

3 Understanding Inequities: The Role of Schemas

“I don’t care who they are; I just want the best person.”

Perhaps the most common statement we hear from faculty and administrators is their commitment to the merit principle. Everyone affirms that people’s advancement should be based purely on their merit, not their gender, color, sexual orientation, religion, class, or any other feature that is irrelevant to merit. Hiring and advancement should be based solely on excellence. We agree.

In this chapter we introduce the notion of schemas (especially gender and race schemas) and the role they play in our perceptions and evaluations of people (for review, see Valian, 1998).¹ In brief, gender and race schemas result in our slightly undervaluing the accomplishments and competence of women and people of color, and slightly overvaluing the accomplishments and competence of White men. Many of the examples we discuss are small and can occur outside awareness. We note the covert nature of many of our perceptions and discuss the importance of even small examples of erroneous judgments. Schemas, and their effect on the policies and procedures that affect hiring, retention, and promotion, explain how, despite good intentions, colleges and universities end up hiring and promoting fewer women and underrepresented minorities than would be expected on the basis of the available pool.

We agree that decisions should be based solely on excellence. We disagree when people tell us that the social group someone belongs to plays no role in their evaluations. We think the data are clear that someone’s social group does play a role, overall, whether people think it does or not. We say “overall” because many factors affect evaluations. On any given occasion, it can be hard to know what the effect of someone’s social group is.

Three Mistakes in Judgment That People Make

We think people make three mistakes (at least!) about their judgments.

Mistake 1: People Think That Their Explicit Intentions to Be Fair Completely Determine Their Behavior

People incorrectly believe that their explicit intentions completely guide their evaluations and observations of others. People think that consciously saying and believing something—in the area of evaluation—make it so. At least, they think it is so for *them*: *they* are not biased, or, even if they are slightly biased, they do not behave in a biased way. They have a blind spot where their own behavior is concerned (Pronin, Lin, & Ross, 2002; West, Meserve, & Stanovich, 2012). Nor are people aware of the studies we describe in this chapter that indicate that good intentions are not enough. Those studies suggest that all of us at least occasionally make decisions influenced by the gender, race, ethnicity, and other characteristics of the person we are assessing. Perhaps the worst aspect of our belief in our good intentions is that the very belief that we can evaluate people independently of the group to which they belong can actually make it easier to judge people on the basis of those irrelevant characteristics. We present data on belief in one's own good intentions.

Mistake 2: People Think That Judging Merit Is Straightforward—At Least for Them

People see the criteria they are using as valid on the face of it and transparent, even though their criteria may be somewhat idiosyncratic. They do not consider that some of their criteria may be subjective nor that their criteria might shift so that they choose someone who seems like a “good fit”—that is, someone who will fit in well with *them*, as we described when discussing homophily. People also seldom ask whether their criteria inadvertently exclude people whose contributions they would value. Finally, people do not see that their own behavior may affect how others display their merit. We present data on judgments of merit.

Mistake 3: People Do Not Understand How Success Happens

People associate success with banner-headline achievements, tending not to see the slow accumulation of small successes that make large achievements possible. They also fail to see how a steady drumbeat of

small bits of praise or criticism can encourage or suppress people's efforts to be successful in a given field. If, more often than not, you receive praise for doing a good job, only profound modesty will prevent you from concluding that you are capable and talented. Conversely, if, more often than not, your efforts are either ignored or criticized, only profound aplomb will protect you from concluding that your aspirations are too high.

Data Examined in This Chapter

In this chapter we draw on experimental data—some that were collected in the field and some that were conducted in the laboratory—to determine whether and how people get advantages on the basis of irrelevant characteristics. The data show that Whites are advantaged compared to people of color, and that men are advantaged compared to women, especially in domains related to work. Those asymmetries in advantage occur even if others think they are judging on the basis of merit alone.

Although we have just spoken of “Whites” and “people of color,” of “women” and “men,” we realize that people fit multiple categories simultaneously, and that the groups are large and very heterogeneous. In short, we realize that there are subtleties that the experimental data we report do not always capture and that experiments have both strengths and limitations. Our general strategy is to use both experimental and observational data. When the two converge in their findings, we are more confident that the phenomena we are examining are strong. In this chapter we concentrate on experiments. In chapter 4 we concentrate on observational data—data on numbers and percentages of men, women, and people of color in different types of academic institutions, and data on salaries.

Three Studies to Set the Stage

The three studies we begin with all use real-life events, such as employer responses to people applying for a job, faculty responses to people requesting an appointment, and bus driver responses to people with insufficient fare.

Study 1

Excellent credentials do not fully offset the disadvantages that African Americans experience in job hunting. That was tested by submitting identical job applications to employers that differed in the kind of first name the applicant had. The names Lamar or Ebony suggest a different race than the names Ronny or Anne. But although prospective employers pay attention to the prestige of a candidate's degree, a prestigious degree (from, say, Harvard) for an African American is similar in effect to a less prestigious degree (from, say, the University of Massachusetts at Amherst) for a White (Gaddis, 2015). Lamar from Harvard receives the same number of responses as Ronny from UMass (13% of the time). African Americans with *low*-prestige degrees received a response only 7% of the time.

When applying for jobs, candidates submit their dossiers and hope for the best. Prospective employers use various cues to ethnicity and gender (consciously or nonconsciously), along with information about job-relevant characteristics, and judge accordingly.

Social class matters too, though there is less research on it, and our first example here does not concern employment. Mental health practitioners, in another audit study (Kugelmass, 2016), replied positively to ostensibly middle-class people who requested an appointment on average 23% of the time but replied positively to ostensible working-class people only 8% of the time. Some of those class effects may be driven by a reluctance to deal with the bureaucracy of Medicaid (Asplin, Rhodes, Levy, Lurie, Crain, Carlin, & Kellermann, 2005). Within the group of apparently middle-class people, African Americans received an offer of an appointment 17% of the time, while Whites received an appointment 28% of the time. Another suggestion that race is more important than class comes from the finding that, within race, names associated with higher social class are likely to receive as many invitations for a job interview as names associated with lower social class (Bertrand & Mullainathan, 2004).

Study 2

Lest faculty think that they are immune to cues about race, ethnicity, and gender, they can absorb the sobering results of a study of 6,548 professors showing a preference for Whites (Milkman, Akinola, & Chugh, 2012). Fictitious e-mails, purportedly from prospective doctoral students, were sent under different names. Some names sounded Caucasian, some African

American, some Hispanic, some Indian, and some Chinese. Some were from women and some were from men. In one condition, the prospective students asked for a meeting that same day; in the other they asked for a meeting in a week. The e-mails were otherwise identical.

The experimenters predicted little or no preferences on the basis of race, ethnicity, or gender for a same-day request, on the grounds that the faculty member would think primarily about whether they had time in their schedule. As predicted, faculty treated different groups the same when they were asked for a same-day meeting. For the more distant request, the experimenters predicted a preference for White males, on the grounds that the faculty member would probably try to control how they allocated their time a week hence and could thus think about whether they actually wanted to meet the person. That prediction was confirmed.

Note that in both Study 1 and Study 2, prospective employers or faculty had no information about the applicant other than what was provided in the applicant's dossier or e-mail. Thus, they had no additional information that gave them a sense of what the person was like, information that will be particular to that person. That information is sometimes called individuating information. Many real-life situations are exactly like the conditions in Studies 1 and 2: evaluators have little to go on, so they rely on preconceptions that they would deny that they have. Study 3 examined what happened when observers had some individuating information.

Study 3

In a field study of bus drivers in Australia, the drivers actually saw an individual who boarded the bus and said they did not have enough fare for the trip they wanted to take (Mujcic & Frijters, 2013). White bus drivers gave a free ride to Whites twice as often as they gave a free ride to Blacks (76% vs. 38%). Drivers gave both Whites and Blacks more free rides when they were dressed in business clothes instead of casual clothes, but they still gave Whites more free rides. Thus, business clothes offset some of the disadvantage Blacks face but did not eliminate the cost of being Black. Here, too, although the drivers saw the individuals, they had only cursory information about them, as is common in real life. There are many cases where one gives or withholds opportunities, politeness, and kindness on the basis of minimal cues.

What if You Have Even More Information about the Person? Gender Still Matters

The cautious reader will note that we often have richer information about people than that presented in the previous studies. So we might think that if an observer has enough information to get a full idea of the person, then the person's gender, race, and ethnicity would be relatively unimportant. People would be judged on the basis of who they are.² And, indeed, "stereotypes typically have no ... effects when the individual is also known to have engaged in an unambiguous behavior that is clearly diagnostic of the trait in question" (Kunda & Thagard, 1996, p. 291). The problem is that in most situations in academia (and in many situations in life more generally), we do not encounter unambiguous behaviors that are "clearly diagnostic of the trait in question." We are usually faced with ambiguous behaviors that can be interpreted in multiple ways.

Research on parents' impressions of their children and teachers' impression of their students give two examples showing that even when people have a great deal of individuating information, they are still affected by their beliefs about gender differences.³

The simplest example comes from a study of mothers' impressions of 11-month-old infants' crawling abilities. There are no gender differences in early motor development. Crawling and walking, for example, occur at similar ages in boys and girls and develop in a similar way (see Mondschein, Adolph, & Tamis-LeMonda, 2000, for a review).⁴ During the toddler years, boys' motor skills outshine girls', but in infancy the two are equal.

Infants' mothers offered estimates of their children's crawling ability and estimates of how well the infants would navigate a sloped walkway (Mondschein et al., 2000). Despite their personal experience with their children's behavior, mothers of boys overestimated and mothers of girls underestimated how well their children would perform on the slope. Boys and girls made the same number of attempts, they navigated the same degree of slope, and they were equally unlikely to attempt slopes that were too steep for them. Yet mothers of boys thought their sons would succeed when the likelihood of success was actually 0%, and mothers of girls thought their daughters would fail when the likelihood of success was actually 100%.

Individuating information may nevertheless be having an effect on mothers. The experiment did not test how adults without children would

rate infant girls and boys. Mothers may be more accurate than nonparents, while still being inaccurate along the lines that schemas would predict. The information they have may reduce, even though it does not eliminate, the impact of schemas.

The cautious reader may bring up another point: not everyone responds the same. We are reporting group averages. People vary from one to another, and within themselves from one occasion to another. For that reason researchers try to have large samples, so that the variability will not obscure the regularities. Even though there is variability, there are also average differences. Men are on average taller than women, for example. That phenomenon does not vanish because there are some very short men and some very tall women. Similarly, we tend to judge men and women, boys and girls, differently, even though we do not judge them differently on every occasion. One of us has often enough observed herself making judgments favoring White men to think that the experiments apply to her as well as to participants in experiments. See box 3.1 for one example.

The parents in the study we described spent months viewing their children's crawling and were likely motivated to think the best of their children. Nevertheless, their children's gender influenced evaluations of their children's abilities. Parents' beliefs about what children are capable of in turn influences what parents expect of their children; that in turn influences what children think they can do. Parents want to, and believe they do, assess their children simply as individuals. And in many areas they

Box 3.1

Gender Enters Perception, but Effort Can Prevent Its Influence on Behavior

One of us is embarrassed to report the following example:

I walked into a computer store in search of an arcane peripheral. As I approached the relevant department I saw two salespeople, a young man and a young woman, standing and chatting together. I had to stop myself from addressing only the man. I had to force myself to look equally at both salespeople as I asked my question. I could see that I thought the man was likely to be more knowledgeable about computers than the woman, even though they were both working in the same department. Had it not been for my knowledge of the literature, I probably would have been unaware of my assumption that the man would be more knowledgeable and I would have focused my attention on him without thinking about it. (Neither of them, it turned out, had any knowledge of the arcane peripheral—but it did take the man quite some time to say so.)

do (see Valian, 1998, for a review). But not all. In the case of motor development, parents may be basing their judgment about different motoric skills on differences that will eventually distinguish boys and girls. Parents' knowledge about general gender differences makes it difficult for them to see their children neutrally, even if they have individuating information about their children that would support correct estimates of their abilities.

Crawling might not matter much. But math and science matter a great deal. If a little girl is an astounding whiz at math—and assuming that she has had an opportunity to demonstrate and develop her proficiency—she will probably correctly be perceived by her parents and teachers as a whiz. In that sort of case, the girl's outstanding characteristics may outweigh any beliefs about gender differences in math that her parents and teachers have. (The same holds for adults. People who are astounding and outstanding are likely to be perceived as such. Most of the people in any population, however, are by definition not astounding and outstanding.) But in the case in which the girl is simply very good at math, she may not be exceptional enough to override adults' beliefs that girls do not excel at math. A boy who is equivalently good will be recognized because he confirms his parents' beliefs that boys do excel at math.

Data on White middle-class parents in both the United States (Yee & Eccles, 1988) and Finland (Räty, Vänskä, Kasanen, & Kärkkäinen, 2002) are suggestive. Parents view their young sons as better at math than their daughters (which they may be, but school grades show an advantage for girls, Voyer & Voyer, 2014). And parents are more likely to see their sons as excelling because of their ability. When explaining the performance of children whom they had rated as highly competent at either math *or* reading, parents tended to understand the children's achievements differently. Parents were more likely to attribute boys' very high competence to talent, but ascribe girls' very high competence to effort (Räty et al., 2002). Thus, boys are talented and girls work hard. Regardless of the domain, parents see boys' and girls' academic excellence differently, despite the large amount of individuating information they have. Similarly, as we describe later, college students are more likely to view their male (vs. their female) professors as brilliant.

Girls do in fact possess better work habits than boys, habits like paying attention, working hard, and being self-disciplined (Duckworth & Seligman,

2006; Lubienski, Robinson, Crane, & Ganley, 2013). However, parents and teachers appear to be overinfluenced by girls' good work habits when they rate children's abilities. They seem to see talent and effort as inversely related to each other, even though they are compatible (Dweck, 1999). Parents are fond of talking about how different each of their children is, without considering the extent to which their treatment of their children, or subtle ways that they talk to their children, might have contributed to those differences.

To summarize the results about individuating information for gender: people attend to individuating information, but they attend differently, depending on the sex of the child or person who is displaying the behavior (Chan & Mendelsohn, 2010) and depending on the nature of the information (Kunda & Thagard, 1996). Gender can play a role even when we are evaluating individuals about whom we have a great deal of information. Nevertheless, individuating information—of the right kind and in the right amount—will serve to reduce inaccuracies based on schemas. It thus behooves evaluators to seek as much diagnostic information as possible, and it behooves those who are being evaluated to provide as much diagnostic information as possible.⁵

In the case of race and ethnicity, the data are more equivocal than the data for gender, with more work suggesting that individuating information overrides beliefs about race and ethnicity for some types of judgments (e.g., Chan & Mendelsohn, 2010). One difference between gender and ethnicity is that men and women are equally represented in the population, have had many interactions with each other, and dress and behave in ways that differentiate them on the basis of their gender. The same is not true for race or ethnicity.

We conjecture that it is difficult to counter gender in part because our view about gender differences are exceptionally detailed. If, for example, we ask people who is more likely to take baths, men or women, people answer, "women," without hesitation, even if their personal experience is very limited. And the question seems natural. But if we ask people who is more likely to take baths, African Americans or Whites, people are likely to say "Huh?" The very question seems odd. Gender pervades every area of existence, but race and ethnicity do not. Even in areas where there are known differences, people are unlikely to be able to say more than what they themselves do

and extrapolate from there. Thus, if asked who is more likely to refrigerate their ketchup, Whites or African Americans, people will answer on the basis of what they themselves do. (It appears that Whites in general refrigerate their ketchup more than African Americans do; Page, 2007.)

Schemas—How We Evaluate Others

How can we understand our judgments about people? We emphasize the role of schemas.⁶ Schemas are hypotheses that we use to interpret people and social events (Fiske & Taylor, 1991). Schemas are cognitive constructs that represent knowledge and beliefs about social groups and roles (such as scientist). They can be accurate and predictive (Jussim, 2017; McCauley, Stitt, & Segal, 1980). Gender schemas are hypotheses about what it means to be male or female; race and ethnicity schemas are hypotheses about what it means to be a particular race or ethnicity. In our use of the term, schemas are similar to stereotypes. We prefer the term “schema” because it is broader and more neutral. The term “stereotype” often carries with it the idea of a negative evaluation (though there are positive as well as negative stereotypes) while the term “schema” does not. All stereotypes are schemas, but not all schemas are stereotypes.

The term “schemas” brings out the protoscientific nature of the social hypotheses that we all share. Categorizing objects and events helps us keep track of them and understand them. We have schemas about everything, including inanimate objects. We have a schema for what a chemistry lab looks like.

Why We Have Schemas

We need schemas to make sense of our world. When applied to people, they help us have expectations of others, predict others’ behavior, and orient our own behavior. Schemas do some valuable work for us. For example, if we twist our ankle and someone in medical scrubs walks by, we are more likely to ask them for help than another random stranger. The person may not be a doctor, but our schema about doctors will be invoked. We can’t respond to every person we see as if we know nothing about the social groups to which they belong. Schemas are at least partially diagnostic: they do tell us something about people. Students have schemas about professors: faculty are knowledgeable about their subject matter, they come to class prepared

and on time, they write exams that aren't "tricky," and they grade fairly. When professors violate any of those expectations, students complain. Schemas can thus be both descriptive and prescriptive—they describe what is generally so and they describe what should be so.

The Role of Physical Appearance

People's physical appearance helps us categorize them. For example, when either of us talks to university audiences, we have a pretty good chance of picking out the upper level administrators, just on the basis of their dress and hair. Administrators on the whole dress in a more corporate and formal style than most academics; they look as if they spent time and money to create their appearance. Their garments are coordinated and pressed, their hair is tidy, and they are well-groomed. The men tend to wear ties and suits. The women tend to wear skirt suits more than pants suits, but either way there are a lot of jackets; they wear scarves and jewelry, especially necklaces and earrings. Those cues do not *always* identify upper level administrators. There are exceptions in both directions. But they work often enough for us to keep on categorizing. People may use cues like dress without being aware of them. The fact that such cues are not part of people's conscious awareness does not prevent people from using them. In general, much of the content of schemas is nonconscious.

The physical differences between men and women help underpin the idea of a psychologically "nurturant" woman: women, unlike men, can literally breastfeed an infant (see Valian, 1998, for a hypothesis about the cognitive basis of gender schemas). Similarly, women's lesser physical strength can help underpin the idea of women as "the weaker sex," not only physically but in other domains as well. There is also some evidence that women who are less feminine looking are viewed as more competent than women who appear more feminine (Heilman & Stopeck, 1985). The physical differences among different ethnicities, in contrast, have more limited extensions. The physical characteristics of Jewish individuals, for example, have no relation to the contents of schemas about Jews. It is possible, however, to create an arbitrary link between visual cues and beliefs about personality, behavior, and intelligence. People can, for example, come to associate how Asians "look" with math and science ability, even though there is nothing in how Asians look that directly supports the association.

At the negative end, both African Americans and Whites see African American men who look more “African-American” (e.g., by having darker skin) more negatively than they perceive African American men who are less racially phenotypical (Maddox, 2004). African Americans who “look” African American receive harsher sentences for the same crimes than do African Americans who “look” less African American (e.g., Burch, 2015). (In connection with our earlier statement that stereotypes are neutralized in the presence of a clearly diagnostic event, it appears that commission of a crime need not be seen as clearly diagnostic.) Non-African Americans are also more likely to feel threatened when interacting with African Americans compared to non-African Americans (Mendes, Blascovich, Lickel, & Hunter, 2002). The perceptions extend to school personnel’s observations of children: African American children receive more punishment in school than White children of equivalent social class who show similar behaviors (Skiba, Michael, Nardo, & Peterson, 2002; Skiba, Horner, Chung, Rausch, May, & Tobin, 2011). Immigrant workers with lighter skin tones earn higher wages than those with darker skin tones, even when controlling for education, proficiency in English, and other relevant factors (Hersch, 2008).

Although schema formation is ubiquitous and helpful *overall*, schemas are a form of “fast” thinking (Kahneman, 2011) that can lead to error and, when applied to people, can lead to inappropriate perception and treatment of others. Once we know the results of experiments demonstrating the ways in which schemas lead to errors, we can sometimes catch ourselves in the act of drawing inappropriate conclusions.

Are Schemas Shared between Groups?

In the case of gender, men and women are similar in how they see the sexes. The content of male schemas, as we describe below, is shared by both men and women, as is the content of female schemas. That is probably a function of the fact that men and women grow up together as part of the same community. In the case of race and ethnicity, however, the groups in question sometimes make similar judgments and sometimes differ. For example, while the schemas of African Americans and Whites about Whites include some elements in common (e.g., “good at school”), they may not share others (e.g., “likely to be racist”; see also Conley, Rabinowitz, & Rabow, 2010).

Similarly, African Americans and Whites do not necessarily have identical schemas about African Americans. Studies of Hispanics and Asian Americans

similarly find conflicting evidence, but there is some evidence that members of both groups internalize—to some extent—the views of themselves that Whites have of them (Jones, 2001; Pittinsky, Shih, & Ambady, 1999; Steele, 2011).⁷

The Available Data and Our Account

We are concerned in this book primarily with people in professional life. Scholars have investigated the objective performance of women and men as well as women's and men's perceptions of their performance. That research is both quantitative and qualitative and includes both objective measures and reports of subjective experiences. Considerably less research exists on the professional performance of different races and ethnic groups and on perceptions of their performance. Nevertheless, the accounts that we offer to explain the lower prestige and power of women and of non-Whites are similar in four ways.

1. People perceive others in terms of surface characteristics that are often unrelated to actual achievement.
2. Surface characteristics activate schema-based perceptions and evaluations.
3. Schemas allow quick decisions but also result in inaccurate judgments and the unintentional advantaging of one group over another in professional settings.
 - Some of the different treatments and evaluations are major, such as failures to hire or promote.
 - Many of the different treatments are small, daily incivilities (Cortina, Magley, Williams, & Langhout, 2001) that may not be noticed or may be seen as unimportant by either the person experiencing them or the person who engages in them.
 - Disadvantages accrue to women and people of color more often than to White men.
4. Different treatments and evaluations, including the ones that are small and hard to detect, not only directly create different outcomes when they should not, but can create different performances.

Schemas have a broad range of consequences, some of which we describe in this chapter, and some of which we describe in our discussions of hiring and promotion in part II.

Stereotype Threat

One consequence for people of being seen as not having the abilities that are necessary for success is lower performance. When people's knowledge of a negative stereotype about them is activated, and the stereotype is in an area that is important to them, their performance suffers. That phenomenon is called stereotype threat. Women for whom math is important perform less well on a math test when they are told that women typically perform worse on it than when they are told that there are no gender differences (Spencer, Steele, & Quinn, 1999). Tests of intellectual performance also pose stereotype threat to African Americans and Hispanics who value academic achievement, reducing their performance (see reviews in Spencer, Logel, & Davies, 2016; Steele, 2011).⁸

Part of the insidiousness of stereotype threat is that it can lead to confirmation and amplification of the contents of the stereotype. If women or minorities who are susceptible to stereotype threat perform poorly, that confirms their lack of ability both to others and to themselves. Expectations do not need to be directly communicated to people to impair their performance.

Gender and ethnicity work somewhat differently, in that subtle cues lead to larger effects for women in math than do more blatant cues, while the reverse is true for ethnicity (Nguyen & Ryan, 2008). Gender and ethnicity also combine to produce complex effects (Gibson, Losee, & Vitiello, 2014; Pittinsky et al., 1999; but see Moon & Roeder, 2014). For example, if Asian women are primed with their ethnicity (i.e., given a cue that brings their ethnicity to mind), they tend to perform above expectation (stereotype "lift"), but if they are primed with their gender, they tend to perform below expectation (stereotype threat).

Explanations for the effects invoke three different underlying mechanisms, all of which consume cognitive resources that would otherwise be used to perform at capacity (Schmader, Johns, & Forbes, 2008). There is a physiological stress reaction, a tendency to think about one's performance as one is doing it, and an attempt to suppress negative thoughts (Murphy, Steele, & Gross, 2007). All of those processes take attention and resources away from the task itself, reducing performance.

The presence of stereotype threat is itself evidence of schemas. If there were no stereotype or schema, or if people were unaware of it, or did not

care whether or not it was true, there could be no threat. Thus, the dimensions on which stereotype threat operates tell us something about the content of schemas.

Stereotype threats—or their absence—also affect the extent to which people feel they belong in a setting. If you *do* belong, it is hard to appreciate how much it matters not to belong. Research stemming from belongingness theory (Baumeister & Leary, 1995) has delineated the importance of a feeling of belonging for academic success (e.g., Glass & Westmont, 2014). Most people, regardless of their gender, race, or ethnicity, want to feel that they belong, and a positive feeling of belonging is correlated, among college students, with academic success.

Stereotype threat effectively announces that the person does not belong. For example, environmental cues that indicate that a given domain is “White” or “male” affect how much women and racial-ethnic minorities feel that they belong in academic settings (Cheryan, Plaut, Davies, & Steele, 2009; Murphy et al., 2007; Walton & Cohen, 2007, 2011). A corrosive sense of not fitting in, of being unwelcome, is usually accompanied by depressed performance—and confirmation of negative stereotypes. White men can feel similarly uncomfortable and out of place in a group of African American women. That is not to say that people cannot overcome stereotype threat. They can. But it requires cognitive resources to do so. People who are in environments where they feel—and observers feel—that they fit have an easier time.

When African Americans and Whites are having a conversation together, different threats are activated. African Americans are concerned about appearing competent, believing as they do that Whites are likely to think they are not competent, while Whites are concerned about appearing likable and nonracist, believing as they do that African Americans are likely to think they are racist (Bergsieker, Shelton, & Richeson, 2010). Both Whites and African Americans have different goals in cross-race interactions than in same-race interactions. And their concerns about what impression they are making takes a cognitive toll for both parties (Richeson & Shelton, 2007).

Both gender and race-ethnicity produce complex expectations in ourselves and in observers about what we are good at and what we can achieve (Steele, 2011). Those expectations in turn can enhance or diminish our performance. Because the dimension of competence or intelligence is so

consistently important in the construction of gender and racial-ethnic stereotypes, the academy—where this kind of talent is the one at stake—is a particularly charged domain.

The Specific Content of Gender Schemas

We have spoken generally thus far about schemas. Gender schemas are hypotheses about what it means to be male or female, hypotheses that we all share, male and female alike. Gender schemas assign different psychological traits to males and females (Martin & Halverson, 1987; Spence & Helmreich, 1978). We think of males as capable of independent action (agentic), as focused on the task at hand (task oriented), and as doing things for a reason (instrumental). Thus, we expect men to demand compensation for their activities. As mentioned earlier, we think of females as nurturant; we also see them as being expressive and behaving communally. Thus, we expect women to “labor for love” and see them as selfish if they demand compensation; they have violated the schema that women are nurturant and communal. In brief: men act; women feel and express their feelings.

Our beliefs about gender differences have some support. First, people differ in how strongly they see different characteristics as applying to them. While men and women overall are roughly similar on the characteristics we associate with men—traits like “stands one’s ground,” women do endorse more of the characteristics we associate with women—traits like “kind”—than men do. Second, the sexes agree on what *other* people are like, seeing college students as more polarized in “masculine” and “feminine” characteristics than they have said they themselves are. There is a gap between how people see themselves and how they see the average person (Ridgeway, 2011).⁹

Since men are associated with the competent capable norm, it is safe for women to acquire some of the hallmarks of the male gender schema, and in fact women must do so if they are going to be perceived as competent. However, if women do not include some of the hallmarks of the female gender schema in their behavior, they will experience backlash (Heilman, 2012; Heilman & Okimoto, 2007; Ridgeway, 2011; Rudman & Phelan, 2008). Men run the risk of seeming incompetent if they appear too “feminine,” and they experience backlash if they appear modest (Moss-Racusin, Phelan, & Rudman, 2010) or succeed at “feminine” tasks (Rudman & Fairchild, 2004).

Men who succeed at “feminine” tasks are also seen as wimpy (Heilman & Wallen, 2010).

As a result of schemas, people overrate men’s performances and underrate women’s. The differences are usually small but consistent. For example, college students see male college professors as having more innate ability than female college professors. An analysis we ran of adjectives used in Rate My Professors descriptions shows that the word *brilliant* is used more often for men than women in every academic field, even those like accounting, where *brilliant* is seldom used (sorry, accountants). Similarly, the word *genius* is more often used for men than women, with an even bigger gender disparity than *brilliant*. As one might expect, women are rated as “helpful” and “sweet” more than men are—in every field. But negative adjectives related to student treatment, such as “mean” and “strict,” are also more common for women, possibly because women who are not “helpful” and “sweet” are violating gender schemas. Men are rated as much more “arrogant” than women in every field. (See <http://benschmidt.org/research-portfolio/>, Gendered Language in Teacher Reviews, to put in your own search terms.)¹⁰

The more male dominated a field is, the greater the belief about the importance of innate ability compared to motivation and hard work (Leslie, Cimpian, Meyer, & Freeland, 2015). Faculty and graduate students in different academic disciplines rated how much their field required abilities that could not be taught. The fewer women that the field contained, the more people thought their discipline required innate talent. Our field, psychology, which grants PhDs primarily to women (more than 70%), was rated by people in the field as requiring little field-specific ability compared to hard work. In contrast, philosophy, which has a relatively low proportion of women PhDs compared to other humanities fields (around 30% of 2011 PhDs), was rated as requiring more field-specific innate ability than any other field.

Overall, then, gender schemas skew our perceptions and evaluations of men and women, causing us to overrate men and underrate women. Gender schemas affect our judgments of people’s competence, ability, and personal characteristics. Many of our judgments are small everyday events, such as not listening when a woman talks to us or not congratulating a woman on an achievement. Moreover, these small but frequent occurrences, which seem too petty to notice, accumulate to advantage men and disadvantage women.

Experimental Data about Perceptions of Sex Differences

Experimental data demonstrate that we do not see other people simply as people; we see them as males or females. Once gender schemas are invoked, they work to disadvantage women by directing and skewing our perception, even in the case of objective characteristics like height. In one example (Biernat, Manis, & Nelson, 1991), the experimenters exploited the fact that our schemas include the—of course correct—information that men are on average taller than women. In this experiment, college students saw photographs of other students and estimated their height in feet and inches. The photos always contained a reference item, such as a desk or a doorway, so that height could be accurately estimated.

Unbeknownst to the students who were doing the estimating, the experimenters had matched the photographs so that for every photograph of a male student of a given height there was a female student of the same height. But the students were affected by their knowledge that men are on average taller than women. They judged the women as slightly shorter than they really were, and the men as taller. This finding is reminiscent of the study we referred to earlier, where parents underestimated their 11-month-old daughters' motor ability and overestimated their sons' (Mondschein et al., 2000).

In the experiment on height, there were no differences in male and female observers. That is the usual finding. We all have hypotheses, some conscious and some not, about males and females, and we all use those hypotheses in perceiving and evaluating others. The important point about these two studies is that genuinely objective characteristics—height and motor ability—are not immune from the effects of gender schemas. Observers are influenced by what is generally the case, as in the height example, or what *will* generally be the case, as in the example concerning infants' motor ability. And it makes cognitive sense to be influenced by the average. Absent other information, the average is a good measure to go by. At the same time, that strategy has a cost for the individuals who are misjudged. When one underestimates people's abilities or skills, one tends to give those people less credit than they deserve, to ask less of them, and to rob them of the potential for growth.

Four Examples of Judgments of Professional Competence

In the case of professional competence, which differs from height in being harder to measure, our perceptions are similarly prone to error. We are likely to overvalue men and undervalue women. We can see why that would be the case: gender schemas will play a large role in evaluations (1) whenever schemas make a clear differentiation between males and females, and they do for professional competence as much as for height, and (2) when evidence is ambiguous and open to interpretation, as is the case with professional competence. It is tempting to think excellence is straightforward, but it is not. Four experiments on judgments of women's competence demonstrate the effects of gender schemas and the costs for women in professional life.

The first study investigated how males and females rated people who were described as being an Assistant Vice President in an aircraft company (Heilman, Wallen, Fuchs, & Tamkins, 2004). The evaluators read background information about the person, the job, and the company. The information they read about the person included the person's birthplace, the college they attended, their grades, how long they had worked in the company, how much management training they had received, how many employees they currently supervised, and their personal interests. In half the cases, the person was described as about to have a performance review. Thus, in this condition, evaluators did not know how well the person was doing in the job. In the other half of the cases, the person was described as having been a stellar performer. The evaluators' job was to rate the employees' competence and likability.

When evaluators had no information about how well people were doing in the job, they rated the man as more competent than the woman and rated them as equally likable. The finding that both men and women rate women as less competent than men, even when given identical information, is a typical finding in the social psychology literature. What was new about this experiment was looking at what happened to likability.

When the background information made clear that the individuals were extremely competent, evaluators rated the man and the woman as equally competent, showing that evaluators are responsive to clear information. It is only when the information is ambiguous that men have an advantage in competence ratings. But evaluators rated the competent woman as much

less likable than the competent man. They also perceived the woman as considerably more hostile than the man.

Thus, in evaluating a woman in a male-dominated field, male and female observers alike see her as less competent than a similarly described man unless there is clear information that she is competent. And in that case, they see her as less likable than a comparable man. Both males and females see competence as the norm for men and as something that has to be demonstrated unequivocally for women. Both males and females see competent men as likable. Neither males nor females see competent women in male-dominated positions as likable.

And likability matters: in a follow-up experiment, the experimenters described the people to be evaluated as high or low in competence and high or low in likability. People rated those who were high in likability as better candidates for being placed on a fast track and as better candidates for a highly prestigious upper level position. It would be poor advice to tell women just to be competent because likability can make the difference in whether or not people get rewards. Again, there were no male-female rater differences.

A second study shows that there is a trade-off for women between competence and femininity (Phelan, Moss-Racusin, & Rudman, 2008). Observers heard fictitious interviews for a computer lab manager job in which both the interviewer and interviewee were actors who used exactly the same scripts. The observers were told that the job required strong technical skills as well as social skills—because of the need to help students and faculty. In one condition, the interviewees adopted an assertive style emphasizing their competence; in the other, they adopted a style that emphasized how communal they were. Observers rated the interviewees on their competence, their social skills, and how hireable they were.

Evaluators generally gave more weight to competence than social skills. The notable exception was women who were assertive. In that case observers gave more weight to social skills. Since assertive women were seen as not having social skills, they were also seen as less hireable than assertive men. Again, there was no difference in judgments on the part of female versus male observers. Women are thus in a difficult position. If they are not perceived as competent, they will not get the job. But if they make their competence clear by behaving assertively, they will be seen as lacking social skills and will be downgraded for that reason.¹¹

The third study demonstrated how people shift their standards in order to justify a choice that seems a priori reasonable to them (Norton, Vandello, & Darley, 2004). In this experiment, gender schemas determined what seemed reasonable. The experimenters asked male undergraduates to select a candidate for a job that required both a strong engineering background and experience in the construction industry. The evaluators rated five people, only two of whose résumés were competitive. One candidate had more education—both an engineering degree and certification from a concrete masonry association—than the other, who only had an engineering degree. The other candidate had more experience—a total of nine years—than the first, who only had five years. Thus, there was a trade-off between accreditation and experience.

In the control condition, the candidates were identified only by initials. Here, the evaluators chose the candidate with more education three-quarters of the time and education was the reason most often cited as important for their decision. In one of the two experimental conditions, a male name was given to the résumé that had more education and a female name to the résumé that had more experience. Here, too, evaluators chose the candidate with more education three-quarters of the time and also rated education as very important. In the second experimental condition, a female name was given to the résumé with more education and a male name to the résumé with more experience. Now, less than half of the evaluators picked the person with more education and less than a quarter cited education as the most important characteristic.¹²

Men look more appropriate than women for the job of construction engineer, whether they have more education *or* more experience. The standards by which we judge people shift depending on our a priori judgments about their goodness of fit. Gender schemas help determine goodness of fit. This study demonstrates how easily people can shift their standards if they have not antecedently decided what the criteria are.

The fourth study bears on a possible objection to the studies just reviewed: the participants are (generally) college students who are not trained in evaluating others. Working scientists, the thinking goes, would be able to judge men and women accurately, both because scientists are trained to objectively and dispassionately evaluate data and because they subscribe to the merit principle. Thus, any differences in the representation of men and women in science would be due to the different choices

that men and women make, due to factors that are internal to men and women, rather than due to the social structure or evaluators' judgments. But it turns out that scientists—chemists, physicists, and biologists—are just like undergraduates.

University scientists were given a description of a possible lab manager—an undergraduate student who would manage a laboratory (Moss-Racusin, Dovidio, Brescoll, Graham, & Handelsman, 2012). The description of the candidate was ambiguous, indicating weaknesses as well as strengths. An ambiguous description was used because, as we have noted, observers are able to recognize outstanding talents in women. It is when people are less than outstanding (which most people are) that schemas have more room to operate. Given the ambiguous description, we would expect scientists to respond by overrating the man and underrating the woman. And that is what they did—they rated the female candidate less positively than the identical male candidate.¹³ As usual, there were no differences between male and female scientists in the ratings.

The scientists also completed a Modern Sexism Scale designed to tap subtle beliefs about gender equality (Swim, Aikin, Hall, & Hunter, 1995). It included eight items on which the scientists indicated the extent of their agreement, such as "Society has reached the point where women and men have equal opportunities for achievement" and "Women often miss out on good jobs due to sexual discrimination."

The scale taps "modern" sexism in two ways. First, participants indicate the extent to which they think that problems related to gender and achievement are in the past, and second, beliefs like equality of opportunity do not appear sexist on their face. They are not the same as believing that women are intellectually inferior to men. Scientists are reluctant to voice such attitudes, but thinking that sexual discrimination is in the past seems like a benign belief. Despite how mild such sentiments seem to be, the more the scientists endorsed statements indicating that women faced no particular hurdles, the more likely they were to rate the female candidate poorly. (Scientists' evaluation of the male candidate was unaffected by their score on this scale.) There were no differences between men and women scientists in the extent to which they believed sex discrimination was a thing of the past.

These results suggest that beliefs that women and men are on an equal footing make it more likely that an evaluator will judge a woman poorly. It appears that people go from endorsing the desirability of judging people

without regard to gender to believing that the world actually does judge people without regard to gender.¹⁴ Part of the problem, then, is that people's conviction that gender equality is a fact, and not simply a desired state, impairs their judgment of women. Their very belief in the existence of equality is associated with more negative evaluations of women.

Perceptions of Outstanding Women

We have alluded several times to the differences between people's responses to outstanding individuals and individuals who are closer to the average. There are signs of progress in the perception of leaders, at least among some people in some settings. For example, in a survey that was available online to visitors to the MSNBC site, people were invited to rate their boss. Over a ten-day period, roughly 60,000 people contributed usable data (Elsesser & Lever, 2011). Although differences in ratings were very small, women with male managers rated them as more competent than did women with female managers; men were more evenly split, but men with female managers rated them as very slightly more competent than did men with male managers. That effect appeared to be driven by competition: whichever gender people reported competing with at work was the gender that got lower competence ratings. Those results suggest that people do not necessarily rate their manager's competence along gender lines.

However, a somewhat different picture emerged when people were asked to indicate which sex they would prefer to work for. Among the slightly less than half who indicated a preference, men were preferred at more than double the rate at which women were preferred. The preference for men was even stronger in male-dominated occupations, with no difference between male and female participants.

Why might people's preferences in the abstract differ from their evaluations of their own managers? It may be easy for people to see their own manager as an exception to a more general tendency—to engage in subtyping (Richards & Hewstone, 2001). In subtyping, seeing someone who is very different from one's expectations may result in an accurate perception of that person without a change in one's overall view. The person is an exception who leaves the rule intact. And that is a rational way to behave: a single experience should not overturn a view based on multiple experiences. In the domain of social perception, however, where perceptions are

affected by a person's group membership, and where beliefs can substitute for firsthand experience, the "rule" may itself be incorrect.

Businesspeople's perceptions of men and women as leaders suggest that women are seen as excelling in some of the traits associated with the gender schema for women, such as supporting and rewarding others, while men are seen as excelling in traits associated with the gender schemas for men, such as delegating and influencing one's superiors (Prime, Carter, & Welbourne, 2009). Other traits, such as the "masculine" trait of problem-solving, are seen by women as more common for women but are seen by men as more common for men. A meta-analysis of 99 studies of perceptions of male and female leaders suggests that, overall, both men and women rate their female leaders more highly than their male leaders, but when rating themselves as leaders, men rate themselves much more highly than women rate themselves (Paustian-Underdahl, Walker, & Woehr, 2014). One reason for the superior rating of women in top positions may be that women in those positions are seen as having overcome, and may actually have overcome, strong odds against them through superior competence (Foschi, 2000; Paustian-Underdahl et al., 2014).

Taken together, the results suggest that people can acknowledge outstanding female performance under certain conditions. Consider, for example, a study in which students rated individuals applying for a position of either a football or a tennis photographer on a magazine (Heilman, Martell, & Simon, 1988). Football photography is considered a male-dominated field compared to tennis photography. In one condition, the dossier was labeled as awaiting review, and in the other the person was labeled as a finalist, having been chosen by experts as worthy of being in final consideration.

In both the tennis case and the football case, when evaluators had no information about how the candidates had been rated, female candidates were seen as less competent than male candidates. Absent information to the contrary, people rate men more highly than women. However, when evaluators had information that the women were finalists, ratings of tennis and football photographers were not the same. When women were finalists in tennis, there was no difference in ratings for male and female candidates; legitimation by experts made men and women equal. When women were finalists in football, they received ratings that were *more* favorable than men's (Heilman et al., 1988). Evaluators seemed to think that the woman

must be a star to have received a favorable rating from experts in a highly male-dominated field.

A similar finding arose in a study of rankings of job candidates for a professorial position in science (Williams & Ceci, 2015), where evaluators were presented with descriptions of three candidates, two of whom had already received very high ratings from a search committee. One of the two was male, and one was female. In line with the study on sports photographers, evaluators—in this case academics—ranked the woman as the person to hire.¹⁵ When women are stars, they will be recognized.

This review of gender schemas in action demonstrates that both men and women are likely to overrate men and underrate women in settings where professional competence is at issue. Different reactions to the same behavior, occurring because of the sex of the person displaying the behavior, happen among college students, business school students, working adults, and working scientists. Part of what makes schemas so powerful is their very subtlety. On implicit measures, both men and women evaluate men more positively than women, even though most individuals demonstrate a commitment to judging individual merit rather than basing judgments on group membership.

Laboring for Love

A female full professor, whom we will call Carol, provided the following example. Carol had spoken to her chair, whom we will call John, about the differential treatment of two new faculty who had been hired two years earlier and whose research overlapped. John was a friend of Carol's; she liked him and respected him. The male assistant professor had taught the same two courses every semester, one of which was a graduate seminar in his area. He recruited students to work with him from that seminar. The female assistant professor had taught different courses every semester, always undergraduate "service" courses. Carol suggested that John should make more equitable assignments. John agreed that the inequity was helping the junior male and hurting the junior female. But, he said, the junior male hadn't been very cooperative and had resisted John's request that he teach other courses. The junior male also wasn't a great teacher. The junior female didn't put up a fight about her teaching and did a good job at it. He had to think about the department and the students.

Relatively little quantitative data from 2010 on are available to show how teaching and service responsibilities are divided among women and men, but women appear to teach more (Carrigan, Quinn, & Riskin, 2011) and perform more service, regardless of rank (Guarino & Borden, 2017; for a review, see Mitchell & Hesli, 2013). White women and men and women of color perceive that they spend more time than their White male colleagues mentoring women and students of color and spend more time, relatively speaking, on committees so that those committees can be more diverse. One study of over 1,000 faculty in political science found that women say no to committee assignments less often than men do (Mitchell & Hesli, 2013). Women are asked to serve more often than men are, and they accept more invitations than men do. They do not volunteer more often than men. In addition to serving on more internal committees, the women also had more undergraduate students than the men (though they did not supervise more projects) and fewer postdocs. Examination of faculty workload reports across a range of disciplines at a single university found similar results: women performed more service than men (O'Meara, Kuvaeva, & Nyunt, 2017). Women are also asked to perform more service, according to diary entries of individual faculty at 13 of the 14 Big Ten universities (O'Meara, Kuvaeva, Nyunt, Waugaman, & Jackson, 2017).

Time allocation studies provide inconsistent reports about whether men and women of color spend more time teaching and mentoring than their White peers (for discussion, see Martinez & Toutkoushian, 2014). Latinx faculty are similar to White faculty in their time allocations, except that Latinx faculty spend more time than do Whites on unpaid work in the community (Martinez & Toutkoushian, 2014).

The request for women's services is expected since the schema for women portrays them as nurturant and communal. Women are expected to, and will, labor for love. We see this explanation as part of the gender schema story. Institutions can ameliorate the problem of unequal demands on different demographic groups by reviewing their teaching and committee assignments for fairness. Chairs can learn how to handle refractory faculty. If White women and men and women of color are doing more than their fair share of committee service, they can be compensated for that extra service via course releases or research assistance.

When Good Intentions Go Awry

What is most sobering for evaluators who wish to be fair is that simply explicitly disagreeing with an overtly sexist statement makes it more likely that an individual will then go on to pick a man for a stereotypically male job (Merritt, Effron, & Monin, 2010). People were asked to say whether a job in the construction industry—a job typically associated with men—would be better for men, better for women, or equally suited to either sex. One group of participants was first asked to agree or disagree with statements like “Most women are not very smart.” Another group did not read any sexist statements ahead of time. Male participants who read the statements were likely to disagree with them, but that very disagreement made them more likely to choose a male for the construction job than the participants who had not read any statements (Monin & Miller, 2001). (Women were unaffected.)

It appears that once individuals have announced their bona fides, they think they have demonstrated a lack of bias. That in turn makes it more likely that, on their next opportunity in a situation that activates gender schemas, they will demonstrate bias. People can reassure themselves that their decision was made on the basis of merit because they have at hand judgments that they had just made that suggested to them that they were free of bias. Since academics routinely espouse the merit principle, this moral “licensing” should give everyone pause. The statement with which we opened the chapter—“I don’t care who they are; I just want the best person”—is a good example. Once people have assured themselves that they judge fairly, they are less concerned about their behavior in any particular case.

Again, the propensity to make errors in evaluation can best be handled by having explicit valid criteria and procedures.

Race and Ethnicity Schemas

In the same way that individuals are egalitarian and meritocratic in their explicit comments about women, they are egalitarian and meritocratic in their explicit views of African Americans, Hispanics, and other under-represented minorities. And, as with gender, subtle tests suggest a more

complicated picture. When people say that they “don’t see race,” they are not taking into account data about how evaluations work. Whether judging males or females, or African Americans or Whites or Hispanics, we generally take people’s social identities into account. Yet it is hard to say what the content of our race and ethnicity schemas is. In the case of gender, we can articulate the core of each schema. For the male gender schema, agency is the core; for the female gender schema, nurturance is the core. For ethnic groups, it is more difficult to say what the core is. For example, what is the core of the White ethnicity schema?

Different schemas characterize different racial-ethnic groups. African Americans and Hispanics are viewed as lower than Whites on competence and warmth, while Asians are rated as high on competence but low on warmth (see, e.g., Zhang, 2015). Schemas about different ethnic groups are related to schemas about job prestige. Hispanics are assumed to hold lower status jobs than non-Hispanics, and in turn to have less of a work ethic and to be more “traditional,” that is, in having a greater family orientation and religiosity (Jones, 2001). Both African Americans and Whites see Whites as having high status (see the review in Fiske, Dupree, Nicolas, & Swencionis, 2016). Similarly, in a study of adolescents’ stereotypes, beliefs that African Americans were not successful in school, and that Hispanics engage in manual labor occupations, underlay adolescents’ stereotypes about their intelligence, compared with those of Asians (Kao, 2000).

Status enters via the terms used to describe different groups. Whites evaluate “Blacks” more negatively than “African Americans” or “Whites,” and see them as having lower status and as being less competent (Hall, Phillips, & Townsend, 2015). Whites also see “Blacks” as having less education than “African-Americans” and as making less money. Further, in newspaper articles, “Blacks” are associated with more negative emotions than “African Americans,” especially more terms conveying anger. For Whites, the term “Black” conveys less respect than the term “African American.”

Americans of African descent are about evenly split in their preferences between “Black” and “African American,” but different subgroups show a strong preference for “African American” (Sigelman, Tuch, & Martin, 2005). For example, the combination of having been in a grammar school that was roughly 50–50 African American and White, plus being younger, plus living in a city, plus not living in the South, plus identifying strongly with an African American identity was associated with a strong preference for

“African American.” (See Agyemang, Bhopal, & Bruijnzeels, 2005, who argue that there are health treatment consequences to what labels are used, and Philogène, 1999.)

Ethnicity in the Workplace

Since we are primarily concerned with professional competence, we can concentrate on the data concerning schemas of different ethnicities in the workplace. In rating a series of adjectives taken from various studies of race schemas, a group of business school students (African Americans, Whites, Asians, Hispanics, and other ethnicities) perceived hypothetical African American and White managers differently. Whites were seen as more competent, ambitious, and manipulative; African Americans were seen as more skillful interpersonally but less polished. (There were not enough participants to determine whether African Americans and Whites examined separately would differ in their judgments.) When the participants were asked to rate successful managers, rather than just managers, Whites continued to be seen as more manipulative and African Americans as more skillful interpersonally and less polished (Block, Aumann, & Chelin, 2012). In a related study, similar perceptions of Whites came from ratings by minority students—African Americans, Asians, and Hispanics. Some of the positive traits that all three ethnic groups saw as characteristic of Whites were having more opportunities, being intelligent, being confident, and being outgoing (Conley et al., 2010).¹⁶

Since competence is the key characteristic for professional success—at least when White men possess it—the fact that Whites are seen as more competent gives them an advantage in the same way that the perception of men as more competent than women gives them an advantage. For example, when men and women of varied ethnicities (but primarily White) believed that they had performed poorly on a task, they were more likely to relinquish leadership to a White male than to an African American male (Ratcliff, Vescio, & Dahl, 2015).¹⁷ Similarly, participants were more likely to cede power to a White man than a White woman (Ratcliff et al., 2015).

Differential assumptions about ability have implications for evaluations of success and failure. In one study, participants read about a company that had succeeded or failed either because of a leader’s performance or because of external conditions that were independent of anything the leader did

(Rosette, Leonardelli, & Phillips, 2008). The company leader was sometimes given a “Black” name (Tyrone) or a “White” name (Todd). A picture was provided of the leader. Thus, evaluators had some detailed information about the leader. Observers rated the leader on competence, confidence, intelligence, and competitiveness.

In the condition in which the organization had been successful, Whites were rated as more effective than African Americans when internal reasons were given for the success, but when external reasons were provided for success, the two groups were rated as equally effective. Thus, Whites are seen as having earned their success, even when identical language is used to indicate that African Americans are responsible for their success. When the organization had not been successful, there were no differences in ratings for Whites and African Americans. In this study, there were no differences as a function of the participant’s race. Thus, it appears that non-Whites are as likely as Whites to see successful leadership as White.

However, race does not have uniform effects on non-African Americans’ perceptions of African American men and women. Non-African American participants rated a leader in one of eight conditions: the leader was White or African American, male or female, and showed dominant or communal behavior. Participants rated the leader’s effectiveness on several dimensions (Livingston, Rosette, & Washington, 2012).

African American female leaders were seen as equally effective whether they were dominant or communal, but White female leaders were penalized for showing dominant behavior rather than communal behavior. Among male leaders, African Americans were penalized for being dominant rather than communal, but Whites were rated equally highly whether they were dominant or communal. Thus, non-African American observers “allow” African American women and White men to be dominant or communal but do not allow African American men or White women to be dominant (Livingston et al., 2012).

It is worth noting that—partly because of the long and painful history of slavery and race relations between African American and White Americans—more is known about African American stereotypes than about those of other racial and ethnic groups. One meta-analysis found that in the three social psychology journals that published much of the research on schemas about different races and ethnic groups, only 7% of that research focused

on stereotypes about Hispanics, while 61% focused on stereotypes about African Americans, even though Hispanics are a larger percentage of the population (Dovidio, Gluszek, John, Ditlmann, & Lagunes, 2010).

Why We Need to Pay Attention to Race and Ethnicity

We understand why Whites find the idea of paying attention to race aversive. At first blush, attending to race seems like a violation of the merit principle. Shouldn't one judge people completely on their merits? If one could really be blind to race, color, and ethnicity, wouldn't the world be a better place?

We have argued that it is not possible to perceive people independently of their social identities. For that reason, it is better to be on one's guard about the perceptions one is likely to have than to deny their existence. Until the world really *is* fair, we should not act as if it is. When there are clear and valid standards for performance, and when it is possible to evaluate people without knowledge of their social identity, then people can judge impartially (e.g., Goldin & Rouse, 2000; Self, Mitchell, Mellers, Tetlock, & Hildreth, 2015). However, in keeping with our earlier discussion of the importance of visual cues, listeners rely on them even when they are evaluating musical performance (Tsay, 2013). As long as there are cues to social identity, social identity has an impact on judgments.

Trying to adopt a race-neutral approach in a world that sees race and ethnicity can backfire (Apfelbaum, Norton, & Sommers, 2012). For example, children who heard a story about a teacher who adopted a color-blind approach were less likely to subsequently recognize overt discrimination than were children who heard a story about a teacher who did not adopt a color-blind approach (Apfelbaum, Pauker, Sommers, & Ambady, 2010).

If we think that race and ethnicity *can't* play a role, we are unlikely to see the occasions when it is playing a role. We recognize that one is not always correct in seeing a role for social identity. If one is always looking for evaluation errors, one may see some where they do not exist. Whatever hypothesis we hold is likely to affect how many confirming examples we see. We argue that it is healthy to start with the possibility that one is choosing the White over the African American for reasons that are incidental to their actual performances. That will make it more likely that we perceive

evidence in favor of the hypothesis. Whites are already highly motivated to find that hypothesis to be wrong, so they need consciously to look for evidence that the hypothesis is right.

Another reason for Whites to pay attention to race and ethnicity is that underrepresented minorities prefer to have differences acknowledged rather than ignored and prefer a multicultural approach in which different orientations are valued, rather than a color-blind approach or an assimilationist approach in which minorities are assimilated to the majority culture (Dovidio, Gaertner, Ufkes, Saguy, & Pearson, 2016; Ryan, Casas, & Thompson, 2010; Ryan, Hunt, Weible, Peterson, & Casas, 2007). Whites could thus recognize difference and value it instead of denying its existence or effects.

When Leaders Fail

All leaders occasionally fail. When women or people of color fail, their failure is likely to be attributed to their gender or ethnicity. When White men fail, in contrast, they are more likely to be seen as failing as individuals (or as failing because the task was very difficult). People do not say, "Oh, so-and-so was a poor president because he was White." They say, "Oh, so-and-so was a poor president because he lacked initiative," or "He didn't have people skills," or "He was a poor delegator." When women or people of color fail, they are seen as failing because of their group membership (Miller, Taylor, & Buck, 1991). They failed because they were women (and women as a group don't have the qualities necessary for success), or because they were African American (and African Americans as a group don't have the qualities necessary for success).¹⁸

Accumulation of Advantage

Many of the slightly too negative evaluations that women receive and the slightly too positive evaluations that men receive seem minor. Women and men notice some of them but not others in daily life. As we mentioned in the preface, the most commonly noticed examples arise in meetings. A woman makes a suggestion, and no one pays any attention to it. A little later a male colleague makes the same suggestion, and everyone says, "Joe, what a great idea." What makes it hard for the woman herself to evaluate this event is that everyone's ideas are ignored some of the time; she doesn't

want to be paranoid. Neither she nor anyone else has been keeping a running tally. Is it really a systematic phenomenon? Did she not express her idea in an optimal way? Did the idea just need time to percolate through the group? And, should she mention the phenomenon to a colleague, that colleague might or might not agree that women weren't attended to quite as much but would probably suggest that it's not worth her time to think about it, let alone protest it. She shouldn't make a mountain out of a molehill (Valian, 1998).

One woman told us about giving a keynote address at a major international conference. Her department put that honor on their website for two days. It was then superseded by an announcement about a male colleague who would be giving a keynote address at a much smaller conference the following year. His notice stayed up for over a week. (In fact it stayed up until she approached her chair about the reason for his being featured.) Her major achievement was briefly recognized; his minor achievement—which had not even happened yet—was given much more attention. In another website example, a woman was elected as a board member for a prestigious, national board. Although the department was informed about the honor, it did not put it on the website. Later, when a man was a candidate for the same board, his candidacy was put on the website. (He did not get elected.) Since these data are not systematic, we do not know if there were equally many examples that went in the opposite direction; the accumulated data we have presented thus far would suggest not.

Again, someone might say about such events, don't make a mountain out of a molehill. Giving the keynote was the honor; being elected was the honor. Those were the important things. It doesn't matter that much that your department hardly noticed.

One feature of small negative effects is that they tend to be more frequent than big negative effects. They are chronic. They may be like gnats, but the effort of swatting away gnats on a frequent basis is stressful. A meta-analysis comparing the negative effects of subtle and overt discrimination found that both were correlated with negative effects in a variety of areas (Jones, Peddie, Gilrane, King, & Gray, 2016). Subtle discrimination referred to the kinds of small events we have been discussing (Hebl, Foster, Mannix, & Dovidio, 2002). The negative effects ranged from decreased success in one's job (as measured by promotions and productivity) to substance abuse. Correlations were as high for the subtle measures of discrimination as for the

more overt measures. Small negative experiences turn out to have negative effects on the person that are comparable to those from large negative experiences.

But they also have negative effects over time. That is where the notion of the accumulation of advantage becomes relevant. The Matthew effect (Merton, 1988; Zuckerman, 1997) was developed to explain disparities in scientific recognition: to oversimplify, the rich (or recognized) get richer (even more recognized). (See Rigney, 2010, for a recent summary of effects in a range of fields and for a discussion of the limits of the accumulation of advantage; see also DiPrete & Eirich, 2006.) It lets us see that mountains *are* molehills, piled one on top of the other (Valian, 1998). Most success is the accrual of small gains that compound over time, like interest on investments. If one fails to get one's fair share of the interest, one ends up with less in the long run.

A computer simulation of this phenomenon showed that a tiny amount of bias—an amount accounting for only 1% of the variability in determining who would get promoted—resulted in a striking difference over time in the composition of the uppermost level of the simulated organization (Martell, Lane, & Emrich, 1996). The simulation assumed an eight-level hierarchically organized company, with 500 positions at the lowest level and 10 positions at the highest. It was staffed initially by the same number of men and women.¹⁹ There was an attrition process, so that 15% of incumbents were periodically removed. The promotion process assumed equal performance by men and women, but men received bonus points—for being men—that amounted to 1% of the variance in scores. The promotion process was repeated until the organization had completely turned over. At that point, the uppermost level was 65% male. (Twenty iterations were run, so that the average was 65%.) Thus, even a tiny amount of bias, an amount most institutions would be delighted to achieve, adds up over time to a mountain of disadvantage.

Summary

Our presentation of the literature on gender schemas and race and ethnicity schemas shows that, without intending to, we treat people differently on the basis of their gender, their race, their ethnicity, and a number of other factors. We are likely to undervalue the competence and professional

ability of women and members of underrepresented groups. An analysis of schemas looks underneath the more common reasons suggested for the relative paucity of certain groups and shows how some of those apparent differences—including apparent differences in interests—could arise. Those differences have consequences for the opportunities that we make available to people, the expectations that we have of them, the support that we provide them, and the evaluations that we give them. Although we believe we are attending only to merit when we evaluate others, our impressions are shaded by our schemas. We are unlikely to eradicate the cognitive strategies that give rise to those impressions. The first step is to become aware of them. The second step is to create procedures that will protect us from the mistakes resulting from those strategies.

Notes

1. Early sources of information about schemas in the sense that they are being used here include Bakan, 1966; Fiske & Taylor, 1991; Martin & Halverson, 1987; and McCauley, Stitt, & Segal, 1980. These researchers link schemas or probabilistic stereotypes to general cognitive concept formation.
2. In Bayesian terms, schemas are the equivalent of what people in some fields call “priors” or expectations. They can thus be modified, in any individual case, by the information presented, depending on the strength of the priors, the nature of the information, and other factors. We thank Charles Brown for pointing out this parallel.
3. In characterizations of people’s personalities, race and ethnicity seem to recede in importance as more information is provided, but race and ethnicity remain important when judging leadership abilities, as we discuss below.
4. When possible, we refer to very recent studies, so that we can avoid discussing previously observed phenomena that may no longer hold. It seems unlikely, however, that parents have shifted their views about the motoric abilities of their 11-month olds since 2000.
5. A study of male conscripts in the Norwegian Armed Forces found that minimal individuating information did not alter their lower evaluations of a woman as a possible squad leader, but that intense exposure by sharing a room and working in the same squad as a female soldier did (Finseraas, Johnsen, Kotsadam, & Torsvik, 2016).
6. Schemas are somewhat different from implicit attitudes, as measured by, for example, the Implicit Association Test (IAT; Banaji & Greenwald, 2013). We emphasize the cognitive content of schemas and the continuity between social schemas

and schemas for inanimate objects and asocial events that lack intrinsic affect. Affect may be attached to schema content and often is in the case of schemas about human groups. There is disagreement in the literature about whether content and affect can be separated (e.g., Fiske, 2015). IAT measures correlate with behaviors toward members of different groups, though the extent of the relation and its importance is debated (Greenwald, Banaji, & Nosek, 2015; Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2015). It is difficult to alter IAT judgments (Lai et al., 2014, 2016), though some approaches have appeared successful (Carnes et al., 2015; Devine, Forscher, Austin, & Cox, 2012). Short-term procedures that have been used to change IAT performance have not reliably mediated changes in behavior (Forscher et al., 2017).

Whether the IAT measures associations, propositions, beliefs, attitudes, or some other mental structure, is not clear. For discussion of the nature of implicit bias, see, *inter alia*, Gendler (2008), Levy (2015), Madva and Brownstein (in press), and Mandelbaum (2016). We do not use the term *implicit bias*, despite its popularity and utility, and despite our agreement that much of human cognition is not conscious. The reason is that we do not equate rapid associations with bias.

7. People also make judgments about people's warmth and competence (for more detail, see the stereotype content model of Fiske, Cuddy, & Glick, 2007) and their moral character (Goodwin, 2015; Goodwin, Piazza, & Rozin, 2014). All three attributes are at least partially independent. People can, for example, have a fine moral character but be cold and incompetent. Status is highly correlated with perceived competence, especially by those who believe that people get what they deserve, with the result that low-status individuals are seen as incompetent. Gender and ethnicity play a role. In the United States, Asians, Jews, the rich, and professionals, even if they are female or ethnic minorities, are seen as competent and cold (Fiske et al., 2007). In contrast, older people, and the physically and mentally disabled, are seen as incompetent and warm. Higher-status groups, such as middle-class people, Christians, and straights, are perceived as both competent and warm. Lower-status groups, such as poor Whites, Hispanics, and Blacks, are seen as incompetent and cold.

8. The extent to which people are knowledgeable about the stereotype, the extent to which they believe it, the subtlety with which the stereotype is activated, and other factors affect the strength of the phenomenon (see Kiefer & Sekaquaptewa, 2007; see papers in Inzlicht & Schmader, 2012).

9. In the 1970s the differences in how men and women rated themselves on various traits were more marked than they were in 2000 (Spence & Buckner, 2000). The change over time was due primarily to women's seeing themselves as having more characteristics that we associate with men—such as “has leadership.” Men changed less in the extent to which they endorsed “feminine” characteristics. One reason for women to have changed more than men is that men define the norm in the professions, where it is not just acceptable but desirable to be capable of independent

action, oriented to the task at hand, and doing things for a reason. Even though kindness is socially desirable, our schemas about what is required to be a successful professional do not represent kindness as a useful characteristic. Another reason for women to have changed more is that they became a larger and larger percentage of the workforce between the 1970s and 2000.

People apply gender schemas less to themselves than to others (see also Gill, 2004; Heilman, 2012; Ridgeway, 2011). Males and females describe their traits more similarly than people's beliefs about gender differences would suggest. Even as the sexes move closer together in reporting that they share traits classically associated with one sex or the other, schemas continue to guide our perceptions and evaluations of others. We see people as more different than they actually are and penalize people when they deviate from what is expected.

10. Teaching evaluations of instructors by students show that when students in an online course believe their instructor is male, the instructor is rated more highly than when the students believe their instructor is female (MacNeill, Driscoll, & Hunt, 2015). See also Boring, Ottoboni, and Stark (2016) for further analysis of these U.S. data and French data, plus analyses showing that student evaluations do not measure teaching effectiveness. A meta-analysis similarly suggests that student evaluations do not correlate well with teaching effectiveness (Uttl, White, & Gonzalez, 2017).

11. A related study also found that women appear to be punished if they make their skills clear. Undergraduate participants read a résumé for a candidate for a job as a project manager. In one condition, the résumé presented the candidate in terms that were stereotypically masculine (competitive and self-aggrandizing); in another condition, it was stereotypically feminine (sensitive to others' needs). The same qualifications were present in both conditions. Participants evaluated a single candidate on multiple dimensions: likability, competence, social skills, interview probability, and hireability. Since the résumé was identical in all conditions, participants' evaluations of the candidate were based on the candidate's sex and whether the description "fit" the candidate's sex (Tyler & McCullough, 2009). If women's problem is a failure to promote themselves as strongly as men do, they should receive higher evaluations when their cover letter is stereotypically masculine rather than stereotypically feminine. Men rated the female applicant more negatively on each measure compared to the male applicant if her cover letter displayed typically masculine characteristics. On every dimension, men rated the stereotypical male candidate highest and the counterstereotypical female lowest. When the cover letter was more communal in tone, men still tended to rate the male applicant more highly on every dimension. Women aren't as good job candidates as men, period, where men are concerned. Women participants, however, were unaffected in their judgments of the candidates by the nature of the cover letter and showed no consistent preference for a male or female candidate. (The difference between male and female reactions in this study may be due to women's heightened sensitivity about job applications.)

12. A related experiment used a community sample of people in a public park instead of undergraduates. The findings were similar, except that women were less likely to shift their standards than men were (Uhlmann & Cohen, 2005). This is one of the rare studies in which women and men responded differently.

13. The scientists rated either a male or a female candidate. Scientists who saw the male candidate's qualifications rated him as more competent and more hireable than did those who rated the female candidate. Scientists were willing to provide more career mentoring for the male candidate than for the female candidate. Scientists said they would pay the male more than the female, on average \$3,000 more.

14. A related finding is that people who do not see beliefs as sexist are also more likely to endorse them (Swim, Mallett, Russo-Devosa, & Stangor, 2005).

15. There were several methodological limitations to the study, as well as a lack of full presentation of the findings, making the study difficult to fully evaluate. For example, the study also varied descriptors used for each of the two competitive candidates, using either male- or female-stereotyped terms, sometimes for the male and sometimes for the female. Data were not presented separately for those conditions. As another example, ranking, which the investigators used, is a less preferable system than rating because it forces a search for extraneous reasons in candidates who might otherwise be evaluated as the same. As another example, both candidates were described and rated unrealistically highly, and to the same degree, in the paragraph that was the basis for ranking; similarly, in a control condition where evaluators saw CVs, the accomplishments were also unrealistically high. There was, unfortunately, no baseline condition—a condition without gender-based descriptors. There were no questions probing evaluators' understanding of the task. The descriptions of the candidates included irrelevant (and illegal) information about their marital and parental status. See Williams and Smith (2015) for methodological points.

Another study, in France, suggested that women were more likely to be chosen as instructors than equivalently competent men in fields where women were underrepresented (Breda & Hillion, 2016). The instructors were by and large being chosen for subuniversity (high school) positions, where they are referred to as professors but are not professors in the sense understood in the United States. It is not clear that evaluators would use the same metrics for judging university faculty. In addition, the report lacked necessary information about the structure of the examinations and the basis for evaluators' decisions. It is thus difficult to determine whether there are any implications of the study for college and university hiring.

16. Most studies examining race and ethnicity schemas do so with the aim of understanding more general attitudes that different groups hold of each other. Thus, investigators have concentrated on how positively or negatively people view others of the same and different ethnicities, how prone they are to interpret an ambiguous photo

(e.g., where an individual is holding something) in a way that suggests fear (e.g., as a weapon rather than a cell phone), and the extent to which Whites are implicit racists. We will discuss such studies when we consider the perceptions and behaviors of interviewers and other gatekeepers.

17. No data on ethnicity differences among participants were provided, but the majority of participants were White.

18. Although one experiment we reviewed earlier suggested that people saw Whites and African Americans similarly when they failed in leadership roles, that experiment explicitly provided evaluators with reasons. Either the failure was the fault of the person being evaluated, or the failure was due to external circumstances. The hypothesis we are presenting here is that when evaluators have to decide themselves what is behind a failure, they are likely to see Whites as responsible for specific reasons and African Americans as responsible because they are African Americans.

19. The two types of individuals were arbitrarily labeled males and females. They could have been labeled anything. The results are as applicable to race-ethnicity as to gender. Small amounts of preference amount, over the long haul, to large differences.

References

Agyemang, C., Bhopal, R., & Bruijnzeels, M. (2005). Negro, Black, Black African, African Caribbean, African American or what? Labelling African origin populations in the health arena in the 21st century. *Journal of Epidemiology and Community Health, 59*(12), 1014–1018.

Apfelbaum, E. P., Norton, M. I., & Sommers, S. R. (2012). Racial color blindness: Emergence, practice, and implications. *Current Directions in Psychological Science, 21*, 205–209.

Apfelbaum, E. P., Pauker, K., Sommers, S. R., & Ambady, N. (2010). In blind pursuit of racial equality? *Psychological Science, 21*, 1587–1592.

Asplin, B. R., Rhodes, K. V., Levy, H., Lurie, N., Crain, A. L., Carlin, B. P., & Kellermann, A. L. (2005). Insurance status and access to urgent ambulatory care follow-up appointments. *Journal of the American Medical Association, 294*(10), 1248–1254.

Bakan, D. (1966). *The duality of human existence*. Chicago, IL: Beacon Press.

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

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin, 117*(3), 497–529.

Bergsieker, H. B., Shelton, J. N., & Richeson, J. A. (2010). To be liked versus respected: Divergent goals in interracial interactions. *Journal of Personality and Social Psychology, 99*(2), 248–264.

Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review, 94*(4), 991–1013.

Biernat, M., Manis, M., & Nelson, T. E. (1991). Stereotypes and standards of judgment. *Journal of Personality and Social Psychology, 60*(4), 485–499.

Block, C. J., Aumann, K., & Chelin, A. (2012). Assessing stereotypes of Black and White managers: A diagnostic ratio approach. *Journal of Applied Social Psychology, 42*(S1), E128–E149.

Boring, A., Ottoboni, K., & Stark, P. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. *ScienceOpen Research, 1*–11. <https://www.scienceopen.com/document/read?vid=818d8ec0-5908-47d8-86b4-5dc38f04b23e>

Breda, T., & Hillion, M. (2016). Teaching accreditation exams reveal grading biases favor women in male-dominated disciplines in France. *Science, 353*(6298), 474–478.

Burch, T. (2015). Skin color and the criminal justice system: Beyond Black-White disparities in sentencing. *Journal of Empirical Legal Studies, 12*(3), 395–420.

Carnes, M., Devine, P. G., Manwell, L. B., Byars-Winston, A., Fine, E., Ford, C. E., et al. (2015). Effect of an intervention to break the gender bias habit for faculty at one institution: A cluster randomized, controlled trial. *Academic Medicine, 90*(2), 221–230.

Carrigan, C., Quinn, K., & Riskin, E. A. (2011). The gendered division of labor among STEM faculty and the effects of critical mass. *Journal of Diversity in Higher Education, 4*(3), 131–146.

Chan, W., & Mendelsohn, G. A. (2010). Disentangling stereotype and person effects: Do social stereotypes bias observer judgments of personality? *Journal of Research in Personality, 44*(2), 251–257.

Cheryan, S., Plaut, V. C., Davies, P. G., & Steele, C. M. (2009). Ambient belonging: How stereotypical cues impact gender participation in computer science. *Journal of Personality and Social Psychology, 97*(6), 1045–1060.

Conley, T. D., Rabinowitz, J. L., & Rabow, J. (2010). Gordon Gekkos, frat boys and nice guys: The content, dimensions, and structural determinants of multiple ethnic minority groups' stereotypes about White men. *Analyses of Social Issues and Public Policy (ASAP), 10*, 69–96.

Cortina, L. M., Magley, V. J., Williams, J. H., & Langhout, R. D. (2001). Incivility in the workplace: Incidence and impact. *Journal of Occupational Health Psychology, 6*(1), 64–80.

Devine, P. G., Forscher, P. S., Austin, A. J., & Cox, W. T. (2012). Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of Experimental Social Psychology, 48*(6), 1267–1278.

DiPrete, T. A., & Eirich, G. M. (2006). Cumulative advantage as a mechanism for inequality: A review of theoretical and empirical developments. *Annual Review of Sociology, 32*, 271–297.

Dovidio, J., Gluszek, A., John, M., Dittmann, R., & Lagunes, P. (2010). Understanding bias toward Latinos: Discrimination, dimensions of difference, and experience of exclusion. *Journal of Social Issues, 66*, 59–78.

Dovidio, J. F., Gaertner, S. L., Ufkes, E. G., Saguy, T., & Pearson, A. R. (2016). Included but invisible? Subtle bias, common identity, and the darker side of “we.” *Social Issues and Policy Review, 10*(1), 6–46.

Duckworth, A. L., & Seligman, M. E. P. (2006). Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. *Journal of Educational Psychology, 98*(1), 198–209.

Dweck, C. S. (1999). *Self theories: Their role in motivation, personality, and development*. Philadelphia, PA: Psychology Press/Taylor and Francis.

Elsesser, K. M., & Lever, J. (2011). Does gender bias against female leaders persist? Quantitative and qualitative data from a large-scale survey. *Human Relations, 64*, 1555–1578.

Finseraas, H., Johnsen, Å. A., Kotsadam, A., & Torsvik, G. (2016). Exposure to female colleagues breaks the glass ceiling—Evidence from a combined vignette and field experiment. *European Economic Review, 90*, 363–374.

Fiske, S. T. (2015). Intergroup biases: A focus on stereotype content. *Current Opinion in Behavioral Sciences, 3*, 45–50.

Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences, 11*(2), 77–83.

Fiske, S. T., Dupree, C. H., Nicolas, G., & Swencionis, J. K. (2016). Status, power, and intergroup relations: The personal is the societal. *Current Opinion in Psychology, 11*, 44–48.

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

Forscher, P. S., Lai, C., Axt, J., Ebersole, C. R., Herman, M., Devine, P. G., & Nosek, B. A. (2017, October 5). A meta-analysis of change in implicit bias. Retrieved from <http://psyarxiv.com/dv8tu>

Foschi, M. (2000). Double standards for competence: Theory and research. *Annual Review of Sociology, 26*, 21–42. doi:10.1146/annurev.soc.26.1.21

- Gaddis, S. M. (2015). Discrimination in the credential society: An audit study of race and college selectivity in the labor market. *Social Forces*, *93*(4), 1451–1479.
- Gendler, T. (2008). Alief and belief. *Journal of Philosophy*, *105*(10), 634–663.
- Gibson, C. E., Losee, J., & Vitiello, C. (2014). A replication attempt of stereotype susceptibility (Shih, Pittinsky, & Ambady, 1999). *Social Psychology*, *45*, 194–198.
- Gill, M. J. (2004). When information does not deter stereotyping: Prescriptive stereotyping can foster bias under conditions that deter descriptive stereotyping. *Journal of Experimental Social Psychology*, *40*(5), 619–632.
- Glass, C. R., & Westmont, C. M. (2014). Comparative effects of belongingness on the academic success and cross-cultural interactions of domestic and international students. *International Journal of Intercultural Relations*, *38*, 106–119.
- Goldin, C., & Rouse, C. (2000). Orchestrating impartiality: The impact of blind auditions on female musicians. *American Economic Review*, *90*(4), 715–741.
- Goodwin, G. P. (2015). Moral character in person perception. *Current Directions in Psychological Science*, *24*, 38–44.
- Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, *106*, 148–168.
- Greenwald, A. G., Banaji, M. R., & Nosek, B. A. (2015). Statistically small effects of the Implicit Association Test can have societally large effects. *Journal of Personality and Social Psychology*, *108*(4), 553–561.
- Guarino, C. M., & Borden, V. M. (2017). Faculty service loads and gender: Are women taking care of the academic family? *Research in Higher Education*, *58*, 672–694.
- Hall, E. V., Phillips, K. W., & Townsend, S. S. (2015). A rose by any other name? The consequences of subtyping “African-Americans” from “Blacks.” *Journal of Experimental Social Psychology*, *56*, 183–190.
- Hebl, M. R., Foster, J. B., Mannix, L. M., & Dovidio, J. F. (2002). Formal and interpersonal discrimination: A field study of bias toward homosexual applicants. *Personality and Social Psychology Bulletin*, *28*, 815–825.
- Heilman, M. E. (2012). Gender stereotypes and workplace bias. *Research in Organizational Behavior*, *32*, 113–135.
- Heilman, M. E., Martell, R. F., & Simon, M. C. (1988). The vagaries of sex bias: Conditions regulating the undervaluation, equivalence, and overvaluation of female job applicants. *Organizational Behavior and Human Decision Processes*, *41*(1), 98–110.
- Heilman, M. E., & Okimoto, T. G. (2007). Why are women penalized for success at male tasks? The implied communality deficit. *Journal of Applied Psychology*, *92*(1), 81–92.

Heilman, M. E., & Stopeck, M. H. (1985). Attractiveness and corporate success: Different causal attributions for males and females. *Journal of Applied Psychology, 70*(2), 379–388.

Heilman, M. E., & Wallen, A. S. (2010). Wimpy and undeserving of respect: Penalties for men's gender-inconsistent success. *Journal of Experimental Social Psychology, 46*(4), 664–667.

Heilman, M. E., Wallen, A. S., Fuchs, D., & Tamkins, M. M. (2004). Penalties for success: Reactions to women who succeed at male gender-typed tasks. *Journal of Applied Psychology, 89*, 416–427.

Hersch, J. (2008). Profiling the new immigrant worker: The effects of skin color and height. *Journal of Labor Economics, 26*(2), 345–386.

Inzlicht, M., & Schmader, T. (Eds.). (2012). *Stereotype threat: Theory, process, and application*. New York, NY: Oxford University Press.

Jones, K. P., Peddie, C. I., Gilrane, V. L., King, E. B., & Gray, A. L. (2016). Not so subtle a meta-analytic investigation of the correlates of subtle and overt discrimination. *Journal of Management, 42*(6), 1588–1613.

Jones, M. (2001). Stereotyping Hispanics and Whites: Perceived differences in social roles as a determinant of ethnic stereotypes. *Journal of Social Psychology, 131*(4), 469–476.

Jussim, L. (2017). Accuracy, bias, self-fulfilling prophecies, and scientific self-correction. *Behavioral and Brain Sciences, 40*, 1–65.

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux.

Kao, G. (2000). Group images and possible selves among adolescents: Linking stereotypes to expectations by race and ethnicity. *Sociological Forum, 15*, 407–430.

Kiefer, A., & Sekaquaptewa, D. (2007). Implicit stereotypes and women's math performance: How implicit gender-math stereotypes influence women's susceptibility to stereotype threat. *Journal of Experimental Social Psychology, 43*, 825–832.

Kugelmass, H. (2016). "Sorry, I'm not accepting new patients": An audit study of access to mental health care. *Journal of Health and Social Behavior, 57*(2), 168–183.

Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychological Review, 103*(2), 284–308.

Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J. L., Joy-Gaba, J. A., et al. (2014). A comparative investigation of 17 interventions to reduce implicit racial preferences. *Journal of Experimental Psychology. General, 143*, 1765–1785.

- Lai, C. K., Skinner, A. L., Cooley, E., Murrar, S., Brauer, M., Devos, T., et al. (2016). Reducing implicit racial preferences: II. Intervention effectiveness across time. *Journal of Experimental Psychology. General*, *145*(8), 1001–1016.
- Leslie, S. J., Cimpian, A., Meyer, M., & Freeland, E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science*, *347*(6219), 262–265.
- Levy, N. (2015). Neither fish nor fowl: Implicit attitudes as patchy endorsements. *Noûs*, *49*(4), 800–823.
- Livingston, R. W., Rosette, A. S., & Washington, E. F. (2012). Can an agentic Black woman get ahead? The impact of race and interpersonal dominance on perceptions of female leaders. *Psychological Science*, *23*(4), 354–358.
- Lubienski, S. T., Robinson, J. P., Crane, C. C., & Ganley, C. M. (2013). Girls' and boys' mathematics achievement, affect, and experiences: Findings from ECLS-K. *Journal for Research in Mathematics Education*, *44*(4), 634–645.
- Maddox, K. B. (2004). Perspectives on racial phenotypicality bias. *Personality and Social Psychology Review*, *8*(4), 383–401.
- Madva, A., & Brownstein, M. (in press). Stereotypes, prejudice, and the taxonomy of the implicit social mind. *Noûs*.
- Mandelbaum, E. (2016). Attitude, inference, association: On the propositional structure of implicit bias. *Noûs*, *50*(3), 629–658.
- Martell, R. F., Lane, D. M., & Emrich, C. (1996). Male-female differences: A computer simulation. *American Psychologist*, *51*, 157–158.
- Martin, C. L., & Halverson, C. F. (1987). The role of cognition in sex role acquisition. In D. B. Carter (Ed.), *Current conceptions of sex roles and sex typing: Theory and research* (pp. 123–137). New York, NY: Praeger.
- Martinez, S., & Toutkoushian, R. K. (2014). Decomposing the differences in time allocation and research output between Latino and non-Latino White faculty. *Latino Studies*, *12*(4), 566–595.
- MacNell, L., Driscoll, A., & Hunt, A. N. (2015). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education*, *40*(4), 291–303.
- McCauley, C., Stitt, C. L., & Segal, M. (1980). Stereotyping: From prejudice to prediction. *Psychological Bulletin*, *87*(1), 195–208.
- Mendes, W. B., Blascovich, J., Lickel, B., & Hunter, S. (2002). Challenge and threat during social interactions with White and Black men. *Personality and Social Psychology Bulletin*, *28*(7), 939–952.

- Merritt, A. C., Effron, D. A., & Monin, B. (2010). Moral self-licensing: When being good frees us to be bad. *Social and Personality Psychology Compass*, 4, 344–357.
- Merton, R. K. (1988). The Matthew effect in science. II: Cumulative advantage and the symbolism of intellectual property. *Isis*, 79(4), 606–623.
- Milkman, K. L., Akinola, M., & Chugh, D. (2012). Temporal distance and discrimination: An audit study in academia. *Psychological Science*, 23(7), 710–717.
- Miller, D. T., Taylor, B., & Buck, M. L. (1991). Gender gaps: Who needs to be explained? *Journal of Personality and Social Psychology*, 61(1), 5–12.
- Mitchell, S. M., & Hesli, V. L. (2013). Women don't ask? Women don't say no? Bargaining and service in the political science profession. *PS, Political Science & Politics*, 46(02), 355–369.
- Mondschein, E. R., Adolph, K. E., & Tamis-LeMonda, C. S. (2000). Gender bias in mothers' expectations about infant crawling. *Journal of Experimental Child Psychology*, 77(4), 304–316.
- Monin, B., & Miller, D. T. (2001). Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology*, 81, 33–43.
- Moon, A., & Roeder, S. S. (2014). A secondary replication attempt of stereotype susceptibility (Shih, Pittinsky, & Ambady, 1999). *Social Psychology*, 45, 199–201.
- Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences of the United States of America*, 109(41), 16474–16479.
- Moss-Racusin, C. A., Phelan, J. E., & Rudman, L. A. (2010). When men break the gender rules: Status incongruity and backlash against modest men. *Psychology of Men & Masculinity*, 11(2), 140–151.
- Mujcic, R., & Frijters, P. (2013). Still not allowed on the bus: It matters if you're Black or White! (Institute for the Study of Labor [IZA] Discussion Paper No. 7300). Bonn, Germany: IZA.
- Murphy, M. C., Steele, C. M., & Gross, J. J. (2007). Signaling threat: How situational cues affect women in math, science, and engineering settings. *Psychological Science*, 18(10), 879–885.
- Nguyen, H. H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology*, 93(6), 1314–1334.
- Norton, M. I., Vandello, J. A., & Darley, J. M. (2004). Casuistry and social category bias. *Journal of Personality and Social Psychology*, 87, 817–831.

O'Meara, K., Kuvaeva, A., & Nyunt, G. (2017). Constrained choices: A view of campus service inequality from annual faculty reports. *Journal of Higher Education, 88*(5), 672–700. doi:10.1080/00221546.2016.1257312.

O'Meara, K., Kuvaeva, A., Nyunt, G., Waugaman, C., & Jackson, R. (2017). Asked more often: Gender differences in faculty workload in research universities and the work interactions that shape them. *American Educational Research Journal*. Advance online publication. doi:10.3102/0002831217716767

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2015). Using the IAT to predict ethnic and racial discrimination: Small effect sizes of unknown societal significance. *Journal of Personality and Social Psychology, 108*(4), 562–571.

Page, S. E. (2007). *The difference: How the power of diversity creates better groups, firms, schools, and societies*. Princeton, NJ: Princeton University Press.

Paustian-Underdahl, S. C., Walker, L. S., & Woehr, D. J. (2014). Gender and perceptions of leadership effectiveness: A meta-analysis of contextual moderators. *Journal of Applied Psychology, 99*(6), 1129–1145.

Phelan, J. E., Moss-Racusin, C. A., & Rudman, L. A. (2008). Competent yet out in the cold: Shifting criteria for hiring reflect backlash toward agentic women. *Psychology of Women Quarterly, 32*, 406–413.

Philogène, G. (1999). *From Black to African American: A new social representation*. Westport, CT: Praeger/Greenwood.

Pittinsky, T. L., Shih, M., & Ambady, N. (1999). Identity adaptiveness: Affect across multiple identities. *Journal of Social Issues, 55*, 503–518.

Prime, J. L., Carter, N. M., & Welbourne, T. M. (2009). Women “take care,” men “take charge”: Managers' stereotypic perceptions of women and men leaders. *Psychologist Manager Journal, 12*(1), 25–49.

Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin, 28*(3), 369–381.

Ratcliff, N. J., Vescio, T. K., & Dahl, J. L. (2015). (Still) waiting in the wings: Group-based biases in leaders' decisions about to whom power is relinquished. *Journal of Experimental Social Psychology, 57*, 23–30.

Räty, H., Vänskä, J., Kasanen, K., & Kärkkäinen, R. (2002). Parents' explanations of their child's performance in mathematics and reading: A replication and extension of Yee and Eccles. *Sex Roles, 46*(3–4), 121–128.

Richards, Z., & Hewstone, M. (2001). Subtyping and subgrouping: Processes for the prevention and promotion of stereotype change. *Personality and Social Psychology Review, 5*, 52–73.

Richeson, J. A., & Shelton, J. N. (2007). Negotiating interracial interactions: Costs, consequences, and possibilities. *Current Directions in Psychological Science*, *16*, 316–320.

Ridgeway, C. L. (2011). *Framed by gender: How gender inequality persists in the modern world*. Oxford, UK: Oxford University Press.

Rigney, D. (2010). *The Matthew effect: How advantage begets further advantage*. New York, NY: Columbia University Press.

Rosette, A. S., Leonardelli, G. J., & Phillips, K. W. (2008). The White standard: Racial bias in leader categorization. *Journal of Applied Psychology*, *93*(4), 758–777.

Rudman, L. A., & Fairchild, K. (2004). Reactions to counterstereotypic behavior: The role of backlash in cultural stereotype maintenance. *Journal of Personality and Social Psychology*, *87*, 157–176.

Rudman, L. A., & Phelan, J. E. (2008). Backlash effects for disconfirming gender stereotypes in organizations. *Research in Organizational Behavior*, *28*, 61–79.

Ryan, C. S., Casas, J. F., & Thompson, B. K. (2010). Interethnic ideology, intergroup perceptions, and cultural orientation. *Journal of Social Issues*, *66*(1), 29–44.

Ryan, C. S., Hunt, J. S., Weible, J. A., Peterson, C. R., & Casas, J. F. (2007). Multi-cultural and colorblind ideology, stereotypes, and ethnocentrism among Black and White Americans. *Group Processes & Intergroup Relations*, *10*(4), 617–637.

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, *115*(2), 336–356.

Self, W. T., Mitchell, G., Mellers, B. A., Tetlock, P. E., & Hildreth, J. A. D. (2015). Balancing fairness and efficiency: The impact of identity-blind and identity-conscious accountability on applicant screening. *PLoS One*, *10*(12), e0145208.

Sigelman, L., Tuch, S. A., & Martin, J. K. (2005). What's in a name? Preference for "Black" versus "African-American" among Americans of African descent. *Public Opinion Quarterly*, *69*(3), 429–438.

Skiba, R. J., Horner, R. H., Chung, C. G., Rausch, M. K., May, S. L., & Tobin, T. (2011). Race is not neutral: A national investigation of African American and Latino disproportionality in school discipline. *School Psychology Review*, *40*(1), 85–107.

Skiba, R. J., Michael, R. S., Nardo, A. C., & Peterson, R. L. (2002). The color of discipline: Sources of racial and gender disproportionality in school punishment. *Urban Review*, *34*(4), 317–342.

Spence, J., & Buckner, C. (2000). Instrumental and expressive traits, trait stereotypes, and sexist attitudes: What do they signify? *Psychology of Women Quarterly*, *24*(1), 44–62.

- Spence, J., & Helmreich, R. (1978). *Masculinity and femininity: Their psychological dimensions, correlates and antecedents*. Austin, TX: University of Texas Press.
- Spencer, S. J., Logel, C., & Davies, P. G. (2016). Stereotype threat. *Annual Review of Psychology, 67*, 415–437.
- Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*(1), 4–28.
- Steele, C. M. (2011). *Whistling Vivaldi: And other clues to how stereotypes affect us*. New York, NY: Norton.
- Swim, J. K., Aikin, K. J., Hall, W. S., & Hunter, B. A. (1995). Sexism and racism: Old-fashioned and modern prejudices. *Journal of Personality and Social Psychology, 68*, 199–214.
- Swim, J. K., Mallett, R., Russo-Devosa, Y., & Stangor, C. (2005). Judgments of sexism: A comparison of the subtlety of sexism measures and sources of variability in judgments of sexism. *Psychology of Women Quarterly, 29*, 406–411.
- Tsay, C. J. (2013). Sight over sound in the judgment of music performance. *Proceedings of the National Academy of Sciences of the United States of America, 110*(36), 14580–14585.
- Tyler, J. M., & McCullough, J. D. (2009). Violating prescriptive stereotypes on job resumes: A self-presentational perspective. *Management Communication Quarterly, 23*, 272–287.
- Uhlmann, E. L., & Cohen, G. L. (2005). Constructed criteria: Redefining merit to justify discrimination. *Psychological Science, 16*, 474–480.
- Uttl, B., White, C. A., & Gonzalez, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation, 54*, 22–42.
- Valian, V. (1998). *Why so slow? The advancement of women*. Cambridge, MA: MIT Press.
- Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin, 140*(4), 1174–1204.
- Walton, G., & Cohen, G. (2007). A question of belonging; Race, social fit, and achievement. *Journal of Personality and Social Psychology, 92*(1), 82–96.
- Walton, G., & Cohen, G. (2011). A brief social-belonging intervention improves academic and health outcomes of minority students. *Science, 331*, 1447–1451.
- West, R. F., Meserve, R. J., & Stanovich, K. E. (2012). Cognitive sophistication does not attenuate the bias blind spot. *Journal of Personality and Social Psychology, 103*(3), 506–519.

Williams, J., & Smith, J. (2015). The myth that academic science isn't biased against women. *Chronicle of Higher Education*. <http://www.chronicle.com/article/The-Myth-That-Academic-Science/231413>

Williams, W. M., & Ceci, S. J. (2015). National hiring experiments reveal 2:1 faculty preference for women on STEM tenure track. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(17), 5360–5365.

Yee, D., & Eccles, J. (1988). Parent perceptions and attributions for children's math achievement. *Sex Roles*, *19*, 317–333.

Zhang, Q. (2015). Perceived ingroup stereotypes, threats, and emotions toward Asian Americans. *Howard Journal of Communications*, *26*(2), 115–131.

Zuckerman, H. (1997). Accumulation of advantage and disadvantage: The theory and its intellectual biography. In C. Mongardini & S. Tabboni (Eds.), *Robert K. Merton and contemporary sociology* (pp. 139–161). London, UK: Routledge.

