

6 Evaluating Job Candidates: Choosing the Short List and Treating Interviewees Equitably

Once the recruitment phase has ended, the evaluation phase begins. In this chapter we focus on the process that occurs after the applications are in—the process of evaluation that eventually leads to a new faculty hire. In this context, everyone wants to hire the “best” candidate. Everyone also wants to evaluate people fairly. Most people pride themselves on being impartial and objective observers of other people’s competence, independent of the person’s gender or race. As we have argued, however, considerable research suggests that evaluations are subject to error. We have outlined experiments that demonstrate that people—from undergraduates to practicing scientists—are likely to overrate men’s competence and underrate women’s, unless the woman’s performance was extraordinary. People are similarly likely to overrate Whites’ competence and underrate that of some racial-ethnic minority groups. *Wanting* to be gender and race neutral does not, by itself, prevent people from making decisions influenced by gender and race schemas. We start with an example of the different attributions people can make for the same behavior, depending on the gender of those being observed:

One man said he had noticed that when he saw a paper with a young male scientist as the first author and the man’s male supervisor as the second author, he presumed the young man had contributed the main ideas and had led the research. The young man had earned his place as first author. Otherwise, the supervisor would have put his name first. But when he saw a young female scientist as the first author, he assumed that the male supervisor was doing her a favor. Maybe she had earned the position, but equally likely, her mentor was trying to help her.

We like to think that people’s records of accomplishment do not really require interpretation, but in fact we make many interpretations along with our judgments. With someone’s CV in hand, we think we can judge his or

her merits objectively. But we need to consider how we go about assessing merit.

We will argue here that a confluence of factors—gender and race schemas, age schemas, socioeconomic origin schemas, a bias toward high prestige universities, and a superfluity of candidates—results in fewer women and people of color being considered than would be appropriate. Establishing merit is not straightforward, so we need to adopt practices that will increase the likelihood that our judgments are fair and unbiased.

Fairness in Evaluations

The evaluation process divides into four rough stages: initial screening of candidates, construction of a short list (or medium-short list), selection of interviewees from the short list, and decision-making about candidates after the interviews. Some issues will play a role throughout the process. For example, at the initial screening and at all subsequent points, a preference for individuals with a PhD from a high-prestige institution will result in underrating women (slightly) and people of color (more). Another issue throughout will be the likely slight overrating of White men's credentials compared to those of women and people of color. Some issues, however, only arise later in the process, depending on how the institution conducts its searches.

At the interview stage, faculty beyond the search committee are involved: they go to the candidate's talk and may meet with the candidate. For those faculty, who may not have read any of the candidate's work, their impression is based on a very small sample of behavior, much of which is irrelevant to the work itself, such as the faculty member's physical appearance, manner of presentation, and response to questions. Since faculty tend to have high confidence in their evaluation decision, their reliance on interview performance leads to giving insufficient weight to years of past performance and too much weight to a small slice of behavior.

Although everyone wants to be fair (and usually believes they are fair), several factors make fairness somewhat difficult to attain. We review some of those factors here. The belief that it is possible to make gender- and race-neutral decisions can get people in trouble. Some people take their commitment to merit to mean that, *using the criteria that are currently in place*, everyone should be judged in exactly the same way, regardless of

their sex or the color of their skin. We too subscribe to the merit principle, and we too subscribe to the principle of treating people similarly, but we do not endorse the idea that our current criteria are the best ones, nor do we endorse the idea that current notions of similar treatment will maximize everyone's potential or be equally revealing.

We have two goals in this section. One is to spell out the problems with using the criteria that are currently in place. The other is to question whether a commitment to the merit principle ensures following it.

Using the Criteria Currently in Place

For any job, if that job has primarily been held by one type of person, with general approval, observers will tend to see those characteristics as necessary for the job. In fact, one way of trying to determine whether individuals are a good fit for a job is by giving them a test to see if they share the interests and characteristics of people currently in the job (Nye, Su, Rounds, & Drasgow, 2012). The logical problem with this is obvious (Valian, 2014). There might be many satisfactory ways to do a job, only a subset of which are displayed by current jobholders. For example, faculty can be surprised by how well a colleague performs as a chair, noticing that his or her focus and style are very different from a previous admired chair's, but equally, if not more, beneficial to a department. Using criteria derived from one set of successful occupants of a position to evaluate new candidates tilts a faculty against innovation.¹ If a prior occupant who was a woman or a member of a minority group was not successful, illusory correlation can bias observers against others who are demographically similar to the person who failed.

Ratings of job performance depend on the raters' conceptions of what the job requires. Those ratings in turn depend in part on the characteristics of the people in the job. A rater can use standards that are irrelevant but in place because they were at one time typical of most successful jobholders (Valian, 1998, 2014). Some criteria that might look right for the job of university professor—ambition, self-promotion, competitiveness, assertiveness—are characteristics that are double-edged swords for White women and minority women and men. Women who are ambitious, self-promoting, competitive, and assertive are seen negatively by both men and other women, as we have discussed in chapter 3. A further point is that as the potential pool for an occupation slowly becomes less demographically homogeneous, raters run the risk of taking extraneous properties that

the original, demographically similar, jobholders shared as relevant to job performance. Having seen only a limited number of ways of solving a problem, raters may see those ways as the sole ways to solve a problem. If, for example, a long line of department chairs had a decisive and assertive style, a candidate who was known for being cooperative and likable might not be seen as sufficiently “tough.”

One senior professor told us about trying to interest her colleagues in hiring an extremely talented woman at another university. Her colleagues said the woman was “difficult” and were opposed to pursuing her. Later in the year one of her male colleagues proposed hiring an extremely talented man at another university. Our senior professor said that the man was “difficult.” Her male colleagues said, “We can work around that.” People notice that difficult men are difficult, but their being difficult is not a deal breaker. For women, being “difficult” is a deal breaker, as shown by work on backlash. And that assumes that the same criteria were used for deciding that these two people were difficult. The result for women is that being a good fit is complicated. Even if a woman meets the informal criteria by which men are judged, she will not be seen as positively as a man is.

Setting the Stage for Shifting Standards

In chapter 3 we reviewed experiments demonstrating that people shift their evaluation standards, depending in part on their prior beliefs about who will do a better job in a particular area. Men look right for the job of construction engineer, with the result that whatever attributes that men might possess to a greater degree than women will give men an advantage. Those attributes will come to look as if they are necessary for the job, whether they are actually relevant or not.

Because there are often competing ideas about what a department needs, the stage is set for shifting standards to play a role. It is in the department’s interest to determine impartially who will best satisfy the department’s needs and interests, but not everyone has the same ideas about what a department’s needs and interests are. Even if there is agreement about the overall needs and interests, people differ in which values they weigh more heavily than others. Some may care only about the likelihood that a faculty member will be a star. Some may see a new area as the wave of the future and others may see it as currently fashionable but unlikely to last. Some might value filling an instructional gap more than

strengthening an existing area. Some might prefer a new hire if they are likely to collaborate with that person or if they think the new person is likely to support their views about desirable future directions of the department. To the extent that people are successful in arguing for their candidate, they gain power and advantage in their department—another form of the accumulation of advantage. Having a job description (the creation of which can involve difficult conversations, as we detail in chapter 5) is not the same thing as agreeing on what the department needs or *should* need.

Those different preferences set the stage for people to adopt shifting standards. If likelihood of collaboration is the top priority for one member of a search committee, she may ignore information, such as number of publications, that she might otherwise use to evaluate a group of people all or none of whom she might collaborate with. Thus, even before issues of gender and ethnicity come into play, search committee members are likely to differ in what is important to them in a candidate and to use as a measuring rod whatever qualities the person they want happens to have. Only an agreed-upon set of criteria can provide a buffer to shifting standards.

Gender Issues

The combination of gender schemas and, in some fields, much smaller numbers of women than men applicants, can make it very difficult for women to get to the interview stage. Other factors unique to women, such as motherhood, underline women's identity as women, with the result that mothers are perceived more negatively than childless women and are less likely to be offered as good a job (Correll, Benard, & Paik, 2007; Correll, Kelly, O'Connor, & Williams, 2014; Heilman & Okimoto, 2008). In line with gender schemas, it appears to be difficult for us to imagine that mothers can be committed to their jobs. "Working mothers" have a special title, while "working fathers" do not. And if mothers are seen as exceptionally good professionals, they appear to correspondingly be seen as bad parents (Okimoto & Heilman, 2012) and can be viewed negatively for that reason.

Race–Ethnicity Issues

For underrepresented minorities, the situation is even more extreme. As we mentioned in chapter 3, field studies have found that job applicants with names that sounded African American, such as Lakisha, were less likely to be called for an interview than were applicants with names that sounded

“White,” such as Emily (Bertrand & Mullainathan, 2003). There is little reason to think that search committees in academia are less subject to such influences than managers in the general workplace.

Parental and Partner Statuses and Their Role in the Evaluation Process

We indicated in chapter 5 that institutions do well to signal their family-friendly policies to applicants. They should not, however, include consideration of applicants’ family situation in the hiring process. As one résumé study demonstrated, parental status plays a different role in the evaluation of men compared to women (Correll et al., 2007). For women it results in perceptions of lesser suitability for compensation, advancement, and leadership; for men it results in perceptions of greater suitability for the same outcomes. In short, there appears to be a “motherhood penalty” and a “fatherhood bonus.” For this reason, we recommend that job applicants *not* signal their parenthood status in their applications, and that search committees avoid requesting or permitting discussion of any information they may have about either the parenthood or the partner status of applicants. It is illegal to ask for such information.

Often search committees are tempted to discuss the partner status of their applicants, believing that it is important to take account of it in the process of selecting candidates if they are to be successful. While this thinking is logical from the perspective of the hiring institution, acting on it is likely to trigger both gender and parental schemas that may lead committees to underestimate the qualifications of women and overestimate those of men, as we have just described. For an example, see the kinds of schemas expressed by department members in departments that fail to hire women even in the presence of an institutional commitment to doing so (Stewart, Malley, & Herzog, 2016).

Because of the difficulty of remaining unaffected by personal knowledge about applicants, we recommend that information about candidates’ personal lives be neither sought nor considered in the course of the selection process. Whether one penalizes mothers (or mothers-to-be) or benefits fathers or fathers-to-be (because they are seen as more serious and responsible, and as needing to do well at their job to provide for their family), it is hard to undo the effects of irrelevant information. It is like telling a jury to disregard information that is not strictly relevant. A qualitative study of twenty-four search committee deliberations at a research-intensive university found that

search committees tended to assume that a woman would not accept a position if she had a partner, but that a man would (Rivera, 2017). Women's relationship status was discussed more often than men's, and committee members believed that it was legal to use such information as long as they had not solicited it from the candidate (which it is not). Personal questions also affect applicants' views of the institution. One study of individuals who withdrew from searches late in the process or turned down offers demonstrated that interpretations of even small comments by faculty members in recruiting departments led to counterproductive interpretations of the underlying biases in the departments by candidates, even if those comments were not the direct cause of candidates' withdrawing (<http://advance.umich.edu/resources/STEMTurnDownStudy-2009.pdf>).

Commitment in Principle versus Commitment in Practice

Commitment to the merit principle as a principle does not entail abiding by it, as we detailed in chapter 3 when we described how schemas work. Worse, the fact that everyone explicitly subscribes to the principle can impair people's ability to discern the extent to which they do not act in accord with it. The dual experience of holding a principle and of having some data that seem to suggest that one abides by it can close one's eyes to data that suggest this isn't true. Everyone wants to think that they act fairly, leading to their paying more attention to confirming data and less attention to disconfirming data.

We reviewed experimental results on moral licensing in chapter 3. A "license" to act contrary to one's explicit beliefs arises if one has recently performed an action that seems to be in accord with those beliefs. We permit ourselves to lapse into more stereotypic thinking if we have recently performed an action that seems fair and nonstereotyped to us. Similarly, it appears that doing any task that reminds us of what a sterling character we are—such as writing a description of ourselves using morally positive words provided by the experimenter—makes us less generous afterward than does writing a story that reminds us of our negative traits. When we write positively about ourselves, we are less generous in giving money to a worthy cause. When we write negatively about ourselves, we are more generous (Sachdeva, Iliev, & Medin, 2009). The effect is not simply due to writing positively or negatively since it does not occur when people write positively or negatively about someone else, only when they write about themselves.

The effect of moral licensing is not limited to an individual's own behavior. Identifying with a group that one believes is moral can provide license to the individual to behave less morally. Moral licensing can even work vicariously (Kouchaki, 2011). In one of a series of studies with undergraduates, participants in the experimental condition were told that students at their university were more moral than students at other universities (Kouchaki, 2011). In the three control conditions, participants were told that students at their university were not more moral than students at other universities, or that they were more competitive, or that they were more intelligent. All participants then read a scenario in which they were a police chief hiring a new officer in a small town and had to judge whether a White or an African American officer would be better for the job. Those in the more moral and more intelligent groups were more likely to choose the White officer over the African American one than were those who were told that students at their university were more competitive or told they were not more moral (Kouchaki, 2011). Although it appears paradoxical, our reminders to ourselves of our positive ethical qualities make us freer to act selfishly and freer to act in ways that are consistent with group stereotypes.

Strategies to Reduce Schemas' Effects

Two types of strategies can help reduce the effects of schemas (see Monteith, Ashburn-Nardo, Voils, & Czopp, 2002, and Moskowitz & Li, 2011, for reviews and discussion). One strategy acknowledges that schemas are easily activated and concentrates on reducing the biased decisions that one might make on the basis of the schemas (e.g., Devine, Forscher, Austin, & Cox, 2012; Forscher, Mitamura, Dix, Cox, & Devine, 2017; Monteith et al., 2002). In some cases the strategy requires training to develop ways of countering schemas, such as learning to think of a person in terms of his or her individual characteristics rather than in terms of his or her group membership (e.g., Carnes, Devine, Baier Manwell, Byars-Winston, Fine, Ford, et al., 2015; Devine et al., 2012). In other cases the strategy might be to explicitly call one's goals to mind—for example, *I want to be fair*—or reflect on times when one didn't act in accord with one's meritocratic aspirations. Those strategies may help neutralize the consequences of the schema. In these *reactive* strategies (Moskowitz & Li, 2011), acknowledging that one responds on the basis of schemas is a precondition for controlling the behavior that might arise from the almost automatic activation of schemas. The two-part

model that is invoked here—acknowledge the presence of schemas, and act on controlling them—requires *not* patting oneself on the back for one's principles but acknowledging that one may not always live up to them.

The second type of strategy suggests that it is possible to prevent activation of a schema to begin with, through being routinely motivated to avoid prejudice (Moskowitz & Li, 2011). Individuals with “chronic” goals of that sort are less likely to respond in terms of schemas when judging individuals. Goals to avoid prejudice can also be developed by encouraging someone to recall a failure in living up to a meritocratic goal. Individuals (primarily White) who were primed to recall such a failure showed a reduced tendency to associate African Americans with negative traits, such as *lazy* (Moskowitz & Li, 2011). A “proactive” strategy of this sort reminds people of an instance when they have acted counter to their principles and thus removes the moral license that thinking well of themselves would otherwise bestow.

Is It Good to Be Race “Blind”?

Whites' commitment to a race-blind strategy can backfire. A White person refusing to use race as a category when it is relevant can affect African Americans negatively, rather than being neutral or positive. In one study, Whites were asked to pick out a target photo from an array of photos of African Americans and Whites. They had a partner who knew who the target was. They were to ask their partner questions that would allow them to find the relevant photo in the shortest amount of time. Even though an African American's ethnicity would help to quickly identify the person, Whites tended not to ask about the race of the target person, especially if their partner was African American. Whites wanted to appear race blind (see review in Apfelbaum, Norton, & Sommers, 2012). However, African Americans saw Whites who avoid mentioning race as more biased than those who did mention it—when the mention was appropriate. Most Whites have probably had the experience of failing to mention someone's racial or ethnic identity from a desire for race not to matter. But if you are a member of an underrepresented group, that refusal to recognize the obvious can simply seem insulting.

A commitment to being race blind might make it difficult for Whites to see discrimination when it exists. A study with children suggests how that might work. Children ages 8 to 11 who were provided with different

assumptions in thinking about race differences either perceived or failed to perceive actual discrimination, depending on their assumption. The children heard one of two stories about a teacher who wanted to ensure racial equality in her class by having a class performance, either via a color-blind approach or by means of an approach affirming the value of differences (Apfelbaum, Pauker, Sommers, & Ambady, 2010). Then the children heard different vignettes about forms of exclusion or mistreatment and were asked to evaluate them. In one case, the vignette was ambiguous—a White child didn't invite an African American child to a party. That example is very like what happens to underrepresented minorities and White women in departments—they aren't invited to collaborate, they aren't invited to lunch or coffee, they are at the margins—even though there may be no overt intention to exclude. In another case, the vignette was clear. A White child tripped an African American child in order to take a ball away from him during a game and justified his behavior by saying that the African American child played rough because he was African American.

Children were asked to describe each vignette in their own words and to say whether discrimination had taken place (Apfelbaum et al., 2010). The children who heard the color-blind story were much less likely to detect discrimination, even when there was strong overt evidence of it, than were the children who heard the story emphasizing that different races have value. The study suggests that an assumption that discrimination is unlikely can lead to a failure to see it when it occurs. Thus, we caution search committees to remember that race, ethnicity, and gender can affect our judgments, even if we wish they did not, and to realize that congratulating oneself for being gender and race blind can open the door to making biased judgments. The most helpful attitude is to assume at least skepticism about one's ability to make judgments that are unaffected by someone's race, ethnicity, and gender.

Evaluating Applicants to Create a Short List

The First Pass

The first categorization of candidates' applications is usually made on the basis of (1) information that can be gleaned from a curriculum vitae (CV); (2) information that the evaluator may already have about applicants from meeting them, hearing them present their work, or hearing about them

from someone else; and (3) letters of recommendation. (Recommendation letters are not always requested as part of someone's initial application packet. Medical schools, for example, often do not request recommendation letters until after the search committee has created a "long" short list.) CVs include the names of the institutions where the person has received his or her training, and, often, the names of the people the person trained with; a list of publications; a list of conference and other presentations; a list of grants, prizes, and awards; and, often, teaching experience and administrative service. The candidate's cover letter, teaching statement, and research statements, if they exist, may also be consulted, especially the cover letter. Some searches may have 50 candidates; others may have 400. Some search committees may have only three members; others may have six or more. A search committee member who is faced with a multitude of applications will look for ways to extract the maximum amount of information with the minimum amount of effort.

At universities and colleges that put a premium on research, research accomplishment is the single most important criterion. That criterion can be evaluated in terms of a person's publications and in terms of his or her success in obtaining funding for research. In the natural sciences, most publications take the form of articles in journals that only accept papers for publication if they have been successfully reviewed by peers. In the social sciences and the humanities, depending on the field, publications may take the form of journal articles, books, book chapters, or all three. At institutions that put a premium on teaching, research will often still be relevant, but teaching experience and dedication will play an important role as well.

The densest and most substantive information about the quality of someone's research is in the publications themselves. But it is time-consuming to read individuals' work when a search committee is faced with 50 applicants, and even more time-consuming when a committee must review 400 applicants. In addition, many search committees, especially in smaller departments, are composed of members with varying degrees of knowledge in different subfields. Committees will thus be tempted to use at the outset markers like prestige of the person's PhD-granting institution or prestige of the person's primary supervisor as criteria for keeping or eliminating candidates. We take up these issues again in chapter 9, where we consider evaluations of faculty for tenure and promotion.

Relying on Proxies in Assessing Quality

When search committee members base their classifications of applicants on CVs and letters of recommendation instead of reading their work directly, they are using proxy information. The information does not directly provide information about the quality of the person's scholarship—how the candidate presents the background to their research and frames the problem or question; the candidate's methods and materials, the research plan, treatment of the raw material (texts or data), conclusions drawn and inferences made about how the analysis changes one's understanding of the initial problem. Direct information about how the candidate approaches a problem is only available through reading the person's work or, to a lesser extent, through hearing the candidate present the work. Similarly, direct information about the person's teaching is typically acquired through proxy information unless the candidate provides an example of teaching.

Time constraints lead search committee members to begin with proxy information—the CV and the judgments of other people. While the CV is a factual record of a candidate's achievements, achievements are themselves dependent not just on the candidate but on other people. Journal editors decide who the reviewers of a paper will be. The editors' knowledge of the authors, or of the authors' institution, or reputation, may influence their choice of reviewers. Journal reviewers, in their turn, may be influenced not only by the paper itself but also by prestige factors—to the extent that they are known or deducible—such as the prestige of a coauthor or the authors' institution. Some journals practice blind review—where reviewers are not told the identity of the author(s), but identity can often be deduced. Search committee members are sensitive to the prestige of the journals in which the candidate has published and the prestige of the candidate's PhD and postdoctoral supervisors.

Relying on proxy information substitutes a history of other people's views of the candidate for one's own independent evaluation. If one believes that all the preceding processes have been impartial and objective, and if one believes that everyone has had the same opportunity to attend the same stellar schools, one will have no qualms about proxy information. If, however, one believes that factors other than merit can influence where someone applies to graduate school, where someone is accepted to be a student, and with whom one studies once at graduate school, one will be less sanguine about the validity of relying on proxy information. Similarly, if

one believes that each achievement is independent of prior achievements, one will assess someone's record differently than if one appreciates the power of the accumulation of advantage.

In general, relying on markers of prestige will advantage men over women, and Whites, Asians, and Hispanics over African Americans, because a lower percentage of women than men, and a lower percentage of African Americans than every other ethnic or racial group, have markers of prestige in their histories. Markers of prestige are more difficult for some groups to acquire, and the lack of those markers then makes further markers harder to acquire. Thus some accumulate advantage, while others accumulate disadvantage.

Search committees have a decision to make at the outset. If they want to maximize inclusiveness, they will have to look beyond proxy information because standard proxies for merit generally favor men over women and favor every other race and ethnic group over African Americans. In some cases the disparities are small, but as Gompers and Wang (2017) point out when discussing the lack of women in venture capital, even a small tendency toward homophily will have an effect on binary outcomes. The same result, we suggest, will hold generally for small disparities.

Prestige of Institution

The prestige of the institution where the candidate received their degree or completed their postdoctoral training or had their first professorial job is objective, in the sense that there is good agreement among academics about which institutions are prestigious; the predictors of prestige are known.² But if search committees rely on prestige of degree-granting institution to reduce their applicant pool, they are more likely to eliminate African Americans than Hispanics, Whites, or Asian Americans. In 2015, only 50% of African Americans' PhDs were from universities with very high research activity (as rated by the Carnegie Foundation for the Advancement of Teaching, 2005), compared to 72% of Whites', 73% of Hispanics', and 77% of Asians'. Similarly, committees will be more likely to eliminate women than men: also in 2015, 68% of women and 74% of men received their degrees from the most prestigious institutions. All comparisons, save the difference between Whites and Hispanics, are significant.³ Putting a premium on prestige of degree-granting institution will result in fewer African Americans and women.

A search committee member attending to prestige might object that prestige of prior institution is diagnostic of future success, as it is and could hardly fail to be, since advantages accumulate, and there are more advantages available at high-prestige institutions. But it is hard to disentangle the effects of individual accomplishment from institutional prestige. As long as highly prestigious institutions preferentially employ people from other highly prestigious institutions, which they do (for political science, see Oprisko, Dobbs, & DiGrazia, 2013; for computer science, business, and history, see Clauzet, Arbesman, & Larremore, 2015), it will be impossible to know whether someone with a degree from a less prestigious institution would fare well if given the resources of a prestigious institution.

Another reason that it is difficult to assess the diagnosticity of a prestigious degree is that success is not just an outcome of the tangible resources that institutions supply more or less of, but an outcome of the professional networks that an individual has access to in different types of institutions. Individuals who graduate from nonprestigious institutions not only have fewer tangible resources at their disposal but also have networks that are less densely populated with influential people, compared to individuals whose degrees are from prestigious institutions.

An analysis of job success in sociology makes plain how prestige works in at least one field (Headworth & Freese, 2015). Sociology departments were placed into four prestige tiers, based on survey answers by department chairs and directors of graduate programs. Over the period from 2004–2005 to 2008–2009, 2,644 PhDs were granted. About 60% (1,617) of those graduates came from the two lower prestige tiers and about 40% from the top two tiers. Overall, graduates of the most prestigious schools had more advantages; as students, for example, they published more. Students from the top two tiers were also much better represented as authors in sociology's two flagship journals than were students from the bottom two tiers. Such achievements were presumably facilitated by access to resources and knowledgeable faculty (though access may not have been allocated equally to students from all demographic backgrounds).

Throughout their PhD training, then, more advantage accumulated for students in the top-tier institutions, resulting in a higher likelihood of obtaining a desirable job. Perhaps the saddest finding was that of the 1,617 students from the two lowest tiers, only two were hired by schools in the top tier. None of the 1,133 students from the lowest tier were hired by a

school at the top tier, and only three were hired by a school at the second-highest tier. That contrasts with the fate of the 663 individuals who graduated with a degree from the top tier—30 of them took a job at a top-tier institution, and another 69 went to the second tier (Headworth & Freese, 2015). Recall that African Americans, in particular, are more likely to get a PhD from a lower tier institution.

In fields where postdoctoral training is the norm, such as biology, prestige can again play a role. An analysis of biology postdoctoral placements found that high-achieving male faculty had fewer women in their labs than did high-achieving female faculty (of whom there are fewer) or other faculty (Sheltzer & Smith, 2014). To the extent that elite male researchers have laboratories that are male dominated, a hiring preference among search committees for individuals with training from elite investigators will be male dominated. There is no reason to assume explicit gender bias on the part of elite investigators or on the part of search committees, but a shortcut that relies on prestige will result in fewer women being hired for academic positions.

Search committees want to maximize the likelihood that their chosen candidate will be productive and will stay at their institution. It is time-consuming and expensive to hire people. In addition to the time and expense of the search, there is the expense of providing new faculty with start-up costs. If the new faculty member leaves, either because he or she is not productive or because he or she receives a better offer elsewhere, the process starts all over again and the department's investment has gone down the drain. Search committees reasonably want to maximize the likelihood of success of someone they are about to hire. Prestige of PhD-granting institution, prestige of the faculty member(s) the candidate worked with, prestige of the journals the candidate published in, how many publications the candidate has—all those variables are interrelated. But if the committee keeps in mind the values that diversity of all sorts brings to academia, as we discussed in chapter 2, they will need more strategies than an initial reliance on prestige in order to find people who can flourish at their institution, while contributing to its excellence.

Reducing Reliance on Prestige

Three specific practices would reduce the reliance on prestige. The first is to limit to three (or fewer) the number of publications that candidates can submit; that would require candidates to submit the work they think is their most important. If the most important paper is short, so much the better.

The second is to ensure that every candidate who fits the basics of the job description has *at least one* of their papers read by *at least one* search committee member. Readers would be on the lookout for work they found particularly creative or deep or substantial. To that end, the job ad might ask the candidate to specify which one of the three papers they submitted they would most like the committee to read. It may be helpful in mitigating bias to take steps to have that paper read without information about the authors' names or institutional setting. In most fields, it is possible to divide up the search committee's labor in that way. With a large number of applications—say 400—and a search committee of five people, that would mean that after an initial cut (always by at least two of the committee members) of those who do not fit the job description well, each person might read about 50 papers. A search committee member might well find that a daunting prospect! In such a case, the first step might be skimming the abstracts and only reading further if the paper seemed noteworthy in some way. Another solution for very large applicant pools is to divide the reading across the department as a whole.⁴ In any case, having a rubric against which the reader evaluates each paper would help standardize assessments both within particular readers and across readers.

The third practice is to look for evidence that candidates from lower prestige institutions could thrive at the searching institution by seeing if they have achieved more than would seem to be the norm for the institution where they received their training or started out. For example, have they applied for external funding? Have they presented at conferences?

For departments that want to increase the diversity of their faculty, the goal is to find women and people of color who could be very successful. Some schools' diversity officers review departments' short lists with an eye to seeing whether any women or people of color might have been overlooked and could be added to the short list. While this procedure is better than nothing, because it ensures some attention to diversity, it is not as good as having the department itself take the responsibility to assess talent in the absence of prestige markers.

Numbers of Publications

Quantity is easy to measure and thus a tempting shortcut for search committees. Journal articles can be evaluated in terms of quantitative measures—the number of publications, the impact factors of the journals that the candidate

publishes in, and the number of citations the person's publications have accrued. Books can also be quantitatively evaluated: their number, the prestige of the publisher (usually an academic press), and the number of citations. Each of these markers is itself subject to prestige effects where women and people of color are likely to fare slightly worse than White (and sometimes Asian) men.

Analysis of a nationally representative sample of almost 1,600 individuals in research-intensive institutions shows that researchers increasingly published papers while still in graduate school from 1970 to 2000 (Pineiro, Melkers, & Youtie, 2014). Before 1980, less than half of graduate students had published a paper pre-PhD. By 2000 and later, a majority of students had published a paper before getting their PhD.

In chemistry, biology, computer science, earth and atmospheric sciences, and electrical engineering, greater productivity as a graduate student is correlated with greater productivity as an assistant and associate professor (Pineiro et al., 2014). Thus, in those fields, counting a candidate's number of publications as a graduate student is a reasonable way of predicting how productive the individual will be as an assistant and associate professor. Here, too, however, there will be an effect of prestige of institution.

Individuals who publish with their advisor (in fields where that is a practice) publish more papers than do those who publish without their advisor (Pineiro et al., 2014). To some extent, greater productivity among those who publish with their advisors may be due to selection factors: advisors may choose particularly talented students to publish with. High-publishing advisors may also be more likely to include students as coauthors, especially if they are asked to contribute book chapters. Publishing with an advisor potentially benefits a student in multiple ways: the advisor has skills and knowledge the student can learn from and has a network that can help the student. Those who published at least one paper with their advisor as a graduate student had almost a third more publications per year postdegree than did those who did not publish with their advisor (Pineiro et al., 2014).

Since the 1970s, men have been more likely than women to publish with their advisors, though the gap is narrowing. Whites have been more likely than Asians and other underrepresented minorities to publish with their advisors, though that gap has narrowed (Pineiro et al., 2014). It is possible that men and Whites are more talented than women and people

of color and are chosen by faculty for that reason. A different possibility is that faculty are more likely to *think* that men and Whites are more talented.

Even when controlling for individual productivity, students who had collaborated with advisors had larger publication records postdegree than did those who had not (Pinheiro et al., 2014). That suggests that collaboration with an advisor has benefits above and beyond those that increase publication rate. Junior collaborators acquire academic capital that serves them well in their subsequent careers. In addition, participation in a research “team” that endures over several or many papers may help an individual look “productive” when it is really the collaboration that is productive. Finally, individuals whose parents were academics also have an advantage.

In sum, academic capital, which includes learned skills in the professional activities required in the academy, plays a role in productivity. Individuals with more experience of academia—whether through parents or through advisors—get a leg up that benefits them for the rest of their career. The small advantage that men and Whites accrue in graduate school ramifies, via the accumulation of advantage, in their productivity after graduate school. Search committees that put a premium on numbers of publications may select individuals who will be productive faculty, but they will miss other individuals who could also be productive. The individuals whom they miss are more likely to be women and people of color. Thus, if a search committee wants to increase the diversity of its interview pool, it will have to take into account the lower opportunity that members of some groups have had to accumulate publications. The committee can also remind itself that it cares more about quality than quantity.

Prestige of Publications

Most academics on search committees know the prestige level of the journals in their field (and some institutions require departments to provide a list of the journals in their field that occupy different tiers). One advantage for graduate students who publish with their advisors is that they can acquire knowledge of the prestige ranking of journals and learn how to determine a good match between their paper and a journal. Journal reputation can be assessed via its impact factor, or the average number of times any given paper in a journal is cited, either over the preceding two years or over the preceding five years. Reliance on impact factors is most common in natural and social science fields, despite serious criticism of the practice

(see, e.g., the San Francisco Declaration on Research Assessment: <http://www.ascb.org/dora/>).

Specialized journals in small fields may have impact factors between 1 and 2, meaning that the average paper in that journal over the preceding two or five years was cited 1–2 times. Impact factors differ from year to year but are fairly stable. A broader journal in the same field might have an impact factor ranging from 3–5, while science-wide journals like *Science* and *Nature* have impact factors of 35 and 38, respectively. For small fields with relatively few researchers, impact factors are low because the number of people who might possibly cite a paper is small. Although impact factor is a rough-and-ready guide, it obscures other relevant differences among journals, such as their theoretical orientation or the size of the field or sub-field they reflect.

Deciding where to submit a manuscript is a complicated process. Experienced researchers have a great deal of informally obtained knowledge about journals, from the types of articles a given journal publishes, to the difficulty or length of the review process, to whether they themselves read and cite articles in the journal. All of those factors go into informal judgments of the prestige of someone's publications. When search committee members are looking at CVs outside their immediate area of expertise, they may be unable to judge the importance of a journal by these possibly unshared metrics. They may downgrade or ignore the quality of the publications in specialty journals and resort simply to looking at the number of publications.

Since most researchers tend to have publications in a variety of journals, one strategy a search committee member may use is to scan the list to see whether there are at least a few of the widely recognized “top-tier” journals represented. If there are, the presence of papers in lower tier or unknown journals may be less important. One woman who had served on many search committees and promotion committees thought that men and women made different trade-offs. Men, she thought, published a lot, with a mix of papers in top-tier and lower tier journals. Women, she thought, published less and tended to emphasize publications in top-tier journals. In her view, those different trade-offs benefited men because evaluators look for both quality and quantity. As long as there are some papers to indicate quality, quantity helps. She went on to speculate that when men had a paper rejected from a top-tier journal, they immediately went to a lower tier

journal, making only minor changes in the manuscript. But when women had a paper rejected from a top-tier journal, they spent time doing extra experiments to buttress their point and then resubmitted to the same journal. Since that is a more time-consuming procedure, people who adopt it will publish less than people who switch quickly to another journal.

Whether those informal observations are correct is difficult to establish. There have been mixed reports in the literature about sex differences in quantity and quality of publications. Overall, women tend to publish less than men, though the gender disparity is smaller than it was in the past, as we outlined in chapter 4, and is smaller in some fields than others. Many studies reporting gender differences in favor of men do not control for prestige of institution (e.g., in academic medicine; Eloy, Svider, Cherla, Diaz, Kovalerchik, Mauro, et al., 2013; Holliday, Jagsi, Wilson, Choi, Thomas, & Fuller, 2014), even though high prestige is related to productivity and affects reviewer judgments positively when prestige is known. Controlling for prestige and other institutional factors has eliminated gender differences in productivity in some studies (Xie & Shauman, 1998). Other studies, such as a Dutch study that used a pool in the social sciences, have found that gender differences in productivity do not exist among active scientists within three years of their degree, but do exist for active scientists within 15 years of their degree (van Arensbergen, van der Weijden, & Van den Besselaar, 2012).

For graduate students, publication rates are linked to prestige of institution, so the fact that a smaller percentage of women and underrepresented minority men and women are at higher prestige schools, compared to White men and Asians, is likely to result in their having fewer publications.

Whether the quality of women's publications (again, as measured in citations) is greater than men's is much more difficult to determine, and reports vary tremendously. For example, one study examined the records of 85 social and personality psychologists whose degrees were completed in 1996 or 1997, who had full-time faculty positions, and who were members of the Social Psychology Network. They found no sex differences in quantity of publications or quality (as measured by the impact factor of the journals they published in), and no sex differences in the number of times their publications were cited. Although quantity and quality were moderately correlated, the correlation was not linear. Rather, at high levels of quantity

or quality, the two were negatively correlated, yielding the implication that publishing very frequently takes a toll on quality, and publishing very high quality work takes a toll on quantity.

A cross-national study of number of papers and number of citations in the sciences found that women were cited less than men, and, in papers published collaboratively, those with a female first or last author were cited less than those with a male first or last author (Sugimoto, Lariviere, Gingras, & Cronin, 2013), though many variables that might be correlated with gender, such as age, were not controlled for. A study of citations in international relations that did control for a large number of variables similarly found that men were cited more than women (Maliniak, Powers, & Walter, 2013), in part because women cite themselves less than men cite themselves and because men (who are still the majority of researchers in this field) cite men more than they cite women. By one estimate, men self-cite 56% more than women do (King, Berstrom, Correll, Jacquet, & West, 2016).

In sum, the CV measures of a candidate's quality will tend to give an edge to White men. The measures are objective, in the sense that they are quantifiable. However, they do not allow the evaluator to independently determine the two most important things—(1) the quality of the person's work and (2) the likelihood that the person can succeed and flourish at the hiring institution. Further, search committee members will find it difficult to mentally adjust for prestige. If two candidates are otherwise similar, the nod is likely to go to the candidate with the more prestigious background.

Letters of Recommendation

Search committee members tend to read letters of recommendation (if they have been submitted) after reading a candidate's CV. Women are at a disadvantage, even when their objective characteristics are the equal of men's (Madera, Hebl, & Martin, 2009; Schmader, Whitehead, & Wysocki, 2007; Trix & Psenka, 2003). Studies have varied in the disciplines they examine—academic medicine (Trix & Psenka, 2003), chemistry (Schmader et al., 2007), and psychology (Madera et al., 2009). They have also varied in whether they examine only individuals who have been hired (Trix & Psenka, 2003) or all candidates (Madera et al., 2009; Schmader et al., 2007). Finally, some studies also assessed candidates' objective attributes via CVs (Madera et al., 2009; Schmader et al., 2007).

In the landmark study that originated this line of research, Trix and Psenka (2003) found that letters for women were shorter than those for men, had fewer “standout” terms (expressions like *brilliant*, *superb*, *best*, *outstanding*, *unique*, *exceptional*), and more “grindstone” terms (expressions like *responsible*, *meticulous*).⁵ The combination of fewer standout terms and more grindstone terms can make someone seem not terribly talented but hardworking.

In chemistry, letters for male and female candidates were very similar in length, substance, and use of grindstone expressions. There were, however, significantly more standout terms for men than women, even when their objective characteristics (such as number of publications) were the same. “Star” quality is more likely to be conferred on men than women. In psychology, letters of recommendation for assistant professorship positions in psychology included more terms associated with women—those that emphasize women’s nurturance and communal characteristics—in letters for women than in letters for men. The letters were similar in length and general substance. Although communal characteristics might seem a plus for professors, they were rated by potential academic employers as actively detrimental (Madera et al., 2009). Thus, there are differences by field in how recommendations for men and women are written; overall the data suggest that women fare less well than men, though details vary from study to study.

Readers of letters are unlikely to be able to perform the kinds of analyses that experimenters do, looking at hundreds of letters from multiple searches. They are also unlikely to know a priori that they might be influenced by the use of standout terms. Standout terms, in our view, even in an era of inflated, say-nothing-bad styles of letter writing, have not yet lost their currency. To the extent that those terms are ones that readers nonconsciously search for and are affected by, women and (by inference) underrepresented minorities will be at a disadvantage.

We have discussed three ways that women and people of color may be at a disadvantage compared to White men during the initial review process: their PhDs are more likely to come from lower prestige institutions, they are less likely to be well-published and well-cited, and their letters of recommendation are likely to be weaker. Each of the effects is relatively small and by no means uniform: some White women and people of color, for example, have degrees from high-prestige institutions, have a strong publication history (which is correlated with being at a high-prestige institution), and

have strong letters of recommendation. Equally, some White men do not have those achievements. On average, however, the proxies that are used to estimate quality favor White men.

Buffering against Bias in Creating a Medium-Short List or Short List

The creation of a medium-short list or final short list of candidates can be affected by all the issues we have raised. What can be done to reduce the effects of less diagnostic information when constructing a medium-short and short list? Deciding on who will screen the applications and how they will do it is crucial (see, e.g., Bauer & Baltes, 2002). In some institutions it appears to make sense to have one individual cut the large applicant pool in half by making a fast pass through the files to exclude those who do not meet eligibility requirements. Without a plan, however, that procedure invites, at a minimum, prestige bias: one individual or a small group may screen out individuals from certain kinds of institutions, or people with advisors they have never heard of, and so on. Since, as we have seen, women and people of color have fewer prestige markers, they have a higher chance of being eliminated at the beginning, before their merits are thoroughly reviewed.

Establish Valid Criteria To buffer against the biases that informal and implicit criteria might introduce, we recommend establishing and using explicit criteria before review of candidates' materials. The process of adopting a consensual set of criteria to rate candidates provides an opportunity for the search committee or the department to be explicit about the relevant bases for judging applicants and potential applicants. The criteria should be valid and thus should relate to what the department wants in a faculty member. Perhaps published work will be considered more important than grant support; perhaps the reverse will be the case; perhaps both will be weighted equally. Perhaps ability to direct an MA or PhD program will be important. It is, of course, not enough to develop criteria. It is necessary to actually use the criteria and to assess applicants in terms of those agreed-upon criteria rather than in terms of some overall impression. All candidates must be evaluated by the same criteria to the same extent. What applies to recruitment (see chapter 5) also applies to the interview: for the formal part of the interview, all candidates must be asked the same questions, even if that seems slightly awkward.

Ensure Accountability In general, accountability (e.g., to colleagues) and adequate time for making informed judgments are hallmarks of good decision-making processes (Arthur & Doverspike, 2005; Koch, D’Mello, & Sackett, 2015; Roberson, Galvin, & Charles, 2007). To facilitate evaluators’ accountability, we recommend a formal list of the criteria on an “applicant evaluation sheet” (see box 6.1) that every reviewer completes for each applicant. The committee can then consider applicants with high average ratings by reviewers on any, some, or all criteria, as well as those rated high by *any* reviewer. That allows applicants whose strengths may be overlooked or underestimated by some reviewers to nevertheless surface in discussion. The tool can be adapted so that it will be appropriate for different searches and for different search stages. One might have a slightly different set of criteria for the first stage than for later stages.

Many institutions have adopted the practice of “short-list review” (see Bilimoria & Buch, 2010), in which search committees share their intended short list with the dean’s office or other administrative office, along with a description of their efforts to generate a pool of diverse qualified applicants and information about the demographic makeup of the applicant pool. If the short list is less diverse than the applicant pool, the dean or the dean’s delegate can then discuss the reasons for this departure from expectations, and, if it seems appropriate, stop what would otherwise be the next stage of inviting candidates to visit the campus. This form of accountability motivates search committees to ensure that their process is a good one. Further, one or two such actions on the part of the dean communicates the administration’s seriousness of purpose and moves less motivated departments to redouble their efforts to attract diverse applicants.

In one case we know of, a department that had a short list with ten male candidates and no females was stopped from inviting any of them and was encouraged to review their pool and short list again. They came back to the dean with a new list that included several women. In the end, their top candidate—and the person they hired—was one of those women added only after the dean’s intervention. Such experiences help persuade departments that they have been overlooking talent that was actually there. Deans and provosts can encourage diversity among interviewees by providing funds to invite more candidates to campus if the pool is a diverse pool.

Box 6.1

Applicant Evaluation Tool

The following offers a method for search committee members to provide evaluations of job applicants. It is meant to be a template that they can modify as necessary for their own uses. The proposed questions are designed for junior faculty candidates; however, alternate language is suggested in parentheses for senior faculty candidates. Committees often need two versions of the form, one for screening (based on less material) and one for reviewing full files for the long short list.

Applicant's name:

Please indicate which of the following are true for you (check all that apply):

- Read applicant's CV
- Read applicant's statements (re research, teaching, etc.)
- Read applicant's letters of recommendation
- Read applicant's scholarship (indicate what): _____

Please rate the applicant on each of the following:	excellent	good	neutral	fair	poor	unable to judge
Evidence of scholarly productivity						
Evidence of creativity and innovation in research/scholarship						
Potential for [Record of] scholarly impact/tenurability						
Potential for [Record of] funding (if appropriate to field)						
Evidence of strong background in [relevant fields]						
Contribution to breadth of areas addressed in department research						
Evidence of teaching experience and interest (including grad mentorship)						
Potential [Ability] to teach courses in core curriculum						
Potential [Ability] to teach and advise diverse students						
Potential [Ability] to contribute new course coverage to the curriculum						
Potential [Ability] to contribute positively to the collegial work environment of the department						
Other comments? _____						

For more information or additional copies of this resource, please contact the ADVANCE Program at (734) 647-9359 or advanceprogram@umich.edu, or visit the ADVANCE Program's website at <http://advance.umich.edu/strideResources.php>

Attend to Pool Composition Given the risk of underrating women and people of color, we recommend that search committees attend at the outset to the ratio of male and female, and White and non-White, applicants in the applicant pool. If the ratios do not match the availability pool, the committee can investigate why that is the case and try to bring the applicant pool numbers up at least to the levels of the availability pool. Depending on the field, different ways of estimating the pool will be appropriate, as discussed in chapter 5. In some fields, the PhD rate will be appropriate. In others, the postdoctoral fellowship rate will be more appropriate.

Only by continuing to monitor the composition of successive pools can the committee determine whether it might be using irrelevant criteria to filter candidates (Sagaria, 2002). Useful equity benchmarks include the number and percentage of women and underrepresented minorities in each successive pool after the applicant pool: the preinterview pool, the medium-short list pool, the interview pool, the offer pool, and the candidate acceptance pool. Those benchmarks help the committee see what criteria are being used—or misused. If the ratios of male to female, and White to non-White, increase as the pools narrow, that is a reason to assess the criteria and their application. That will take time, and committees often feel pressed for time. It may nevertheless be necessary to alter the timeline and create procedures that will broaden the pool, even at a late date.

The Interview Stage

As electronic means of interviewing people become more feasible, more search committees interview as many as 10–12 candidates using videoconferencing or teleconferences of approximately one-half hour each, with all committee members present for each interview, and with all questions determined ahead of time by the committee. Some schools stack a large number of 30-minute interviews at annual meetings. An advantage of a preinterview is that search committee members may be more willing to include nontraditional, “riskier” candidates in a large set of interviews than in a small set. A preinterview also provides an opportunity for search committees to see whether their medium-short list adequately represents the pool. Finally, preinterviews, because of their brevity and the need to compare many candidates, lend themselves to a structured interview approach that is much less likely to be adopted in the later interview stage. Structured

interviews help eliminate demographic effects (Levashina, Hartwell, Morgeson, & Campion, 2014; McCarthy, Van Iddekinge, & Campion, 2010).

If the interview is structured in order to obtain the most job-relevant information from candidates, that information will serve to individuate one candidate from another. The more individuating information interviewers have about applicants, the less likely they are to rely on demographic information (Kunda & Thagard, 1996), although we have reason to believe that gender can survive individuating information (Chan & Mendelsohn, 2010; Mondschein, Adolph, & Tamis-LeMonda, 2000). As a meta-analysis demonstrated, individuating information did not lessen the extent to which decision makers tended to prefer males for male-dominated positions (Koch et al., 2015).

People are not generally motivated to search for individuating information. Everyone is a cognitive miser, hoping to do the most with the least, to make a judgment quickly and with as little effort as possible. Thus, people need external motivation to overcome their internal cognitive miserliness. The presence of panels that perform the structured interviews helps provide that external motivation because panelists are accountable to each other. The benefit of structured interviews can still be limited, however, if a powerful individual on the panel sets a tone that the less powerful individuals on the panel go along with. The benefit can also be limited if the panels themselves are limited. Two White men might find it difficult to treat each candidate similarly in the less quantifiable aspects of their behavior that can affect candidates' performance. Such aspects as speech errors, interruptions, displays of nonverbal dominance, and failure to look at the candidate while he or she is speaking can impair candidates' performance (e.g., Latu & Mast, 2016).

Structured interviews help ensure that every element of the interview is completed for every applicant, which prevents interviewers from stopping the interview too early. In a less structured context, they may be tempted not to complete the interview, or stop the interview early because they already feel confident that the candidate is a good choice or that the candidate is a bad choice. Structured interviews reduce group differences in ratings (Levashina et al., 2014).

Structured Interviews

Both the preinterview and the interview with the search committee during an on-campus visit can be structured. What counts as a good structured interview? One list contains 15 characteristics (McCarthy et al., 2010). We

summarize them in order to demonstrate how few desiderata most interviews of candidates adhere to. Questions in the interview should be targeted to the characteristics that are optimal for success in the job, so that interviewers' attention will be directed to those characteristics rather than demographic characteristics. The intent can be subverted by developing questions that are not genuinely job-success-related ones, so it's best if the questions are developed by a working group that explicitly pays attention to job-relevant characteristics.

Seven of McCarthy et al.'s (2010) characteristics are as follows:

1. the interviews are based on a comprehensive job analysis
2. within the interviews, the same questions are asked of each candidate, and within the experience-based interview, similar questions are asked of each candidate
3. the use of prompts and follow-up questions is limited
4. different questioning techniques are employed (e.g., ones focused on past experience vs. hypothetical situations)
5. each interview allows sufficient time for interviewers to ask several questions
6. ancillary information is controlled
7. candidates are encouraged to ask questions after the structured phase of the interview process is complete

Human resources departments may conduct such job analyses and develop clear and specific questions, but academics seldom go beyond stating that they want someone who will be a productive, grant-funded researcher with good teaching skills and collegial departmental behavior. Once the candidate is on campus, he or she may have several interviews with individual faculty and lunch or coffee with small groups of faculty, in addition to having a meeting with the entire search committee. How the unofficial interviewers acquire information is typically up to each individual to decide. The assembly of information is fairly flexible.

In terms of requirements 1–7, academic interviews generally do well on having enough time (number 5), but perhaps less so on inviting the candidate to ask questions (7) or on limiting prompts and follow-up questions (3), using multiple questioning techniques (4), or even knowing what such techniques are! Results are usually mixed at best on conducting a

comprehensive job analysis, asking the same or similar questions across interviews, and controlling ancillary information (1, 2, and 6). This spotty success suggests that interviewing deserves much more care and attention than it currently receives.

Requirements 8–15 from McCarthy et al. (2010) are as follows:

8. interviewers evaluate each dimension using behaviorally anchored rating scales
9. descriptive scale anchors are derived from...definitions [from] previously developed interviews and responses from previous candidates
10. interviewers are trained on the importance of note taking during the interview process
11. a panel of two interviewers evaluates each candidate
12. the same set of interviewers conducts the interviews for each applicant
13. the interviewers do not discuss candidates between interviews
14. all interviewers are extensively trained to ensure proficiency in conducting and scoring the interview
15. statistical procedures (unit weighting) are used to combine ratings within each interview

Who does all of that? We think it is safe to say: no one. Who would be willing to do it? We think it is also safe to say: in academia, pretty much no one. Although it is encouraging to know that one can eliminate demographic effects with a lot of effort, it is sobering to recognize how likely one is to be affected by gender and race without that effort.

On-Campus Interviews

On-campus visits by the candidate typically last at least one day and often last two days. They include a variety of activities, including, at a minimum, a “job talk” in which the candidate covers his or her work. Often three or four candidates are invited to campus over a period of a few weeks. Many faculty (and students) beyond the initial search committee are likely to participate. At the interview visit, candidates in some science fields may be also asked to give a “chalk talk,” in which they present more technical material than would be appropriate for a talk that the entire department will attend; at least one member of the search committee typically attends. At many schools, candidates also give a sample lecture to a class in their topic area, a

lecture which is attended by at least one member of the search committee. Candidates generally have a session with the entire search committee and are asked an established set of questions. Candidates meet with other faculty, administrators, and undergraduate and graduate students, and they see the facilities. This broader contact, and weaker structure, allows many things to happen that may be irrelevant to assessing the candidate's "merit."

Some people interview beautifully and some interview badly. It is difficult for search committees to properly weigh interview performance against the years of performance testified to by the person's work. The very thin slice of behavior one gets at an interview can dominate the search committee's impression. One can be much more or much less impressed by a candidate in person than one was by their record. Although we know of no way to compensate for our tendency to pay too much attention to an in-person impression, we caution search committees about this tendency. People are very confident about their ability to detect someone's merit. The unstructured interview lends itself to false confirmation because people primarily seek information that will confirm their views. People tend to do worse at predicting behavior when they rely on unstructured information than when they rely on solid information (Dana, Dawes, & Peterson, 2013; Kausel, Culbertson, & Madrid, 2016).

Candidates who come from less prestigious institutions may have been less well prepped for their visit than those who come from more prestigious institutions. Although most candidates know that it's good to practice their talks multiple times ahead of time and develop answers to the questions they might be asked, many fewer candidates, in our experience, undergo mock interviews at their own campus. Some candidates do not know until a few days before the visit what the interview will be like.

Search committees can help level the playing field by posting on their website what a visit will involve and by giving candidates an idea of what they are looking for at each event. Search committees can also suggest how candidates can prepare for their on-campus visit, and they can bring up topics (like disability or lactation needs) that might be difficult for the candidate to raise. At the same time, we acknowledge the value in seeing how a candidate handles an unexpected question. A good search committee will try to bring out the best in each candidate without producing cookie-cutter interviews. It is a delicate balance. Just as there is an art to interviewing well as an interviewee, there is an art to interviewing well as an interviewer.

Especially in the transition from the preinterview (if there is one) to the interview, and in the transition from the interview to the offer, search committee members who differ demographically from the job candidates may be hard-pressed to maintain subtle behaviors that will bring out the best in a candidate. For example, White male interviewers may be less comfortable in interviewing women and people of color compared to other White men. If so, they will be less likely to bring out the qualities that will show the candidate at his or her best and less successful in interpreting a candidate's interest. The very people who come to the interview at a disadvantage, for the reasons we outlined earlier, can also be at a disadvantage because of deficits in interviewers' abilities.

One early study examined White male students' behavior with other students who were, unknown to the participants, confederates of the experimenter (Word, Zanna, & Cooper, 1974). The confederates were White or African American. The students were White. The students' task was to interview the student they were paired with for a job as a peer counselor. Students who were paired with an African American confederate produced more filled pauses (e.g., *uh*, *er*) and other speech errors, sat farther away, and completed the interview faster than did the students who were paired with a White confederate. Although that study was conducted in 1974, recent research suggests that the same processes were at work in 2016, this time in a study that had one student teach another student (Jacoby-Senghor, Sinclair, & Shelton, 2016).

Research in Switzerland suggests that gender schemas play a role when men interview women in a laboratory setting for a fictitious position as a marketing manager (Latu & Mast, 2016; Latu, Mast, & Stewart, 2015). In one suggestive study, men who adopted a dominant style in interviewing, by looking at the female candidate more while he was talking than when he was listening, and by interrupting her more, had an effect on the women who were being interviewed. (Only women were selected to be interviewed.) Both the women's self-evaluation and the men's evaluation of the women were lower when the men were dominant. When women were the interviewers, their dominance level did not have an effect on candidates, but the number of women interviewers was too small to be definitive.

We conclude that White male interviewers are likely to do a worse job of interviewing women than men, and of interviewing African Americans compared to Whites—at least if people use unstructured interviews. Academic

interviews are probably less structured than employment interviews in the business sector, in part because so many different individuals participate in job interviews in academia, most of whom have little or no training in interviewing. In highly structured interviews, as we have mentioned, there are no demographic effects (McCarthy et al., 2010). That is, it does not matter what the sex, race, or ethnicity of the interviewers or interviewees is if the interviewers are adhering to highly structured guidelines. It may also be helpful to ensure that all candidate interviews are conducted with at least two people present, so that problematic questions or assertions from an interviewer can be addressed by the other person present.

How the interviewers represent the school will affect that candidate's interest in coming to it. Interviewers must not ask any illegal questions! Candidates know that questions about age, partners, marriage, and plans for children are illegal. Some interviewers, with the best will in the world, think that casual queries over dinner at the end of an interview day are outside the interview proper and do not count. They do count. The interview starts the moment the candidate is contacted and ends only when someone has been hired. Throughout that period, the candidate is in interview mode. At no point is it permissible to ask about someone's personal life. The interviewer may think that knowledge of the candidate's personal needs will make it easier to let the candidate know what accommodations can be made. Nevertheless, it is illegal.

What interviewers do not anticipate is that some personal questions may be very painful for a candidate. The candidate may have recently ended a relationship, or may have a partner, relative, or friend who is ill or recently died, or may have just learned that they are infertile, or may have experienced any one of a number of other difficult life events. Interviewers are not generally thinking about how a question might be upsetting for a candidate, especially since they see themselves as being friendly and sympathetic. But any personal information that a job candidate does not volunteer should be personal information that the interviewer does not ask the candidate to divulge.

The candidate faced with an illegal question is in a quandary. African Americans, for example, may be asked how they would feel about being in a community that is not racially diverse. Women may be asked whether they are married or are planning on having a child. Candidates do not want to appear churlish, but they also do not want their personal life to

be part of the interview. The late Denise Denton said that when she was asked such questions, she answered, "Does that matter?" Not everyone has the confidence, or perhaps seniority and standing, to pull that off. Another possible answer that candidates who are just finishing their degrees or post-docs could provide, is, "Oh, I'm under strict orders from my supervisor (or mentor) not to answer any personal questions," said with a smile. We suggest that candidates practice an answer ahead of time, try it out on various colleagues, tweak it if necessary, and be prepared to use it to answer any question that seems personal. But no matter how tactful or friendly such a response is, it runs the risk of alienating interviewers.

What *can* interviewers do if they are interested in attracting women and people of color as faculty? We recommend that interviewers treat all candidates as similarly as possible and indicate to *every* candidate—no matter their gender, or race, or ethnicity, or sexual orientation, or religion—that the department and institution value diversity (if that is true!). Departments benefit from providing that information to everyone, not only White women and minority women and men; everyone needs to know that they are going to be in a community that values diversity (if, of course, it does). Departments and other offices at the institution can indicate the kinds of supports that it provides. Examples would be support to faculty partners in the hiring process, support in the form of childcare resources and family-friendly policies, and techniques for addressing work-family conflicts, such as stop-the-clock policies. In the ideal case, the department will provide every candidate with full information about all types of faculty support policies and resources. It will also schedule a meeting between the candidate and someone who is not in the department, of whom the candidate can ask questions about what it is like to live there, without concerns that his or her questions will be relayed back to the department. Such community ambassadors should be knowledgeable about matters of concern to diverse groups. Those efforts will help convince otherwise skeptical candidates that the department and school is committed to inclusion and diversity.

One issue an African American may have about small towns, for example, is the availability of services. Single women and men may wonder whether the local social life is couple centered or more diverse. Gay men and women can wonder whether the local community accepts individuals who are out and whether there are centers of community for gay men and lesbians. Muslims may be concerned about whether there is a mosque or a halal

butcher in the community, as well as whether there are intense pressures to drink alcohol. Jews may wonder about the presence of stores where they can buy kosher food and how far away a synagogue is. Everyone can wonder about something. A school benefits by knowing what it is like for members of different groups to live in their community. One of us was on a site visit to a research-intensive school in a somewhat rural area and heard women faculty talking about what a terrible place it was to be a straight single woman. The culture was couple centered, and there were virtually no straight single men. Straight single women spent a lot of time trying to figure out how to leave—and did leave.

If schools want to increase the diversity of their faculty, they have to spend some time helping to make the community attractive to their faculty. Every location has advantages and disadvantages. Some disadvantages can be ameliorated. For example, a school can advocate with local retail stores for inclusion of a wider range of products for diverse customers, and it can document for faculty and faculty applicants the nearest synagogues, mosques, and community groups that offer cultural support to particular immigrant groups.

Negotiation of the Offer

The final stage of recruitment is critical. It is often the case that issues that arise in the course of negotiations, if not satisfactorily addressed, affect the transition onto campus and future retention success. A good offer takes into account a candidate's needs and desires, is communicated with enthusiasm, and presents terms that can be negotiated.

Person-Specific Hiring

As noted at the end of chapter 5, some faculty recruitment involves person-specific hiring. In such cases—recruiting partners of candidates identified in regular searches or faculty who would bring special diversity and excellence to the department—the goal is to hire someone who will be valued and successful at the institution. For that reason, it is ideal to use procedures that mirror those involved in regular searches as closely as possible. Thus, it is best to engage a group within the department in the process of assessing the candidate's qualifications, using the same kinds of evidence and the same

kinds of evaluation tools that are used in regular searches. Similarly, the visit of the candidate to campus should be conducted in exactly the same way that all candidates' visits are handled, and the postvisit decision should be managed in the same way as well. A detailed account of one institution's procedures aimed for this situation is contained in the University of Michigan's Handbook for Faculty Recruitment and Hiring (advance.umich.edu/resources/handbook.pdf). There are other accounts on other institutions' websites, particularly those that have participated in the NSF ADVANCE program.

Summary

Evaluating job candidates requires skill, and whenever there is a demographic imbalance between the search committee members and the candidates, more skill is required. Everyone is likely to be influenced by prestige, gender, ethnicity, and other schemas when making evaluations. The challenge can be met in part by developing procedures that are designed to protect against unintentional error. No set of procedures will accomplish that perfectly, but the better the procedures, the better the likely outcome.

Recommendations for Practices that Increase the Fairness of the Evaluation Process

Provosts, Deans, and Department Chairs

1. Always express the twin goals of (1) excellence in scholarship, teaching, and service and (2) diversity in faculty hiring. Encourage confidence that the two goals are mutually compatible and indeed mutually reinforcing.
2. Consider the adoption of "short-list review"; if adopted, enforce use of diverse pools.
3. Publicly praise and consider rewarding departments or other units that succeed in increasing the diversity and excellence in their faculty. Describe the ways they accomplished this achievement to encourage others.
4. Provide resources for stages of the process if needed to increase diversity and fairness (e.g., more funding for travel for more candidates, development of materials to reflect the institution's values and commitments in the best light, data about past searches, demographic characteristics of units, etc.).

Search Committee Members

1. Be sure your application packet includes what you want and does not include material you do not want, at each stage. Consider limiting the number of papers or writing samples submitted in order to focus on quality.
2. Consider postponing review of letters of recommendation until later in the process (after initial decisions based on review of the actual work).
3. Set out explicit criteria for evaluating candidates early in the process and continually monitor your own and other committee members' comments for evidence of drift from the criteria or differential application of them to different candidates. Be particularly alert to comments that reflect known gender, race, and family status biases.
4. Avoid relying on proxies for assessing the quality of candidates' work. Read the work, and discuss it with your colleagues.
5. When sharing applicants' work, minimize identifying cues that might lead to reliance on prestige or quantity indicators that are only proxies for quality.
6. Be sure decisions about who will be retained in the pool are made by multiple judges at every stage.
7. During preinterviews (if they occur), use structured interview procedures thoughtfully.
8. During interviews, schedule a structured interview between the candidate and the search committee.
9. Design the candidates' on-campus visits carefully, with attention to issues of differential comfort with different situations. Consider ensuring that interviews always involve more than one person. (People monitor themselves differently if there is an observer.)
10. Monitor the pool at all stages of the process; always be open to extending a stage of the process in the interest of ensuring a broad enough pool of candidates to review.

Department Faculty Members

1. It is optimal if all members of a department understand issues of bias and evaluation. Take steps to educate yourself.
2. Participate in a serious way in the visit process. Review the candidates' dossiers in advance, taking special care to read their work. Attend the talk(s) that you can, and participate in interviewing candidates if you can.

Consider interviewing with another colleague or two who have a different background and approach than yours.

3. Be careful to avoid unintended negative effects of asking personal questions.

Applicants for Faculty Positions

1. White women and underrepresented minority candidates can learn how to address the possible deficiencies in their interviewers' styles.

2. Practice expressing your suitability, and show enthusiasm for the job. It is useful to practice with people who deliberately take a laconic approach, and who deliberately take an aggressive approach, including interrupting.

Notes

1. Ratings of job performance depend on the extent of fit between someone's vocational interest profile and the occupation's profile (Nye et al., 2012). Such congruence indices moderately predict ratings of job performance and employee persistence. To create a congruence index, an individual's three major interest categories and an occupation's three major interest categories are compared. A congruence index measures the extent to which the resulting profiles match. When there is high overlap, ratings of job performance are higher. Thus, interest ratings can be helpful for counseling people about what jobs to pursue. Note, however, that the interest ratings for an occupation are based on attributes of the current practitioners of an occupation.

2. This is on the assumption that *U.S. News and World Report* (USNWR) rankings accurately represent prestige. The size and wealth of a university or college, the salary of the faculty, the student-faculty ratio, and the students' SAT scores together account for almost 90% of the variance in USNWR rankings (Volkwein & Sweitzer, 2006). If one assumes that such indicators appropriately assess prestige, then, indeed, the rankings are valid. Other rankings, such as the Shanghai Jai Tong University *Academic Ranking of World Universities*, are responsive to two main, uncorrelated, factors: the number of world-famous researchers and the overall productivity of researchers (Dehon, McCathie, & Verardi, 2010). Those are factors that search committee members are likely to be responsive to, since fame, by definition, means that someone is well-known, and since the more publications an institution produces, the more likely it is that a random search committee member will have come across one. University ranking is complex. The distribution of universities seems to follow a Zipfian curve, with the top 50 or so universities being discriminable, but the mass of universities varies little one from the other in both number of world-famous figures and productivity (Dehon et al., 2010).

3. We computed these percentages from 2015 NSF data (N=55,006) at <https://nces.norc.org/NSFTabEngine/#TABULATION>. Seventy-seven percent of Asians, 72% of Whites, 73% of Hispanics, and 50% of African Americans received their degrees from institutions classified by the Carnegie Commission as having very high research activity. African Americans comprised only 5% of this in 2015. Fifty percent of the available pool of African Americans would be lost at the outset if search committees used prestige of institution to create their first list of possible candidates. Tests of independent proportions showed that Hispanics and Whites did not differ, but Asians were better represented among schools with very high research activity than were Whites, $z = 11.958$, $p < .0002$. African Americans were significantly less well represented than any other group, z (comparison with Hispanics) = -18.475 , $p < .0002$. A higher proportion of men than women received their degrees from very high research activity universities, $z = 14.789$, $p < 0.0002$.

4. In mathematics it may not be possible to adopt this strategy. One mathematician told us that he might have to spend hours on a single paragraph in a paper to understand it and determine its contribution. For most fields that we are familiar with, that issue doesn't arise.

5. There were too few letters from women to do an analysis of letter writer by gender.

References

Apfelbaum, E. P., Norton, M. I., & Sommers, S. R. (2012). Racial color blindness: Emergence, practice, and implications. *Current Directions in Psychological Science*, 21(3), 205–209.

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

Arthur, W., Jr., & Doverspike, D. (2005). Achieving diversity and reducing discrimination in the workplace through human resource management practices: Implications of research and theory for staffing, training, and rewarding performance. In R. L. Dipboye & A. Colella (Eds.), *Discrimination at work: The psychological and organizational bases* (pp. 325–327). San Francisco, CA: Jossey-Bass.

Bauer, C. C., & Baltes, B. B. (2002). Reducing the effects of gender stereotypes on performance evaluations. *Sex Roles*, 47(9–10), 465–476. doi:10.1023/A:1021652527696

Bertrand, M., & Mullainathan, S. (2003). *Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination* (Working Paper 9873). Cambridge, MA: National Bureau of Economic Research.

Bilimoria, D., & Buch, K. K. (2010). The search is on: Engendering faculty diversity through more effective search and recruitment. *Change: The Magazine of Higher Learning*, 42(4), 27–32.

Carnegie Foundation for the Advancement of Teaching. (2006). *The Carnegie Classification of Institutions of Higher Education* (2005 edition). Menlo Park, CA: Author.

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

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.

Clauset, A., Arbesman, S., & Larremore, D. B. (2015). Systematic inequality and hierarchy in faculty hiring networks. *Science Advances, 1*(1), e1400005.

Correll, S. J., Benard, S., & Paik, I. (2007). Getting a job: Is there a motherhood penalty? *American Journal of Sociology, 112*, 1297–1338.

Correll, S. J., Kelly, E. L., O'Connor, L. T., & Williams, J. C. (2014). Redesigning, redefining work. *Work and Occupations, 41*(1), 3–17.

Dana, J., Dawes, R., & Peterson, N. (2013). Belief in the unstructured interview: The persistence of an illusion. *Judgment and Decision Making, 8*(5), 512–520.

Dehon, C., McCathie, A., & Verardi, V. (2010). Uncovering excellence in academic rankings: A closer look at the Shanghai ranking. *Scientometrics, 83*(2), 515–524.

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.

Eloy, J. A., Svider, P. F., Cherla, D. V., Diaz, L., Kovalerchik, O., Mauro, K. M., et al. (2013). Gender disparities in research productivity among 9952 academic physicians. *Laryngoscope, 123*(8), 1865–1875.

Forscher, P. S., Mitamura, C., Dix, E. L., Cox, W. T., & Devine, P. G. (2017). Breaking the prejudice habit: Mechanisms, timecourse, and longevity. *Journal of Experimental Social Psychology, 72*, 133–146.

Gompers, P. A., & Wang, S. Q. (2017). *Diversity in innovation* (NBER Working Paper No. w23082). Cambridge, MA: National Bureau of Economic Research.

Headworth, S., & Freese, J. (2015). Credential privilege or cumulative advantage? Prestige, productivity, and placement in the academic sociology job market. *Social Forces, 94*(3), 1257–1282.

Heilman, M., & Okimoto, T. G. (2008). Why are women penalized for success at male tasks? The implied communality deficit. *Journal of Applied Psychology, 92*(1), 81–92.

Holliday, E. B., Jagsi, R., Wilson, L. D., Choi, M., Thomas, C. R., Jr., & Fuller, C. D. (2014). Gender differences in publication productivity, academic position, career duration and funding among US academic radiation oncology faculty. *Academic Medicine: Journal of the Association of American Medical Colleges*, *89*(5), 767–773.

Jacoby-Senghor, D. S., Sinclair, S., & Shelton, J. N. (2016). A lesson in bias: The relationship between implicit racial bias and performance in pedagogical contexts. *Journal of Experimental Social Psychology*, *63*, 50–55.

Kausel, E. E., Culbertson, S. S., & Madrid, H. P. (2016). Overconfidence in personnel selection: When and why unstructured interview information can hurt hiring decisions. *Organizational Behavior and Human Decision Processes*, *137*, 27–44.

King, M. M., Bergstrom, C. T., Correll, S. J., Jacquet, J., & West, J. D. (2016). Men set their own cites high: Gender and self-citation across fields and over time. *arXiv preprint arXiv:1607.00376*.

Koch, A. J., D'Mello, S. D., & Sackett, P. R. (2015). A meta-analysis of gender stereotypes and bias in experimental simulations of employment decision making. *Journal of Applied Psychology*, *100*(1), 128–161.

Kouchaki, M. (2011). Vicarious moral licensing: The influence of others' past moral actions on moral behavior. *Journal of Personality and Social Psychology*, *101*(4), 702–715.

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

Latu, I. M., & Mast, M. S. (2016). Male interviewers' nonverbal dominance predicts lower evaluations of female applicants in simulated job interviews. *Journal of Personnel Psychology*, *15*, 116–124.

Latu, I. M., Mast, M. S., & Stewart, T. L. (2015). Gender biases in (inter)action: The role of interviewers' and applicants' implicit and explicit stereotypes in predicting women's job interview outcomes. *Psychology of Women Quarterly*, *39*, 539–552.

Latu, I. M., & Mast, M. S. (2016). Male interviewers' nonverbal dominance predicts lower evaluations of female applicants in simulated job interviews. *Journal of Personnel Psychology*, *15*, 116–124.

Levashina, J., Hartwell, C. J., Morgeson, F. P., & Campion, M. A. (2014). The structured employment interview: Narrative and quantitative review of the research literature. *Personnel Psychology*, *67*(1), 241–293.

Madera, J., Hebl, M., & Martin, R. (2009). Gender and letters of recommendation for academia: Agentic and communal differences. *Journal of Applied Psychology*, *94*(6), 1591–1599.

- Maliniak, D., Powers, R., & Walter, B. F. (2013). The gender citation gap in international relations. *International Organization*, *67*(4), 889–922.
- McCarthy, J. M., Van Iddekinge, C. H., & Campion, M. A. (2010). Are highly structured job interviews resistant to demographic similarity effects? *Personnel Psychology*, *63*(2), 325–359.
- 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.
- Monteith, M. J., Ashburn-Nardo, L., Voils, C. I., & Czopp, A. M. (2002). Putting the brakes on prejudice: On the development and operation of cues for control. *Journal of Personality and Social Psychology*, *83*(5), 1029–1050.
- Moskowitz, G. B., & Li, P. (2011). Egalitarian goals trigger stereotype inhibition: A proactive form of stereotype control. *Journal of Experimental Social Psychology*, *47*(1), 103–116.
- Nye, C., Su, R., Rounds, J., & Drasgow, F. (2012). Vocational interests and performance: A quantitative summary of over 60 years of research. *Perspectives on Psychological Science*, *7*(4), 384–403.
- Okimoto, T. G., & Heilman, M. E. (2012). The “bad parent” assumption: How gender stereotypes affect reactions to working mothers. *Journal of Social Issues*, *68*(4), 704–724.
- Oprisko, R. L., Dobbs, K. L., & DiGrazia, J. (2013). Pushing up ivies: Institutional prestige and the academic caste system. *Georgetown Public Policy Review*. Retrieved from <http://gppreview.com/2013/08/21/pushing-up-ivies-institutional-prestige-and-the-academic-caste-system/>
- Pinheiro, D., Melkers, J., & Youtie, J. (2014). Learning to play the game: Student publishing as an indicator of future scholarly success. *Technological Forecasting and Social Change*, *81*, 56–66.
- Roberson, L., Galvin, B. M., & Charles, A. C. (2007). When group identities matter: Bias in performance appraisal. *Academy of Management Annals*, *1*(1), 617–650.
- Sachdeva, S., Iliev, R., & Medin, D. L. (2009). Sinning saints and saintly sinners: The paradox of moral self-regulation. *Psychological Science*, *20*(4), 523–528.
- Sagarra, M. A. D. (2002). An exploratory model of filtering in administrative searches: Toward counter-hegemonic discourses. *Journal of Higher Education*, *73*(6), 677–710.
- Schmader, T., Whitehead, J., & Wysocki, V. H. (2007). A linguistic comparison of letters of recommendation for male and female chemistry and biochemistry job applicants. *Sex Roles*, *57*, 509–514.

- Sheltzer, J. M., & Smith, J. C. (2014). Elite male faculty in the life sciences employ fewer women. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(28), 10107–10112.
- Stewart, A. J., Malley, J. E., & Herzog, K. A. (2016). Increasing the representation of women in STEM departments: What makes a difference? *Journal of Women and Minorities in Science and Engineering*, *22*(1), 23–47.
- Sugimoto, C. R., Lariviere, V., Ni, C., Gingras, Y., & Cronin, B. (2013). Global gender disparities in science. *Nature*, *504*(7479), 211–213.
- Trix, F., & Psenka, C. (2003). Exploring the color of glass: Letters of recommendation for female and male medical school faculty. *Discourse & Society*, *14*(2), 191–220.
- University of Michigan ADVANCE Program. (2009, May). *Positive and problematic practices in faculty recruitment*. Unpublished report available online: http://advance.umich.edu/resources/PositiveAndProblematic_RecruitmentPractices.pdf
- Valian, V. (1998). *Why so slow? The advancement of women*. Cambridge, MA: MIT Press.
- Valian, V. (2014). Interests, gender, and science. *Perspectives on Psychological Science*, *9*(2), 225–230.
- van Arensbergen, P., van der Weijden, I., & Van den Besselaar, P. (2012). Gender differences in scientific productivity: A persisting phenomenon? *Scientometrics*, *93*(3), 857–868.
- Volkwein, J. F., & Sweitzer, K. V. (2006). Institutional prestige and reputation among research universities and liberal arts colleges. *Research in Higher Education*, *47*(2), 129–148.
- Word, C. O., Zanna, M. P., & Cooper, J. (1974). The nonverbal mediation of self-fulfilling prophecies in interracial interaction. *Journal of Experimental Social Psychology*, *10*(2), 109–120.
- Xie, Y., & Shauman, K. A. (1998). Sex differences in research productivity: New evidence about an old puzzle. *American Sociological Review*, *63*(6), 847–870.