrorists” but on evidence assessing what people actually mean. Because research on generics fails the first two tests necessary to avoid the processistic fallacy (they do not inherently produce inaccuracy and they have not been empirically demonstrated to usually produce inaccuracies), one could not conclude that greater agreement that “Muslims are terrorists” than that “Muslims are women” necessarily means people believe there are more Muslim terrorists than Muslim women. It may simply mean “some Muslims are terrorists” or “Muslim terrorism is more widespread than other forms of terrorism” and that “being female is not an important distinguishing characteristic of Muslims.” Absent data, we just do not know. The bias literature writ large (L. J. Cohen, 1981; Gigerenzer & Brighton, 2009; see also Mousavi & Funder’s commentary and SPSR) and the stereotyping literature in particular is so strongly riddled with invalid researcher assumptions about lay people’s beliefs, that, absent hard empirical evidence about what people actually believe, researcher assumptions of inaccuracy that are not backed up by empirical evidence demonstrating widespread inaccuracy rarely warrant credibility.

Bian and Cimpian acknowledge that statistical beliefs are far more capable of being accurate, but then go on to claim that most stereotypes are not statistical beliefs, or, at least, generically based stereotypes are more potent influences on social perceptions. They present no assessment, however, of the relative frequencies with which people’s beliefs about groups are generic versus statistical, and, given Leslie et al’s (2011) evidence that people do not usually translate generics into absolutes, it may well be that agreement with generics such as
“ducks lay eggs” and “Muslims are terrorists” does not preclude the statistical understanding that fewer than half of all ducks are even capable of laying eggs or that the proportion of Muslims who are terrorists is tiny.

We can, however, consider the implications of their claim that most people’s stereotypes include little or no statistical understanding of the distributions of characteristics among groups. This view leads to another falsifiable hypothesis: laypeople would have little idea about racial/ethnic differences in high school or college graduate rates, or about the nonverbal skill differences between men and women, and are clueless about differences in the policy positions held by Democrats and Republicans. That leads to a very simple prediction – that people’s judgments of these distributions would be almost entirely unrelated to the actual distributions; correlations of stereotypes with criteria would be near zero and discrepancy scores would be high. One cannot have it both ways. If people are statistically clueless, then their beliefs should be unrelated to statistical distributions of characteristics among groups. If people’s beliefs do show strong relations to statistical realities, then they cannot be statistically clueless.

We already know that the predictions generated from the “most stereotypes are generic and are therefore statistically clueless” are disconfirmed by the data summarized in SPSR (see also Jussim, Crawford, & Rubinstein, 2015, for an updated review of stereotype accuracy that includes additional studies). Bian and Cimpian have developed compelling descriptions of the processes that they believe should lead people to be inaccurate. In point of empirical fact, however, people have mostly been found to be relatively accurate. Disconfirmation of such predictions can occur for any of several reasons: 1. The processes identified as “causing” inaccuracy do not occur with the frequency that those offering them assume (maybe most stereotypes are not generic); 2. The processes are quite common and do cause inaccuracy, but are mitigated by other countervailing processes that increase accuracy (e.g., adjusting beliefs in response to corrective information); or 3. The processes are common, but, in real life, lead to much higher levels of accuracy than those emphasizing inaccuracy presume (see Mousavi & Funder’s commentary for exactly such a point). Regardless, making declarations about levels of stereotype inaccuracy on the basis of a speculative prediction that some process causes stereotype inaccuracy, rather than on the basis of evidence that directly bears on accuracy, is a classic demonstration of the processistic fallacy.
Confirmation Bias and Questionable Interpretive Practices

Kahan did not disagree with a single claim in SPSR; he did, however, urge me to consider the issues of bias and accuracy more broadly, and I do so here. Kahan correctly pointed out that there is an extensive literature on confirmation bias especially in politicized judgments that SPSR largely ignores. My goal was to evaluate the literature on social perception – how people view other people, especially individuals and groups; and especially with respect to judgments that could conceivably assessed for their accuracy. To compare bias, self-fulfilling prophecy, and accuracy, it was necessary to focus on judgments could be biased, self-fulfilling or accurate. SPSR purposely excluded people’s beliefs about scientific or social science facts or evidence because I do not consider them social perception in the classic sense of “how people understand specific other people or groups.” SPSR also excluded moral and political beliefs because they often have no criteria for assessing accuracy. I concur with Kahan’s view that confirmation biases can be quite powerful with respect to many of these excluded judgments.

Indeed, the very validity of Kahan’s commentary highlights an interesting irony. Exactly the types of confirmation biases in perceptions of science highlighted by Kahan’s commentary may characterize social psychological science. There is ample evidence that scientists’ confirmation biases about research conclusions are demonstrably powerful in at least many cases. Social psychological perspectives that emphasize the power of lay confirmation biases in person perception do so on the basis of a highly selective review of the evidence. Any review reaching the conclusion that the evidence shows that person perception is powerfully characterized by confirmation biases must be based on researcher confirmation bias because the evidence so overwhelmingly shows that lay person perception is mostly motivated by the desire to be accurate (e.g., Devine, Hirt, & Gehrke, 1990; Trope & Bassok, 1983). Chapter’s 5 and 8 addressed this issue at length. With respect to seeking information that bears on their interpersonal expectations, in general, the evidence shows that people overwhelmingly seek and prefer diagnostic, not confirmatory, information.

Kahan’s perspective, however, which focuses a great deal on the role of confirmation biases in how people evaluate science, exquisitely describes the production of social psychological theories of and conclusions about person perception, and many other topics. Other examples consistent with Kahan’s confirmation bias
perspective applied to how psychologists reach conclusions, from *Social Perception and Social Reality*

include:

- Overstated claims about the power of self-fulfilling prophecies
- Overstated claims about expectancy- or stereotype-induced perceptual biases
- Underestimations of the power of accuracy, especially though not exclusively stereotype accuracy, and/or dismissals of its “importance”
- Decades of misinterpretations of studies such as Hastorf & Cantril, 1955 and Rosenhan, 1973 as demonstrating the power of bias, when, in fact, they demonstrated overwhelmingly the power of accuracy

That science *sometimes* goes wrong is a normal part of science. But when science goes off the rails and fails to self-correct for decades, especially when the evidence is sitting in plain daylight from within the original published reports, something other than “pure science” may be going on. Kahan’s work points to some likely possibilities. Kahan’s work helps explain the prevalence of *questionable interpretive practices* (QIPs) – narrative, conceptual, and interpretive means by which scientists can and do reach unjustified conclusions, even in the completely absence of statistical or methodological errors and flaws, and even when *findings* are replicable, (Jussim Crawford, Anglin, Stevens & Duarte, in press; Jussim, Crawford, Anglin, & Stevens, 2015; Jussim, Crawford, Stevens, & Anglin, in press; Jussim, Crawford, Stevens, Anglin, & Duarte, in press). QIPs captured in *SPSR* include:

*Logical incoherence:* Reaching opposite or contradictory conclusions, as long as both advance one’s preferred narratives, values, theory, or ideology. Simple example: Claiming there are no good criteria for assessing the accuracy of stereotypes yet accepting “known groups validity” as a reasonable way to validate new measures.

*Phantom facts:* Declaring something to be a fact without evidence. Simple example from *SPSR:* Declaring stereotypes to be inaccurate without evidence.

*Blind spots:* Overlooking or ignoring research that contests one’s preferred perspective. Simple example from *SPSR:* Citing Darley & Gross’s (1983) single study that they interpreted as showing that stereotypes lead to their own confirmation, and ignoring Baron, Malloy & Albright’s (1995) two failed replications.
**Double standards:** Subjecting the research producing conclusions one dislikes to withering criticisms, and extolling the virtues and value of research producing conclusions one likes, even when the research one dislikes is of equal or higher methodological quality. Simple example: the common claim that there are no “good” criteria for assessing accuracy, while, at the same time, extolling the power of self-fulfilling prophecies. This is a double standard because both accuracy and self-fulfilling prophecies require showing correspondence between belief and reality, so that the criteria for doing so must be identical.

Exposés of major disconnects between accumulated data and common conclusions have been recently published regarding broad areas within cognitive psychology (Firestone & Scholl, 2015), social psychology (Jussim, Crawford, Anglin et al, in press), social neuroscience (Vul, Harris, Winkielman, & Pashler, 2009), and sociology (Martin, 2015). Over a decade ago Pinker (2002) exposed how political motivations led to invalid claims about education, parenting, crime, personality, evolution, and more.

What is going on here? Is it really possible that trained social psychologists, people with PhDs and years of experience, routinely engage in substantial confirmation bias in interpreting scientific research? Many scholarly perspectives answer this question with a clear, “yes indeed” (for general reviews of scientific susceptibility to confirmation bias, see: Greenwald, Pratkanis, Leippe, & Baumgardner, 1986; Ioannidis, 2012; Lilienfeld, 2010). For a review of how confirmation biases have led social psychology to specific unjustified conclusions in areas such as discrimination, stereotype threat, unconscious influences on sensory perception and more, see Jussim, Crawford, Anglin, et al (in press). “Successful” motivated reasoning driven by the goal of reaching some particular conclusion requires information, experience, and skill with logic and argumentation. People with PhDs and extensive training – especially those with training in telling “compelling narratives” (Bem, 2002; Jordan & Zanna, 2007) – are more able to dismiss findings they do not like and defend findings they do like in the face of challenges than are less intelligent and less well-trained laypeople. Indeed, Kahan himself (Kahan et al, 2012) has found that views about climate change become more polarized as people’s science knowledge increases (see also Haidt, 2012).

An even stronger view is presented by Madison, Dutton, & Stern who highlighted scholarship on the clever sillies – which presents a perspective suggesting just how extremely distorted “scholarly” conclusions
can get. Much of that research suggests that social scientists who are obviously very intelligent and have extraordinary levels of knowledge and expertise manifestly silly claims primarily to signal their intelligence (Charlton, 2009; Dutton & van der Linden, 2015). Because manifestly silly ideas are often presented in high-falutin and sophisticated-sounding language, they can appear rigorous and (to paraphrase Stephen Colbert) high in “scientificiness” and, therefore, can create an illusion of plausibility and validity. In the social sciences, such ideas often include the denial of evolutionary or biological bases of human psychology and behavior (see, e.g., Pinker, 2002 for a broad review), the denial of stereotype accuracy, and, I would argue, attempts to stigmatize and ostracize those who point out that the data does not always advance social scientific narratives that are presumed to advance the interests of the oppressed (Gottfredson, 2010; Pinker, 2002).

I do not doubt that desire to signal one’s brilliance may indeed be one motivation underling the clever sillas, but I do not think it is the only one, and, perhaps, not even the most important one in the social sciences. In addition to signaling intelligence, staking out positions that are logically incoherent or disconnected from scientific evidence can signal not just intelligence, but one’s political allegiances, one’s moral positions, and that one is on the “side” of one’s colleagues fighting “the good fight” (e.g., Kahan, in press). The extent to which scientific distortions, such as the denial of stereotype accuracy or evolutionary influences on psychology, result from motivation to signal one’s egalitarian bona fides to one’s colleagues, the desire to advance one’s politics, values, and morals, or other less politicized sources is an important empirical question for the burgeoning area of meta-science and scientific integrity (e.g., Ioannidis, 2012; Jussim, Crawford, Anglin, et al, in press; Simmons et al, 2011)

The Fundamental Publication Error:

Was Planck Right?

“A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it.” Max Planck.

Self-correction is often taken to be a hallmark of science. Whereas religious, political, or moral beliefs may not be subject to change based on evidence, scientific beliefs, presumably, should be subject to change
when sufficient new evidence contradicts existing conclusions. For example, Jost (2011) wrote: “This is because we, as a research community, take seriously the institutionalization of methodological safeguards against experimenter effects and other forms of bias. Any research program that is driven more by ideological axe-grinding than valid insight is doomed to obscurity, because it will not stand up to empirical replication and its flaws will be obvious to scientific peers.”

If only it were so. SPSR presented numerous cases where: 1. An initial high impact “wow!” study yielding some dramatic result was published; 2. Many follow-up studies revealed that the conclusions based on that “wow” study were mostly not justified; and 3. The “wow” conclusions continued to march on for decades as if the correctives were never published. SPSR documented case after case of just this pattern with respect to self-fulfilling prophecies, biases, stereotypes, and accuracy.

Many of the commentaries (Bonnefon, Hopfensitz & De Neys, Kihlstrom, Little, Madison et al, Martin, Mousavi & Funder, Trafimow, Wagoner) seemed to welcome SPSR as a much-needed corrective to the “stupidism” (Kihlstrom) emphasized by much social psychology and the “clever silly” (Madison et al) perspectives that back up such claims. Many of the rest acknowledged the validity of its main points but raised issues beyond the scope of the book (Church, Kahan, Wilson & Huang). If one is to believe the consensus of the commentaries on SPSR, one might believe that the field’s emphasis on “stupidism” is in decline. Although I hope that is true, based on too much evidence from outside these commentaries, such a conclusion is premature, and not only because two of the commentaries committed the processistic fallacy when attempting to defend claims emphasizing lay “stupidism” (Bian & Cimpian, Terbeck) regarding stereotypes.

My collaborators and I have recently updated the review of stereotype accuracy work that appears in SPSR (Jussim, Crawford, Anglin, Chambers, Stevens, & F. Cohen, 2016; Jussim, Crawford, & Rubinstein, 2015). Over 50 studies have been identified, almost double the number reviewed in SPSR because there has been an explosion of research on the accuracy of stereotypes about national character and political groups. The main conclusions of SPSR were reconfirmed, especially regarding the demographic stereotypes that social scientists generally seem most concerned about. Stereotype accuracy is one of the largest effects in all of social
psychology. It has been replicated in multiple independent labs. Given social psychology’s current crisis of replicability, and widespread concerns about questionable research practices (e.g., OSF, 2015; Simmons et al, 2011), one might expect that social psychologists would be shouting to the world that we have actually found a valid, independently replicable, powerful phenomena.

But if one did think that, one could not be more wrong. Testaments to the inaccuracy of stereotypes still dominate textbooks and broad reviews of the stereotyping literature that appear in scholarly books (see Table 1). The new generation of scholars is still being brought up to believe that “stereotypes are inaccurate,” a claim many will undoubtedly take for granted as true, and then promote in their own scholarship. Sometimes, these manifest as definitions of stereotypes as inaccurate; and even when stereotypes are not defined as inaccurate, they manifest as declarations that stereotypes are inaccurate, exaggerated, or overgeneralized.

**Conclusion:**

**Facilitating Self-Correction Regarding Accuracy, Bias, and Self-Fulfilling Prophecies**

Psychology is abuzz with an internal discussion of how it can do better. Greater transparency, pre-registration, replication, and more have all come to the fore. However, most of the unjustified testaments to the power of self-fulfilling prophecies and expectancy or stereotype biases, and most of the attempts to dismiss the power or importance of accuracy did not result primarily from failed replications or questionable statistical or methodological practices, or even lack of transparency. Instead, they are problems of interpretation and (exactly as Kahan’s commentary and perspective might predict) researcher confirmation biases. Even when failed replications did get published, they were generally ignored. Effect sizes were largely ignored. Simple contextual factors (such as the number of plays in a football game, or the total number of judgments made by staff at psychiatric institutions) that could have reigned in overstated claims of bias were often simply ignored, not just by the original researchers, but by decades of subsequent scientists perpetuating the erroneous testaments to bias. Attention to contextual, statistical, and methodological details was seemingly short-circuited by the ability or desire to tell compelling “wow!” stories about the power and pervasiveness of expectancy effects.

What, then, can researchers who want to present valid and nuanced descriptions of the findings do to limit
their vulnerability to perpetuating false claims that appear in scientific literatures? Unfortunately, psychology does not have a consensus on the answers to this question, and is currently in the process of searching for those answers (e.g., Jussim, Crawford, Anglin et al, in press). Here, I focus specifically on: 1. Identifying general principles that may be broadly applicable; and then 2. Give examples of how they could be applied to the literatures addressed by SPSR:

1. **Resist the urge to tell compelling narratives by glossing over or ignoring contradictory findings and conclusions.**
   - Stop citing Rosenthal & Jacobson (1968) as showing that teacher expectation effects are powerful or pervasive.
   - Do not assume that “story studies” -- famous classics around which compelling narratives can be told – are necessarily true or replicable. Review the entire relevant literature before making claims regarding expectancies and stereotypes.
   - Avoid cherry-picking a biased sample of studies about expectancies or stereotypes (or any other topic) to make an argument.

2. **Focus on the actual results of studies, rather than researcher claims about those results.**
   - One can often find evidence of substantial accuracy and rationality in studies that emphasize or only reported error and bias.
   - Biases and self-fulfilling prophecies may be quite modest, or even contingent on moderators, even when the discussion touts their power and pervasiveness.

3. **Search for skeptical reviews and meta-analyses, and do not depend exclusively on reviews or meta-analyses that appear to have as an agenda persuading the reader.** Avoid repeating conclusions based on famous reviews, without either critically examining the basis for those conclusions, or, at least, searching the literature to find out whether other, perhaps less famous but more persuasive, skeptical or critical reviews or meta-analyses have reached different conclusions. Abide by the Mertonian (1942/1973) norm of universalism, that evaluation of scientific claims hinges **not at all** on the status or prestige of the scientist making them, but on the quality of the
evidence, logic, and argument being put forth.

- For every review testifying to the power of expectancies, there are now others casting doubt on such conclusions. If one must make a point about expectancies, at minimum, one can reflect the state of the literature with statements such as:
  - “Whereas some reviews have concluded that expectancy effects are powerful and pervasive, others have concluded that such effects are weak, fragile, and fleeting.”
  - “Although stereotypes have long been presumed to be inaccurate, several reviews have concluded that, in general, stereotypes are often at least moderately accurate.”
  - “Although social constructionist phenomena undoubtedly occur and can sometimes be powerful and important, at the level of individuals interacting with other individuals, such effects are usually quite modest.”
  - “Although people undoubtedly cognitively construct their social perceptual worlds to a considerable degree, and, sometimes such constructions can be quite biased, this does not mean their constructions are always or even mostly inaccurate.”

4. **In new original studies, be excessively transparent about methods and results.** Provide means, standard deviations, and correlations for all variables. When available and relevant, provide frequency distributions. When reporting regression and SEM results, report standardized and unstandardized coefficients and also the t- and F-values associated with each test of significance. If all this cannot make it into the main report, then at least provide it in supplementary materials. Report effect sizes and confidence intervals. This should be done when reporting new empirical studies; and it should be routine when reviewing empirical literatures.

- This is especially important when making claims about the relative power of bias versus accuracy. Distorted claims about bias could have been detected decades earlier, if, e.g., effect sizes had been routinely reported, and if contextual data (e.g., total number of judgments) had been reported.

5. **Be careful about definitions.** Researchers have great latitude in how they define constructs, but
then have to own the implications of their definitions.

- If one defines stereotypes as inaccurate or as exaggerations, then one must be willing to accept that only beliefs about groups that have been demonstrated to be inaccurate and exaggerations among the sample one is studying can be known to be stereotypes.
- One can avoid this problem by defining stereotypes in ways that permit them to be accurate, avoiding presumptions of inaccuracy, exaggeration, or overgeneralization.

6. **Base empirical claims about the state of the world on actual empirical evidence.**

- This may seem obvious, but researchers have been making claims about stereotype inaccuracy without evidence for decades. See Pinker (2002) for similar claims without evidence regarding a range of issues, such as human malleability and the role of social factors in everything from intelligence to aggression to sex differences.
- Avoid the processistic fallacy. Do not make claims about error, bias, or the inaccuracy of stereotypes on the basis of process studies, even ones that identify faulty processes in the lab that one speculatively presumes will cause inaccuracy in people’s naturally-occurring judgments. Such processes might have theoretical import (Mook, 1983), and they might generate predictions regarding patterns or sources of inaccuracy. But they rarely, if ever, constitute evidence of inaccuracy.
- Reach conclusions about stereotype accuracy on the basis of studies reporting empirical data rather than sources (even “authoritative” ones such as G. W. Allport, 1954/1979; see also Table 1) declaring stereotypes to be inaccurate (or exaggerations) without data.
- Do not claim that characterizing stereotypes as possessing a “kernel of truth” constitutes some sort of acknowledgement that stereotypes are often substantially accurate. This functions as a disingenuous attempt to maintain the emphasis on inaccuracy, which can readily be seen with a “turnabout test” (Duarte, Crawford, Stern, Haidt, Jussim, & Tetlock, 2015; Tetlock, 1994): Would declaring, “Psychological research has a kernel of truth” be a great testament to the validity of psychological science?
• If stereotypes do influence judgments regarding an individual target do not assume that increases inaccuracy without testing for accuracy.

7. **Build rational judgment processes into theoretical perspectives on social perception.**

• Because of social psychology’s infatuation with error and bias, almost any result, no matter how reasonable and rational, has been framed as flawed. However, such conclusions regarding lay judgments require showing that some particular perceptual result deviates from some normative model. In social psychology, this is rarely done, thereby liberating researchers to cast almost any result as irrational.

• Social psychologists should stop casting results as irrational absent development of normative model of rational judgment and an assessment of the extent to which lay judgments both correspond to and deviate from that model.

• Social psychologists studying social perception should start developing models of rational judgment processes if they wish to continue reaching judgments about irrationality.

8. **Be clear and consistent with respect to levels of analysis.**

• If one is discussing perceptions of groups, then accuracy refers to correspondence between beliefs about groups and what those groups are like.

• If one is discussing perceptions of individuals, then accuracy refers to correspondence between beliefs about an individual and what that individual is like.

• Cease confounding levels of analysis by declaring that stereotypes are inaccurate because they do not apply to every individual.

Science can tolerate errors, even a great many errors, if it also has strong and largely successful and efficient mechanisms for self-correction. In this spirit, it is worth pointing out that none of the commentaries, not even those few who most strongly disagreed with my conclusions, presented any data showing that self-fulfilling prophecies or expectancy-based biases were generally large, or that stereotypes were generally inaccurate. The strongest arguments for modifying the conclusions reached in *SPSR*, in my view, came from those suggesting that the emphasis on accuracy and the de-emphasis of bias and self-fulfilling prophecy might
not be quite so applicable beyond the specific types of interpersonal contexts addressed in *SPSR* (Kahan, Kihlstrom, Tappan, Wilson & Huang). Perhaps, therefore, we can agree that, even if *SPSR* does not spell the death knell for social or cognitive constructivism, with respect to the topics that it has addressed – teacher expectations, person perception, beliefs about groups and how those beliefs influence social perceptions – a little scientific self-correction is overdue.
References


Jussim, L., Crawford, J. T., Anglin, S. M., Stevens, S. M., & Duarte, J. L. (In press). Interpretations and
methods: Towards a more effectively self-correcting social psychology. *Journal of Experimental Social Psychology.*


Figure 1: The Reflection-Construction Model (Jussim, 1991)

**Figure 1A: The Full Model**

- **Background Information** (anything on which perceivers may base their expectations)
  - Accuracy
  - Sources of expectations

- **Interpersonal Expectations**
  - Self-fulfilling Prophecy

- **Targets’ Behavior or Attributes**
  - Impression accuracy
  - Bias
  - Perceivers’ Judgments of Targets

**Figure 1B: Constructive Accuracy**

- **Background Information**
  - Predictive accuracy
  - Sources of expectations

- **Interpersonal Expectations**

- **Targets’ Behavior or Attributes**
  - Bias
  - Perceivers’ Judgments of Targets

Constructive accuracy: Even when perceivers are completely oblivious to targets’ behavior or attributes, their judgments of targets’ will still correspond to (correlate with) targets behavior or attributes if 1. expectations are based on background information that 2. predicts targets behavior or attributes; and if 3. expectations influence (bias) perceiver judgments.
<table>
<thead>
<tr>
<th></th>
<th>Explicitly acknowledges strong evidence of stereotype accuracy</th>
<th>Reviews little or no evidence of accuracy and either dismisses accuracy as unimportant or emphasizes stereotype inaccuracy and bias</th>
<th>Defines/declares stereotypes to be inaccurate</th>
<th>Representative Quotes:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Scholarly Books</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Banaji &amp; Greenwald, 2013, <em>Blindspot: The hidden biases of good people</em></td>
<td>✗</td>
<td></td>
<td>P. 74: Because all stereotypes are partly true and partly false, it may seem pointless to debate their accuracy. P. 89: … stereotyping is an unfortunate by-product of the otherwise immensely useful human ability to conceive the world in terms of categories.</td>
<td></td>
</tr>
<tr>
<td>Brown, 2011, <em>Prejudice: Its Social Psychology</em></td>
<td>✗</td>
<td></td>
<td>P. 71: … the question of whether stereotypes are ‘objectively’ (in)accurate is only of marginal interest to most students of prejudice.</td>
<td></td>
</tr>
<tr>
<td>Fiske &amp; Taylor, 2008, <em>Social cognition: From brains to culture</em></td>
<td>✗</td>
<td></td>
<td>P. 282: Stereotyping is the cognitive aspect of bias … and it comes in both blatant and subtle forms.</td>
<td></td>
</tr>
<tr>
<td>Whitley &amp; Kite, 2009, <em>The psychology of prejudice and discrimination</em></td>
<td>✗</td>
<td></td>
<td>P. 100: At the group level, then, stereotypes may have a kernel of truth, but relying on them at the individual level may lead to serious judgment errors</td>
<td></td>
</tr>
<tr>
<td><strong>Textbooks</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aronson, 2011, <em>The social animal</em></td>
<td>✗</td>
<td></td>
<td>P. 309: 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</td>
<td></td>
</tr>
<tr>
<td>Baumeister &amp; Bushman, 2014, <em>Social psychology and human nature</em></td>
<td>✗</td>
<td></td>
<td>The high level of accuracy in modern stereotypes may also indicate that stereotyping has changed.</td>
<td></td>
</tr>
<tr>
<td>Author(s)</td>
<td>Title</td>
<td>Page</td>
<td>Note</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>-------</td>
<td>------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>Crisp &amp; Turner, 2014</td>
<td>Essential social psychology</td>
<td>57</td>
<td>P. 57: Once a category is activated we tend to see members as possessing all the traits associated with the stereotype.</td>
<td></td>
</tr>
<tr>
<td>Greenberg, Schmader, Arndt, &amp; Landau, 2015</td>
<td>Social psychology: The science of everyday life</td>
<td>352</td>
<td>p. 352: Even though this kernel [of truth] might be quite small, with much more overlap between groups than there are differences, as perceivers we tend to exaggerate any differences that might exist and apply them to all members of the group.</td>
<td></td>
</tr>
<tr>
<td>Grison, Heatherton, &amp; Gazzaniga, 2015</td>
<td>Psychology in your life</td>
<td>385</td>
<td>P. 385: Indeed, some stereotypes are based in truth: Men tend to be more violent than women, and women tend to be more nurturing than men. However, these statements are true on average.</td>
<td></td>
</tr>
<tr>
<td>King, 2013</td>
<td>Experiencing psychology</td>
<td>402</td>
<td>p. 402: A stereotype is a generalization about a group’s characteristics that does not consider any variations from one individual to another.</td>
<td></td>
</tr>
<tr>
<td>Schachter, Gilbert, Wegner, &amp; Nock, 2015</td>
<td>Introducing psychology</td>
<td>403</td>
<td>P. 403: … stereotyping is a useful process that often produces harmful results, and it does so because stereotypes have four properties: They can be (1) inaccurate, (2) overused, (3) self-perpetuating, and (4) unconscious and automatic.</td>
<td></td>
</tr>
</tbody>
</table>

Table 1 reprinted from Jussim et al, 2015.
Footnote

1 Strictly speaking, the 14th amendment and Civil Rights acts focus on behaviors (discrimination), rather than beliefs. However, failure to provide Alfonso service because the provider believes Latinos are mostly criminals is a violation of those acts; however, failure to provide Alfonso service because the provider believes Alfonso is a criminal is not. Whether the behavior is based on a stereotype or on individuating information is taken as extremely important, thereby highlighting the perceived value of the distinction in legal contexts.