9 Evaluating and Promoting Faculty

A group of senior faculty arrived at a two-day workshop aimed at formulating recommendations for institutional policies and practices around "faculty evaluation." One of us was a participant-observer of the process. Although all of the faculty participants were tenured, and most were full professors, they were diverse in terms of their disciplines, their time at the institution, their age, gender, race or ethnicity, and sexual orientation. The workshop began with small groups of the assembled faculty—organized to be diverse in composition—sharing their best and worst experiences of being evaluated or evaluating their colleagues across their careers.

Groups were surprised to discover that most people recounted a "worst" experience, rather than a "best" experience, about an unforgettably painful early career review when they felt invisible, unheard, or unvalued in the context of the review process. Often, these accounts included experiences that seemed like gross inequities or injustices. This was true of both women and men, of White faculty and faculty of color. Although these senior faculty also identified "best" experiences, the positive stories arose mostly from evaluations they had participated in on the *other* side of the fence, as evaluators. In these cases, they had striven to implement practices that improved on the ones employed in their own painful experiences.

Why is it that these successful senior faculty—chosen as opinion leaders on campus—could so readily recall incidents in their careers when efforts to assess their performance felt inadequate, unfair, humiliating, or even degrading? We believe it is because most academic "performance reviews" are conducted by individuals who do not have a clear understanding of the goals of the review process or of how such a process might work to benefit both individual faculty and the institution. These individuals—nearly always faculty members and faculty administrators—are neither stupid nor

(usually) aiming to inflict pain (such as humiliation or alienation). However, without a better articulated understanding of the goals of the review process, and an understanding of how best to reach those goals, both faculty reviewers and targets of review may create or participate in reviews that are experienced as either meaningless or harmful. Our aim in this chapter is to offer guidelines for evaluating faculty so that good outcomes and efficiency are maximized, and wasted time and needless pain and alienation are minimized.

Why Do Faculty Get Evaluated?

Quite apart from formal performance reviews, faculty are evaluated frequently in the course of their work lives: their grants and publications are subject to peer review, and their teaching is subject to student and peer evaluations. As a result, many faculty think that they get a lot of feedback—perhaps more than enough!—on the end products of their work but are less likely to think they receive feedback that helps their day-to-day efforts. The labor that is reflected in any given grant proposal or publication, or a given course, takes months or years and usually involves only the indirect or informal feedback that faculty have sought or noticed—for example, by talking ideas or drafts over with colleagues, students, and family.

Institutional evaluation processes—where they exist—occur on a different timetable. They arise not because a project or a course is complete, but because 12 months—or some multiplicative function of 12 months—has occurred since the last evaluation. Moreover, the assessments are not of a single product, but something more, though it is not always clear precisely what. Sometimes they seem like assessments of one's value to the institution, partly because they are tied to salary increases, but also because they are usually presided over—and sometimes conducted by—an institutional official such as a department chair or dean. Alternatively, assessments sometimes seem more like predictions of success at leaping over the next big hurdle, such as looming tenure or promotion reviews. They can also feel, perhaps most often to the least experienced faculty, like measures of one's value as a human being. Some of this ambiguity may be unavoidable, and very personal, but we think it would help for institutions to be as explicit and clear as possible about the *intended* meaning of institutionally mandated evaluations.

Should We Eliminate Faculty Evaluation?

A radical possibility would be to eradicate formal institutional evaluation processes: they take a lot of time, are difficult to do well, and have some counterproductive effects, as the discussion alluded to earlier made clear. While this might seem a tempting solution, it is unlikely to happen. Worker appraisal began to be a regular feature of the U.S. workplace in the 1940s, and faculty evaluation processes followed along. In the 1990s, calls for greater accountability of universities and their faculties were accompanied by more and more intense calls for systematic evaluation of professors (Curry, 2006). Interestingly, though faculty chafe under frequent formal evaluation, they also resent not being evaluated systematically at all. If there is no systematic process of faculty evaluation, it is easy for faculty to feel—and to be—unrecognized for their accomplishments in scholarship, teaching, service, or leadership. In short, to paraphrase Winston Churchill on democracy: while systematic faculty evaluation may be a very bad system, it is better than others that have been tried.

Treating Faculty as Employees

The link between faculty performance reviews and worker evaluation raises another thorny issue: faculty do not generally feel, nor do they want to feel, like employees. We prefer to think of ourselves as "professionals"—more like (self-employed) doctors, lawyers, and entrepreneurs (all of whom are subject to their own evaluative processes, but that's a different issue). Thought about this way, institutional evaluation processes appear to reduce our uniqueness, independence, autonomy, and creativity, both because they treat us as if we are mere workers and not professionals and because they assess everyone with the same metrics on the same timetable. As with the previous issues mentioned, there is some merit to this objection. No faculty member enacts the professor role exactly like any other, and faculty evaluation processes should be sensitive to the differences among us. There are many different ways of doing a good job and of being valuable to the institution.

At the same time, the elements of our work lives do constitute a job, and it should be possible to demonstrate that we are doing that job well—or not. This should be possible for faculty to know for themselves, and it should be possible for the people who evaluate and report on their performance, even if they are temporary leaders among equals, as department chairs and deans. Institutional evaluation processes, done right, could provide crucial feedback

both to individual faculty about how they are doing, and to chairs and deans who are responsible for their collective work lives and for maximizing the satisfaction and success of all of their faculty (whether they see it that way or not!).

Evaluation Processes as Feedback Processes

Most senior faculty freely agree that junior or untenured faculty deserve and need feedback about how they are doing (though not all institutions provide systematic evaluations). The stakes are high for untenured faculty (since they may lose their livelihood if they do not secure tenure), and therefore someone should provide them with feedback about how well they are doing at the features of the job that are associated with successful tenure and promotion. This right to feedback and the need for it strikes us—more controversially—as true for all faculty. The stakes in terms of pride in our work, and a sense of being valued and appreciated for our contributions, are high for all faculty, though some may bristle at the idea of the value of feedback from their local colleagues.

Changes across the Faculty Career

Most professors have long careers; their interests, enthusiasms, and accomplishments change over time, as we discussed in chapter 8. There are some persistent beliefs about career paths, including the notion that major breakthroughs, particularly in science fields, occur early in a person's career. Though this idea has been challenged many times, a recent study that considered evidence across disciplines and institutions showed that only one-fifth of faculty—regardless of field—have a productivity pattern consistent with that notion (Way, Morgan, Clauset, & Larremore, 2017). Most faculty do not, and the variability in other patterns is substantial. For that reason, institutions must adopt evaluation procedures that are sensitive to the wide diversity of patterns in productivity among the faculty.

Moreover, changes in knowledge and interests might be fostered and facilitated by an institution that views evaluation as not merely tied to salaries, or as assessments of adequacy, but as a way to think about research support needs and different teaching and service assignments. In short, the process of evaluation is a form of two-way communication between individual faculty and a department, school, or college about an individual's work, communication that can be satisfying and generative, harmful and alienating, or simply absent.

Are Issues of Equity Relevant in Faculty Evaluation Processes?

Evaluation processes involve judgments, and we have seen how subject human judgment is to cognitive biases. It is reasonable to assume, therefore, that faculty evaluations are no less subject to bias than evaluations of job candidates. For that reason, issues of equity are relevant. Thorngate, Dawes, and Foddy (2009) identify three bad cognitive "habits" that could produce inaccurate judgments: relying on memory of past events rather than making a new judgment ("she had a strong tenure case five years ago; she's a star"); making holistic rather than disaggregated or specific judgments ("he's always been terrific"); and being inconsistent in the weighting applied to different criteria (one person assessed as having published enough despite no grant; another as not having a grant despite enough publications). These three habits alone raise questions about how likely it is that faculty evaluations could be accomplished free of error.

The problem of bias, however, is different; it is based on the specific concern that the accomplishments of some groups of people are overestimated and those of others are underestimated. If this is true when faculty are being evaluated, then issues of equity across groups are a concern. We bring in several lines of research suggesting that equity is a concern.

Reviewing Faculty as Individuals or Members of Groups

We have previously discussed individuating information. In principle, the more we know about people, the less likely we should be to judge them as members of their demographic group and the more likely we should be to judge them as individuals (Landy, 2008). Thus, one might argue that evaluations of people we know (our colleagues) are fundamentally different from evaluations of strangers (entry-level job applicants). On that view, we should not be subject to schemas or stereotypes—and therefore equity issues should not arise—when we are assessing someone we know well as an individual since we possess much more information about the individual and that should prevent us from relying on schemas (Landy, 2008).

Biases in Reviewing People We "Know"

There are several reasons to doubt that we can rely on individuation to protect us against judging people on the basis of their gender or other demographic characteristics (Maynard & Brooks, 2008). We have discussed

them in detail earlier in the book but provide a brief reminder of them here. First, we do not always know every colleague we evaluate very well. At a minimum, in large departments, or when evaluating faculty in other departments, there is a wide range of familiarity; we may know one individual we review quite well as a mentor or colleague (or think we do), and another person may only be someone we have seen in the halls. Second, and more importantly, if we relied on schemas in our judgments about an individual at the beginning of a relationship (e.g., upon meeting or recruiting the person), as we know is likely, two cognitive processes, anchoring and confirmation bias, will make us unlikely to change our views (Maynard & Brooks, 2008). Anchoring refers to how slowly we alter our initial impressions. Confirmation bias refers to our tendency to notice evidence that confirms our preexisting ideas rather than evidence that does not. While evaluations of colleagues are different in some ways from evaluations of job candidates, schemas can play a major role in both, for somewhat different reasons.

The Problem of Shifting Standards

A different argument has been made by social psychologist Monica Biernat (see Biernat, 2012). In reviewing many years of research on the role of stereotypes in people's judgments, she notes that members of commonly stigmatized groups (women, racial-ethnic and sexual minorities, etc.) are often viewed differently when judged on their performance in stereotype-consistent domains (where they are rated higher) than in stereotype-inconsistent domains (where they are rated lower), given the same performance. Biernat's hypothesis is that members of commonly stigmatized groups are implicitly compared with stereotyped expectations in both cases, resulting in different and inconsistent judgments because of reference to "shifting standards." Thus, women or racial-ethnic minority group members may be judged as competent or talented "for a woman" or a minority, but as incompetent or not that talented when judged in terms of qualifications for tenure.

Attributions of Credit

Finally, research on attributions of credit suggests that both gender and race of the employee affect the judgments of employers. For example, supervisors are less likely to attribute women managers' performance to ability than men's. In addition, they are less likely to attribute African American

managers' performance to ability or effort than Whites', and more likely to attribute it to help from others (Greenhaus & Parasuraman, 1993). Women who demonstrate competence and achievement simultaneously suffer by negative judgments, for example, of their likability (Heilman & Okimoto, 2007; Rudman & Glick, 1999). This is sometimes called a "backlash effect" (Phelan, Moss-Racusin, & Rudman, 2008; Phelan & Rudman, 2010).

A particularly relevant study (Heilman & Haynes, 2005) showed that when performance was accomplished through collaboration (as it often is in the academy), the work of a mixed-gender dyad was attributed to the male (with obvious implications for our ability to understand the true contribution of a faculty member). In the study, judges received a packet including a summary of a trained observer's notes describing each of the two workers (one male and one female) and the overall performance on the task. The task chosen was intended to be male sex-typed: each team was to create an investment portfolio that would yield maximum return over 20 years. The two employees were described with characteristics that were equivalent in qualifications (e.g., a degree in accounting-management or accounting-finance).

Judges received ratings of the team's task performance, defined either as "individual assessments" or "group assessments." When the successful performance was the result of joint work (the group assessment), women were rated as less competent, less influential, and as having played less of a leadership role than comparable men. Moreover, because individual performance was also assessed, it was clear that it was the negative impact of collaboration on the women's ratings that drove these results (rather than a positive impact on the men's). In short, there is reason to believe that when it is unclear who did the work (because it was collaborative), men are given the benefit of the doubt and women are not. In a recent field test of the consequences of academic coauthorship among faculty in two fields differing in the ambiguity of contribution, Sarsons (2017) showed that in a field in which coauthorship does not signal degree of contribution (economics), women coauthors fare badly in terms of tenure rates, and in one where it does (sociology), women coauthors do as well as men.

An important study clarifying the limits of "backlash" effects on women showed that they did not apply to African American women, even though they did apply to White women and African American men (Livingston, Rosette, & Washington, 2012). Of course, African American women are

subject to other biases (higher levels of invisibility) that do not as often apply to either White women or African American men (see, e.g., Purdie-Vaughns & Eibach, 2008; Sesko & Biernat, 2010).

The Impact of Group Stereotypes

The "stereotype content model" (Fiske, Xu, Cuddy, & Glick, 1999; see also Fiske, 2011) suggests that the trade-off of beliefs about competence and warmth arise not only in gender stereotypes but also in schemas about race-ethnicity, with some groups (e.g., Asian Americans) viewed as competent but not likable, and some as likable but not competent (African Americans and Latinos). Racial and ethnic schemas affect judgments of both people and of occupations—with high-status occupations and leadership roles generally stereotyped as "White." For example, in one study of applicants for high-status jobs, White male judges rated Asian Americans with low-quality credentials as positively as African Americans with high-quality credentials (King, Madera, Hebl, Knight, & Mendoza, 2006).

The existence of so many studies showing that human judgment processes are subject to bad habits that result in error and are saturated with the influence of group-based schemas or stereotypes suggests that we should be attuned to the possibility of inequity in faculty evaluations. There is more direct evidence as well. There is extensive documentation of salary inequity among faculty, in statistical analyses that control for disciplinary, rank, and experience differences (see, e.g., Barbezat, 2002; Jagsi, Griffith, Stewart, Sambuco, DeCastro, & Ubel, 2012; Porter, Toutkoushian, & Moore, 2008; Toutkoushian, Bellas, & Moore, 2007; Toutkoushian & Conley, 2005). If faculty are hired according to the same criteria, then only differential performance should produce salary differences, and there is little evidence for overall performance differences by gender within types of institutions (Xie & Shauman, 2003).

Equally, there is considerable evidence that women spend longer within a particular rank (especially the associate professor rank) than men (Ash, Carr, Goldstein, & Friedman, 2004; Committee on the Status of Women in the Profession, MLA, 2009; Giesler, Kaminski, & Berkley, 2007; Valian, 2000). There is less evidence about the impact of race and ethnicity on salary and on time within rank (but see Barbezat, 2002; Fang, Moy, Colburn, & Hurley, 2000; Hammond et al., 2010; and Castilla, 2008 on race in organizations), but many faculty of color believe that their work is undervalued,

including in terms of salary and time to promotion (Stanley, 2006; Stanley, Porter, Simpson, & Ouellett, 2003; Turner, Gonzalez, & Wood, 2008). One woman of color described the experience of Black women this way:

Most of us feel ourselves obligated to take on certain kinds of tasks. Sometimes you are called upon to be an advisor to a Black student group. You may be asked to serve on a curriculum committee, and you know your presence on the committee will affect the incorporation of certain types of courses into the program, or to be on an admissions committee or review committee where you know that your voice makes a difference. Whether these are undergraduate or graduate students, your voice would make a difference in the discussion. It is very hard to turn some of these things down…. Some of that I don't think you can or should decline. (Harrison, 2016, p. 57)

The experience of faculty evaluation processes is not always the same for faculty of color and White faculty. One scholar describes tenure as a "contact zone" and suggests,

In this zone, the fate of a professor and the relationships between different ethnic, linguistic, and academic communities with different levels of access to power, knowledge and voice are negotiated through "long-term" and often "intractable" "unequal conflicts," contacts (*encuentros*), and ruptures (*desencuentros*). (Chabram, 2016, p. 95)

It is therefore important that faculty evaluations both *be* fair and *be perceived* to be fair (Curry, 2006).

Principles for Fair Faculty Evaluation Processes

Although there are some specific issues associated with each of the different types of faculty evaluations (annual reviews, third-year or midcareer reviews of assistant professors, tenure and promotion reviews, posttenure reviews), there are common features of faculty evaluation processes that are likely to be fair and to feel fair. Many of these are outlined well by Curry (2006). We will revisit many of these elements as they pertain to particular types of reviews.

First, *the review should be conducted by more than one individual*. Generally, reviews conducted by multiple evaluators are trusted more than those conducted by a single individual: they are subject to consensual validation (Thorngate et al., 2009). The assumption is that the separate formation of impressions by even two evaluators makes it less likely that an individual will form and maintain a biased perspective on the evidence. There is in fact reason to believe that multiple observers may produce more complete

and more accurate judgments than a single observer (see, e.g., Vazire & Gosling, 2004, who show that aggregated judgments, based on limited information but multiple observers, are more accurate than those by single observers).

Second, *the review should be based on clearly articulated criteria applied to consistently provided evidence*. Both the faculty member being evaluated and the reviewers should know and understand the criteria that will be used in evaluating the record. This means that the criteria should be explicit and written down so they can be consulted by all parties. Moreover, the criteria should relate directly to larger institutional values and goals (Perna, 2001; Stewart, Dalton, Dino, & Wilkinson, 1996; Tierney & Bensimon, 1996).

Some institutions clearly express expectations in terms, for example, of average student teaching ratings or the number of publications expected before tenure (see, e.g., this table for the University of Michigan Medical School: http://faculty.medicine.umich.edu/appointments-promotions /promotion-process/promotion-benchmarks/, visited October 17, 2017) or the impact factors of the journals that faculty publish in (Ohio State University College of Medicine (https://medicine.osu.edu/faculty/oecrd/Documents/ed _journals20110727.pdf, visited October 17, 2017). Other departments may spell out benchmarks and criteria for achieving positive outcomes (https:// biology.uiowa.edu/sites/biology.uiowa.edu/files/Biology--Faculty%20Promo tion,%20Review%20and%20Assessment%202012.pdf, visited October 17, 2017). Yet other departments claim that precise statements are impossible and that only the broadest possible statement of demands for "quality" or "productivity" can be made. This strikes us as incorrect because it leaves too much of the decision to the discretion of the evaluators and thus risks evaluators' using different standards, depending on whether they are favorably predisposed to the candidate.

Ideally, a department or school strikes a happy medium, being neither vague in a way that encourages evaluation errors nor overspecific by using criteria that may not fit all candidates (e.g., an exact number of publications, particular funding sources, only certain publication outlets). A moderate degree of specificity will make the criteria clear not only to reviewers but to those being reviewed. Sometimes the records of those who were judged favorably in previous years can be provided as information, either in the form of their actual CVs or essays prepared for the review (many of these are password-protected and aimed only at internal faculty) or in terms

of averages, modal figures, and the ranges (which often vary widely and therefore are not so helpful) across candidates. In the latter case, information about a range of publications (journal articles, monographs, edited collections) or average teaching ratings can provide guidance about the acceptable ranges that have recently been associated, within a particular institution, with higher raises, reappointment, tenure, or promotion. One medical school posts this sort of information on its website (http://faculty .medicine.umich.edu/appointments-promotions/promotion-process /promotion-benchmarks/, visited January 22, 2018).

The evidence to be reviewed (the candidate's dossier) should include evidence that is pertinent to the evaluation criteria. Guidelines for candidates' preparation of the dossiers should spell out the evidence needed in sufficient detail that all candidates can provide complete information. It is the responsibility of the institution or the evaluators to ensure that the form of the dossier is well-defined for all candidates. It is then the responsibility of the person being evaluated to provide careful, accurate information.

Curry (2006) recommends that evaluations assess performance in the context of an individualized faculty career development plan. Thus, for example, an individual's productivity in a given year might not be in terms of published papers or completed book chapters if that was a year of collecting a large set of data, learning a new technique, or working in an archive on a new project. Instead, progress in those domains (data collection, technical skills, or archival work) might be assessed for that year. A highly individualized review process like this maximizes the likelihood that evaluation criteria are tied to an individual's career stage and personal goals. This may be a particularly appropriate element in reviews that include a serious mentoring element (e.g., the reappointment review or annual reviews). However, it is difficult to imagine that most institutions would be willing to give up reliance on some kind of absolute standard of judgment (not one relative to the individual's goals) for decisions like tenure or promotion.

Third, *the review process should be transparent*. Thus, the reviewers, the procedure, and the steps in the process should all be explicitly defined according to a known timetable that is made clear to both the person being evaluated and the evaluators. Opportunities for input, as well as exclusions of material, should be clearly defined.

Fourth, *service as a faculty reviewer, within the institution and outside, should be understood to be a privilege, and one that requires skill, knowledge, and respect*

for confidentiality. All reviewers should be selected or elected based on the judgment that they are capable of fair assessment of the accomplishments of peers. Moreover, they should be willing to be educated—and should in fact *be* educated—about the role of bias in evaluation processes and the steps that must be taken to minimize it. This is obviously the responsibility of the institution with regard to internal reviewers. For outside reviewers, institutions can try to make clear what type of evaluation they want and should pay little or no attention to reviews that have internal evidence of reliance on bias in their review.

Since we have demonstrated that faculty exhibit biases in judgments of each other, it is no surprise that students also express biases in their judgments of faculty (recent studies include Graves, Hoshino-Browne, & Lui, 2017; Storage, Horne, Cimpian, & Leslie, 2016). It is worth noting that in the context of faculty performance reviews, students who provide evaluations of classes are also serving as reviewers of faculty performance, and they are usually uneducated about the ways that group-based biases may enter into their judgments. Some institutions have adopted strategies (e.g., in the directions to students on the actual evaluation forms) that might minimize the intrusion of these biases in the process. In any case, the faculty evaluators should be made aware of the risk of influence of bias in these ratings, particularly for members of some groups.

Fifth, *the review* should not only be conducted according to fair criteria and procedures, but it *should be experienced as fair by the person being reviewed and by those conducting the review*. Reviewers should be held accountable for the quality of their evaluation at the level to which they report (school or college for departments and programs, provost and president for deans).

Sixth, *all faculty who are evaluated should be treated respectfully in the course of the review*. *This respectful treatment includes thoughtful feedback about their performance, regardless of the decision outcome*. Wherever possible, this feedback should be viewed as providing information that fosters an individual's successful career trajectory (Perna, 2001; Tierney & Bensimon, 1996). Recommendations vary about the value of written or oral feedback from a committee or from an individual, but there is no doubt that explicit, respectful, and detailed feedback increases faculty members' trust and investment in the process.

Finally, *criteria should be periodically reviewed* to determine whether the criteria need to be altered or expanded. In recent years, several institutions

have adopted new criteria or expanded their criteria to include elements that had previously been ignored. For example, at Syracuse University, the faculty manual was amended in 2005 to recognize "publicly engaged scholarship" as relevant to tenure reviews along with more traditional approaches:

Syracuse University is committed to longstanding traditions of scholarship as well as evolving perspectives on scholarship. Syracuse University recognizes that the role of academia is not static and that methodologies, topics of interest, and boundaries within and between disciplines change over time. The University will continue to support scholars in all of these traditions, including faculty who choose to participate in publicly engaged scholarship. Publicly engaged scholarship may involve partnerships of University knowledge and resources with those of the public and private sectors to enrich scholarship, research, creative activity, and public knowledge; enhance curriculum, teaching, and learning; prepare educated, engaged citizens; strengthen democratic values and civic responsibility; address and help solve critical social problems; and contribute to the public good. (Syracuse University, 2005, *Faculty Manual*, section 2.34 on Tenure)

The University of California system recommends that criteria must be understood flexibly (see University of California, 2005, Academic Personnel Manual, APM-210, section d):

As the University enters new fields of endeavor and refocuses its ongoing activities, cases will arise in which the proper work of faculty members departs markedly from established academic patterns. In such cases, the review committees must take exceptional care to apply the criteria with sufficient flexibility. However, flexibility does not entail a relaxation of high standards.

The document further specifies that

Teaching, research, professional and public service contributions that promote diversity and equal opportunity are to be encouraged and given recognition in the evaluation of the candidate's qualifications. These contributions to diversity and equal opportunity can take a variety of forms including efforts to advance equitable access to education, public service that addresses the needs of California's diverse population, or research in a scholar's area of expertise that highlights inequalities. (APM-210)

An alternative approach to a similar issue—probably less effective over the long term because not codified in official policy and therefore safely ignored—was taken by the provost at the University of Michigan in 2012 (and restated by later provosts). He sent an e-mail to all faculty pointing out the importance of including attention to "the broad range of entrepreneurial, outreach, and creative activities in which faculty engage" in

all faculty review processes. He noted that these included "creating service learning and action-based learning opportunities for students" and "engaging in community-based research," among other examples (Pollack, 2015).

To summarize, fair evaluations require procedures that will buffer evaluators from the errors that evidence suggests can enter the review process. Having multiple reviewers ensures that more features of a candidate will be assessed; having specific criteria guards against emphasizing or deemphasizing criteria depending on who is being evaluated; a fair, transparent, and respectful process gives everyone confidence in the institution; knowledge about the demographically based errors that are inherent in judgments helps evaluators guard against them; periodic review of criteria keeps an institution up-to-date with the range of valuable knowledge that faculty can provide.

Annual Reviews of Faculty

As with all reviews, the purpose of annual reviews should be clear. If, for example, they are to be used as one basis—or the primary one—for allocating merit-based raises, this should be explicit so faculty understand the implications of this evaluation. If the reviews will be used to provide substantive feedback about how well the individual is performing and thus serve as a guideline to the faculty member about whether they are on track for a long-term appointment, then clear criteria must be used to judge performance and detailed feedback must actually be provided. (Many institutions promise feedback but provide such generic feedback that it feels empty and meaningless to the individual.) Finally, if the feedback will be used as a basis for information conveyed to junior or early career faculty about how they are doing, this too should be explicit and should result in direct conversations about how to reach their goals (reappointment or tenure and promotion), as well as the best estimate of whether those goals are attainable.

Who Should Be Reviewed?

Senior faculty may also benefit from periodic reviews if the climate of the department allows for respectful feedback when faculty members are, and perceive themselves to be, at a decision point regarding research, teaching, or service. Such a review cannot be imposed at random or come from

evaluators whom the faculty do not regard true "peers." But for faculty who are facing complex decisions about how to invest their time, an opportunity to share ambivalence and uncertainty and to get feedback from wise colleagues may be valued highly.

Many institutions, but certainly not all, engage in annual reviews of every faculty member. Some require an annual report, described below, from every faculty member, but conduct formal reviews less often and at that time include a review of two or more annual reports. This kind of practice may be warranted when patterns of teaching, scholarship, and service are relatively stable over more than one year, as for more senior faculty. However, it is wise to require faculty to provide reports annually even when they will be reviewed less often so the reports will be less subject to the biases that affect retrospective reports of activities. Annual reports typically include information that may be on a CV (e.g., publications, conference presentations, and service contributions, in a specified time frame) but go well beyond the outcomes covered there, including, for example, research projects under way but not yet yielding conference talks or publications, formal courses taught in that time frame, as well as guest lectures in others' courses, informal teaching through individual tutorials or even small group processes, activities in the community and the discipline, advising of both graduate and undergraduate students, mentoring activities for students and faculty, and contributions to campus and department initiatives that may not be covered by formal committee service.

Annual Activity Reporting

We believe that annual reporting of activities is generally a sensible practice both for individuals and for institutions. It allows individuals to keep track of their activities on a timely basis and it allows institutions to provide formal feedback (in principle via salary adjustments) and informal feedback about how the individual compares with expectations of faculty activity. However, it should not be required of individuals if those reports will not be used by the institution for any purpose that the individual might be expected to value. Thus, if annual reports are never actually read or evaluated, and if feedback is never provided to the individual, requiring the reports is not defensible. In addition, if reports are only used for punitive purposes (e.g., to assign additional teaching to faculty who are not engaged in research activities), there will be little trust or investment in them on the part of the

faculty. Institutions must be wise in requiring labor and disclosure from faculty, and that wisdom includes respectful use of the material that is generated by individuals in the form of reports.

The criteria that will be used in annual reviews should be developed in terms of both broad aspects of the larger institutional mission and unit-specific values. A particular department may have a strong mission associated with teaching first- and second-year undergraduates (e.g., composition, required math and science courses, etc.), or as a large major, or a major feeding into a profession. The criteria should reflect such missions. Equally, a department's scholarly mission may be self-understood to involve rapid publication in online journals or, instead, mainly in printed monographs. These disciplinary and departmental differences should be reflected in the articulation of criteria for the review and in the development of a reporting form for faculty to use in describing their annual activities. A clear timetable for reporting, evaluation, and feedback should also be explicitly articulated.

Consistent Format for Reporting

A reasonable balance must be struck between a uniform reporting format that enables consistent application of criteria and the need for individuals to have ways to report unusual or developing activities. It is helpful to be explicit about precisely how information should be reported. For example, should only work actually newly in print during a particular year be reported? Should manuscripts in press or under review be reported but in separate sections? Should projects undertaken but not yet written up be described? If so, where and how? At the same time, in each of the major areas assessed, it may be useful to provide faculty with an opportunity to describe new projects or commitments. However, if a report of new projects is invited, reviewers and the faculty member must have a clear idea of how this information is to be used. The evaluation process must have built in some ways to guard against evaluating different faculty according to different criteria (systematically or idiosyncratically shifting standards).

How Feedback Is Communicated

In the interest of encouraging communication about important issues, it may be helpful to include an opportunity for face-to-face interaction in the process. This interaction might be a one-on-one conversation between a "primary" reviewer and the individual, or it could be a conversation between the committee and the individual. The latter kind of procedure

may, however, disadvantage faculty who feel they lack status or standing in the department. It may also be helpful to ask faculty members to summarize the key points they took away from the review in a postconversation e-mail to ensure that there is agreement on what was said.

If reviews will affect salaries, department chairs must recognize the key role of salary in faculty satisfaction and retention. Judgments and the resulting salaries should reflect the behaviors that a chair hopes to encourage (e.g., good citizenship, selfless contributions to the department, institution, or discipline), and address the trade-off between high-performers-take-all approaches and sharing rewards more broadly across the faculty.

Special Issues in Annual Reviews of Untenured Faculty

Finally, annual review processes for untenured faculty should be accompanied by detailed individual feedback, communicated with the aim of maximizing the individual's chances for reappointment and then promotion and tenure. One vehicle for communicating feedback is a written letter (perhaps drafted by one individual but always shared and revised in a group); another is a face-to-face meeting. Both can be used to good effect. We recommend sending a letter and following up with a meeting in which the inevitable ambiguities of interpretation of feedback can be addressed in person. An example of such a letter and the questions it might elicit from an untenured faculty member is provided in box 9.1. The questions are included to help readers imagine what a junior faculty member may find unclear or confusing even in an explicit letter like this one. Only face-to-face conversation with the chair and perhaps other senior faculty can address the questions that inevitably arise.

Reviews for Reappointment of Junior Faculty

The Goals of Reappointment Reviews

Many institutions have created a review process that assesses junior faculty for reappointment on the tenure track before they are reviewed for tenure and/or promotion. These reviews are often intended to provide both the institution and the individual with information about how the individual is progressing toward tenure and promotion. In some institutions there are two or more two-year reappointment reviews before the tenure review; in others there is one midcareer (third- or fourth-year) review during the "probationary period" as an assistant professor. Reappointment reviews

Box 9.1
Evaluation Letter—Untenured Assistant Professor, Prior to Reappointment

Evaluation Letter	Recipient's Questions
Dear Untenured,	
It is my responsibility to annually evaluate the performance of each faculty member, taking into consideration input from the department advisory committee. The following assessments follow the expectations identified in the letter of offer you received prior to your appointment.	
Teaching: Your primary assignment has been to teach the first year theory sequence and the **two sections of XXX 151**, as well as a graduate seminar. Teaching the theory sequence has been a considerable challenge. We have talked about the student evaluation forms in which your composite overall rating for these courses is 3.5 out of 5.0. This is below **the department threshold for being considered excellent, which is 4.5**. The examination of your teaching portfolio does indicate that you have put considerable effort into your courses. However, student comments indicate that your lectures seem to lack variety in teaching techniques and could be better organized. The large class setting posed particular difficulties that you continue to work to overcome. Your willingness to consult with others who have experience teaching large required undergraduate courses seems to have been helpful, and I noticed an **improvement in the student evaluation scores in the spring semester**. It was clear from them that you know the material very well and that you are a dedicated teacher. I suggest that you examine closely the comments made by the peer review teaching committee. **Members of the committee would individually welcome a dialogue** with you about these issues. I would also encourage you to continue videotaping class sessions in the coming academic year. With respect to the graduate seminar, all indications from the student evaluation forms suggest that you did an excellent job.	No comments are made anywhere in this report about the two sections of XXX 151; is it going well? Am I expected to reach that level ("excellent") in all classes? Are the scores obtained this term sufficiently high, if I can maintain them, or are they still too low? Those are all busy people; how can I approach them? Should I?

Box 9.1 (continued)

You should consider whether any of the teaching strategies that work so well in this seminar could be adapted to your other courses, different as they are in level and size. Your teaching performance *has been rated as acceptable*.

How could they possibly work in a large undergraduate class?

That sounds OK; is "acceptable" good enough?

Research: With regard to research, the advisory committee and I agree that you are making good progress. Your article in Journal of MMM and the one accepted in the Journal of RRR are a good start. *These are Tier One journals in our department listing of journals, and you should continue to publish in them*. In looking ahead, however, you need to establish a thematic focus and a research agenda that will be cumulative in building your national reputation. I will ask the mentoring committee to help you in this process. The development of a career plan that includes a *statement about your research direction* would prove helpful here. Your research performance has been rated as excellent.

What department listing? Where is it? Do you mean I have to only publish in these two journals?

Don't these things emerge from my work? I need to just keep publishing and doing projects and the thematic focus will be clear—right?

Service and Outreach: You participate actively in faculty meetings and on the department's admission committee and you were a panel member at this year's major outreach conference. This is a good level of service for a beginning faculty member. Your service performance has been rated as excellent.

Summary and Suggestions, Progress Toward Tenure: Overall you have made a good start. As indicated, you will have to continue to work to improve your teaching in the theory sequence; *good teaching is required for tenure within the department*. Because the tenure decision takes a cumulative look at the total body of your work and its impact on the field, *it is imperative that you articulate your research focus and theme more clearly*. The mentoring committee and I will be glad to help. If this letter does not accurately summarize our discussions, or if you have additional questions or concerns about any of its content, please let me know as soon as possible.

How good? Do I have to be excellent in all kinds of teaching?

I'm not sure how I can articulate a focus when I don't know yet which of my efforts will yield results. Seems to me I should just keep working and the "theme" and "agenda" will emerge as I do. Is that wrong?

Letter reprinted from Curry (2006, p. 11) with permission. Possible questions from junior faculty added.

are generally more extensive and thorough than annual reviews (in part because fewer individuals are reviewed), but there is a great deal of variation in the materials used for these reviews and in the probability of reappointment in different institutions.

The Timing of Reappointment Reviews

The path to a tenure review is only about six or seven years at most institutions that have formal tenure systems (currently about half of higher education institutions; NCES, 2016), and at many schools the tenure decision is a highly consequential "up or out" decision, with no opportunity for rereview (Trower, 2002). We propose that reviews for reappointment are better suited to providing feedback and mentoring to untenured faculty than to making ultimate employment decisions. There may be a small number of faculty whose performance in the early years of a tenure track appointment is so problematic that their department believes they should be terminated before a tenure review. In such cases, it is crucial not only to have a process that allows for a fair and thorough review of the (short) record of accomplishment, but also constructively assesses the prospects for accomplishment in the future. This latter estimate is inevitably based on relatively little data and is therefore subject to significant prediction error: particularly in the early stages of a career, the nature of a faculty member's project and the kind of publication sought may offer little or no reviewable evidence.

Criteria for Reappointment Reviews

Since reappointment reviews are more consequential than annual reviews (if less so than tenure reviews), the criteria for review must be well-defined and clearly communicated to all parties. In addition, as previously noted, candidates need clear guidelines for how to prepare their dossier for review by others. Access to a template or form is often helpful. For example, the University of Vermont offers both general advice about the whole dossier and detailed information about the information that must be provided by the applicant for reappointment (https://www.uvm.edu/~facrsrcs/MASTER _BlueSheetform_Apr2016.docx and https://www.uvm.edu/~facrsrcs/Green BlueSheetInstructions.pdf). Departments, schools, and colleges can often also ask individuals recently successfully reviewed and reappointed to share their dossiers with new applicants so they have a potential model to follow.

The Questionable Role of Outside Letters in Reappointment Reviews

Some institutions include outside letters in interim reviews; many do not. A number of considerations are relevant. One powerful reason not to request external letters is to avoid signaling the trajectory of a highly attractive candidate to individuals at other institutions. The individual at this stage has not yet provided benefits to the institution for the labor and financial investment that has already been made in hiring, recruiting, and support. Another reason is avoidance of adding to the already substantial reviewing burden entailed by tenure and promotion review processes. In addition, external reviews of such an incomplete record of accomplishment may have too much impact on a review process at this stage, when the local institution has much more information than external people have about the junior person's developing career. We will review in the next section how questionable it is to rely on external letters even in the tenure process itself, because of their special vulnerability to biases. For all of these reasons we recommend against reliance on letters at this early stage in a faculty member's career.

The Role of Feedback in Reappointment Reviews

Feedback to the junior faculty member is the single most important element in reappointment reviews (Perna, 2001; Tierney & Bensimon, 1996). We strongly recommend both that feedback be in writing and that it be discussed in a meeting with the candidate. That meeting should include the department chair and at least one senior faculty member other than the chair who knows the faculty member well—ideally someone who has served as an advisor. Early career faculty often have difficulty interpreting feedback, and may both over- and underrespond to what they think they hear. Having seriously engaged senior faculty discuss the feedback with the junior faculty member can help guard against ongoing misinterpretations. Moreover, the senior faculty involved must gauge the relative need of any particular faculty member for both clarity about the seriousness of the criteria for tenure and promotion and their distance from it, and encouragement that the standard can be met. There are large individual and group differences in faculty members' confidence, accuracy of interpretation, and attributions about their own situation (to lack of effort, ability, support, etc.).

Sensitive, attentive senior faculty must take responsibility for trying to convey feedback in a way that fits the receiver and is accurate. Junior faculty have their own responsibilities: to review feedback carefully and to try to calibrate its implications for their efforts. The more constructive the feedback is, in content and style, the more likely it is that the faculty member can interpret it correctly and make use of it. Asking specific questions is the best tactic the faculty member can use to minimize the likelihood of misinterpreting the implications of feedback.

This communication process—like all interpersonal communication—is fraught with complexity. Both senders and receivers are subject to the cognitive biases that we have discussed throughout this volume, but there are some that are particularly likely to be mobilized in this kind of profoundly consequential interaction. Differences in the social groups we all belong to may exaggerate our difficulties: junior faculty are more likely to be women or from another underrepresented minority group (or both) than are senior faculty. Evidence suggests that White women and racial-ethnic minorities are most disposed to trust feedback that conveys that high and transparent standards have been and will be employed, along with confidence that the individual can meet those standards (see Biernat, 2012, for a summary; specific studies include Cohen, Steele, & Ross, 1999; Crocker, Voelkl, Testa, & Major, 1991; Roberts, 1991; Roberts & Nolen-Hoeksema, 1994).

Responsibilities of Evaluators

The senior faculty evaluating the dossier and the individuals conveying the feedback must work to avoid being influenced by group-based schemas in their judgment—either underestimating accomplishments because of reliance on a schema suggesting members of the candidate's group cannot achieve at a high absolute standard or overestimating accomplishments because of reliance on a schema suggesting that the individual is performing better than the group normally would. Low absolute ratings of accomplishment accompanied by high praise from a reviewer should be a red flag that the reviewer is being influenced by using a within-group standard to produce praise and an across-group comparison to provide a low rating. Such a pattern, as mentioned earlier, arises from the influence of group-based stereotypes. Implementation of transparent and consistent standards is the best way to counteract this influence (see Biernat, 2012; Thorngate et al., 2009).

Decisions Not to Reappoint

The painful and infrequent use of reappointment reviews to make a decision not to reappoint must be handled with particular care if confidence in the fairness of the process is to be retained. The fact that this decision must sometimes be made underscores the importance of the composition of the review committee (members must be seen to have the needed expertise and to be fair-minded). In addition, the process must be perceived to have been entirely transparent and based on reasonable criteria if the unsuccessful candidate and the surrounding community of other junior as well as senior faculty are to have confidence in the decision. Finally, humane and considerate treatment of the unsuccessful candidate is crucial, including explicit discussion of the kinds of career advice and mentoring that is still available to them from faculty in the department or school. Decisions not to reappoint inevitably create a rupture between individuals and their colleagues. For that reason, it is crucial to take active steps to ensure that a few individuals are available to provide ongoing support and mentoring as the individual finds an alternate path.

Decisions to Reappoint

Positive decisions to reappoint should also be accompanied by explicit advice about future activities that will maximize the likelihood of tenure and promotion (without suggesting there are any guarantees about future outcomes based on those activities alone). Not only must the feedback about future activities be conveyed carefully and fully, but it is best to outline a plan for the period between the reappointment review and the tenure review. This may be done in consultation with the chair and at least one mentor or by a group of mentors, but it should happen close to the time of the reappointment review and as a result of it.

Tenure and Promotion Reviews

All of the features we have discussed as important in annual and reappointment reviews (e.g., clarity, transparency, fairness of the criteria and the process) apply with special force to tenure and promotion reviews. There is a particular burden on faculty who participate in these high-stakes reviews to be well educated about the potential for biases entering into evaluations of colleagues and for them to be fully capable of committing

to conscientious, fair-minded reviews. Thus, both their selection and their "training" or education about bias and fair procedures deserve special attention and care.

Timing of Tenure and Promotion Reviews

We note that it may be valuable to consider whether the tenure and promotion reviews are well-timed at any given institution. Although we assume in our discussion a model that is specifically designed to accommodate the widest possible range of situations, there are alternative unusual models to consider. For example, one model rejects the classic seven-year review cycle. Instead, faculty members have, say, ten years in which to achieve tenure but may put themselves forward at any time. Faculty can submit a maximum of five publications on which they will be evaluated.[1] One faculty member may think she has five noteworthy publications within three years. Another may think he has such a dossier only after ten years. It is the faculty member's choice, in consultation with more senior faculty. On this model, it is not the quantity but the quality of someone's work that is evaluated.

By allowing ten years, the model obviates the need for any policies concerning stopping the clock or modifying duties. Someone can take as much or as little time as they want within their ten-year period. An intellectual benefit is that faculty would be free to consider longer range projects that might take several years to show a result instead of focusing on short-term projects that will maximize quantity. We are not necessarily endorsing this model, but we put it forward as an example of fresh thinking that could stimulate further thinking about the goals of the tenure review and how they should be reflected in the process.

Formation of the Review Committee

Regardless of the timing of the case, it is optimal for members of any tenure review committee to have a discussion of the issue of cognitive biases at the beginning of their deliberations, without reference to any particular case. We recommend that individuals who dismiss the importance of these biases not serve on review committees on the grounds that they are not knowledgeable enough about the cognitive processes underlying evaluations to conduct them fairly and impartially.

Because most faculty assume they are free of bias, it is important that review committee members be educated about specific practices that will minimize the impact of implicit bias. Among other things, this means that all committee members should understand the importance of maintaining consistent standards both across different candidates and in terms of using a set of clear and well-specified (but not overspecified) criteria. It should be viewed as good and appropriate collegial behavior to point out when others are applying different standards to different candidates. The fairness of the review process is the responsibility of every committee member, not just the chair. Some institutions have adopted mandatory (https://aglifesciences.tamu.edu/about/impacts/diversity/) or voluntary (http://fda.fsu.edu/var/ezwebin_site/storage/original/application/c84e2b1b64d1b9269146f092dd21d15f.pdf) educational sessions for tenure review committee members, conducted either online or in person. These seem to be an excellent practice that will likely raise the quality of the tenure review process of any given campus over time, as knowledge and experience implementing improved practices spreads in the community. Some institutions make one committee member responsible for preventing schemas or bias from entering into discussions (http://www.usu.edu/provost/promotion-and-tenure/ombudsperson.cfm). While this may be tempting, we believe it is a mistake to relieve any committee members of the responsibility for a fair review. For that reason, explicit discussion of the issues, recommendation of specific procedures to maximize fairness, and shared responsibility for the review seems a better course.

Specifying Tenure Criteria

Specifying tenure criteria is a complex matter. Many institutions articulate criteria that are too vague ("excellence in scholarship, teaching, and service"), which means they have failed to indicate any standard at all. Some institutions employ overly specific criteria. One overly specific criterion is a particular level of citations. This is not appropriate because citation rate is strongly influenced by time and subfield size, as well as journal impact and other extraneous matters (Hirsch, 2005; Meho & Yang, 2007; Ruscio, Seaman, D'Oriano, Stremlo, & Mahalchik, 2012). Some departments require that candidates include the impact factor of each journal their work appears in. This, too, is inappropriate, as the San Francisco Declaration of

Research Assessment, primarily aimed at evaluation of researchers in the sciences, spells out. (See http://www.ascb.org/dora) This excellent declaration, presented in box 9.2, makes recommendations to funding agencies, institutions, and researchers to limit the use of impact factors. It emphasizes not using proxies, such as journal impact factor, over a paper's quality. Quality, directly assessed, rather than quantity or other metrics, has been stressed by a range of organizations. For example, the Computing Research Association adopted a "best practices" memo in 2015, for evaluating scholarship, and suggested limiting the number of papers considered to one or two for junior scholars being hired and to three to five for consideration for promotion. (See http://cra.org/wp-content/uploads/2016/02/BP_Memo.pdf.)

Another example of overspecificity is requiring particular student teaching evaluation scores. That is not appropriate because those scores are influenced by field, size of class, and complex aspects of discipline, gender, and race composition of the student body, and all interact with the gender and race of the instructor (Bavishi, Madera, & Hebl, 2010), as well as perceived age and "hotness" (Sohr-Preston, Boswell, McCaleb, & Robertson, 2016). In the past studies of student evaluations of teaching have resisted straightforward conclusions because they relied on field data subject to many uncontrolled influences such as self-selection of courses taught by men vs. women. Recently researchers have used inventive techniques to overcome these limitations.

For example, Boring, Ottoboni & Stark (2016) examined one set of data based on evaluations obtained in a 5-year natural experiment in France, in which students were assigned to mandatory first year courses based solely on scheduling considerations, and students were unable to select their instructor; thus instructor gender was "as if" randomly assigned. Data suggested that gender produced significant and large differences in ratings—so large that they could not be used to assess effectiveness. There were, in addition, non-trivial other irrelevant effects (e.g., by discipline), making it impossible to employ a statistical correction. These results were replicated in a US dataset collected under conditions of random assignment in an online course. In another study in which students were randomly assigned to sections taught by men or women, using identical course materials, women were rated significantly lower than men overall; these results were mostly driven by ratings of male students, and were larger for courses

Box 9.2
San Francisco Declaration of Research Assessment

There is a pressing need to improve the ways in which the output of scientific research is evaluated by funding agencies, academic institutions, and other parties.

To address this issue, a group of editors and publishers of scholarly journals met during the Annual Meeting of The American Society for Cell Biology (ASCB) in San Francisco, CA, on December 16, 2012. The group developed a set of recommendations, referred to as the *San Francisco Declaration on Research Assessment*. We invite interested parties across all scientific disciplines to indicate their support by adding their names to this Declaration.

The outputs from scientific research are many and varied, including: research articles reporting new knowledge, data, reagents, and software; intellectual property; and highly trained young scientists. Funding agencies, institutions that employ scientists, and scientists themselves, all have a desire, and need, to assess the quality and impact of scientific outputs. It is thus imperative that scientific output is measured accurately and evaluated wisely.

The Journal Impact Factor is frequently used as the primary parameter with which to compare the scientific output of individuals and institutions. The Journal Impact Factor, as calculated by Thomson Reuters, was originally created as a tool to help librarians identify journals to purchase, not as a measure of the scientific quality of research in an article. With that in mind, it is critical to understand that the Journal Impact Factor has a number of well-documented deficiencies as a tool for research assessment. These limitations include: (A) citation distributions within journals are highly skewed [1–3]; (B) the properties of the Journal Impact Factor are field-specific: it is a composite of multiple, highly diverse article types, including primary research papers and reviews [1, 4]; (C) Journal Impact Factors can be manipulated (or "gamed") by editorial policy [5]; and (D) data used to calculate the Journal Impact Factors are neither transparent nor openly available to the public [4, 6, 7].

Below we make a number of recommendations for improving the way in which the quality of research output is evaluated. Outputs other than research articles will grow in importance in assessing research effectiveness in the future, but the peer-reviewed research paper will remain a central research output that informs research assessment. Our recommendations therefore focus primarily on practices relating to research articles published in peer-reviewed journals but can and should be extended by recognizing additional products, such as datasets, as important research outputs. These recommendations are aimed at funding agencies, academic institutions, journals, organizations that supply metrics, and individual researchers.

Box 9.2 (continued)

A number of themes run through these recommendations:

• the need to eliminate the use of journal-based metrics, such as Journal Impact Factors, in funding, appointment, and promotion considerations;

• the need to assess research on its own merits rather than on the basis of the journal in which the research is published; and

• the need to capitalize on the opportunities provided by online publication (such as relaxing unnecessary limits on the number of words, figures, and references in articles, and exploring new indicators of significance and impact).

We recognize that many funding agencies, institutions, publishers, and researchers are already encouraging improved practices in research assessment. Such steps are beginning to increase the momentum toward more sophisticated and meaningful approaches to research evaluation that can now be built upon and adopted by all of the key constituencies involved.

The signatories of the San Francisco Declaration on Research Assessment support the adoption of the following practices in research assessment.

General Recommendation

1. Do not use journal-based metrics, such as Journal Impact Factors, as a surrogate measure of the quality of individual research articles, to assess an individual scientist's contributions, or in hiring, promotion, or funding decisions.

For funding agencies

2. Be explicit about the criteria used in evaluating the scientific productivity of grant applicants and clearly highlight, especially for early-stage investigators, that the scientific content of a paper is much more important than publication metrics or the identity of the journal in which it was published.

3. For the purposes of research assessment, consider the value and impact of all research outputs (including datasets and software) in addition to research publications, and consider a broad range of impact measures including qualitative indicators of research impact, such as influence on policy and practice.

For institutions

4. Be explicit about the criteria used to reach hiring, tenure, and promotion decisions, clearly highlighting, especially for early-stage investigators, that the scientific content of a paper is much more important than publication metrics or the identity of the journal in which it was published.

5. For the purposes of research assessment, consider the value and impact of all research outputs (including datasets and software) in addition to research

Box 9.2 (continued)

publications, and consider a broad range of impact measures including qualitative indicators of research impact, such as influence on policy and practice.

For publishers

6. Greatly reduce emphasis on the journal impact factor as a promotional tool, ideally by ceasing to promote the impact factor or by presenting the metric in the context of a variety of journal-based metrics (e.g., 5-year impact factor, EigenFactor [8], SCImago [9], h-index, editorial and publication times, etc.) that provide a richer view of journal performance.

7. Make available a range of article-level metrics to encourage a shift toward assessment based on the scientific content of an article rather than publication metrics of the journal in which it was published.

8. Encourage responsible authorship practices and the provision of information about the specific contributions of each author.

9. Whether a journal is open-access or subscription-based, remove all reuse limitations on reference lists in research articles and make them available under the Creative Commons Public Domain Dedication [10].

10. Remove or reduce the constraints on the number of references in research articles, and, where appropriate, mandate the citation of primary literature in favor of reviews in order to give credit to the group(s) who first reported a finding.

For organizations that supply metrics

11. Be open and transparent by providing data and methods used to calculate all metrics.

12. Provide the data under a license that allows unrestricted reuse, and provide computational access to data, where possible.

13. Be clear that inappropriate manipulation of metrics will not be tolerated; be explicit about what constitutes inappropriate manipulation and what measures will be taken to combat this.

14. Account for the variation in article types (e.g., reviews versus research articles), and in different subject areas when metrics are used, aggregated, or compared.

For researchers

15. When involved in committees making decisions about funding, hiring, tenure, or promotion, make assessments based on scientific content rather than publication metrics.

16. Wherever appropriate, cite primary literature in which observations are first reported rather than reviews in order to give credit where credit is due.

Box 9.2 (continued)

17. Use a range of article metrics and indicators on personal/supporting state-ments, as evidence of the impact of individual published articles and other research outputs [11].

18. Challenge research assessment practices that rely inappropriately on Journal Impact Factors and promote and teach best practice that focuses on the value and influence of specific research outputs.

References

1. Adler, R., Ewing, J., & Taylor, P. (2008). *Citation statistics*. A report from the Inter-national Mathematical Union. Retrieved from http://www.ams.org/notices/200808 /tx080800968p.pdf

2. Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluat-ing research. *BMJ: British Medical Journal*, *314*, 498–502.

3. Editorial. (2005). Not so deep impact. *Nature*, *435*, 1003–1004.

4. Vanclay, J. K. (2012). Impact factor: Outdated artefact or stepping-stone to journal certification? *Scientometrics*, *92*, 211–238.

5. The PLoS Medicine Editors. (2006). The impact factor game. *PLoS Medicine*, *3*(6): e291. doi:10.1371/journal.pmed.0030291

6. Rossner, M., Van Epps, H., & Hill, E. (2007). Show me the data. *Journal of Cell Biology*, *179*, 1091–1092.

7. Rossner, M., Van Epps, H., & Hill, E. (2008). Irreproducible results: A response to Thomson Scientific. *Journal of Cell Biology*, *180*, 254–255.

8. http://www.eigenfactor.org/

9. http://www.scimagojr.com/

10. http://opencitations.wordpress.com/2013/01/03/open-letter-to-publishers

11. http://altmetrics.org/tools/

in mathematics and for more junior female instructors (Mengel, Sauer-mann, & Zölitz, 2017). These studies suggest that student evaluations of teaching—though often an important part of faculty evaluation--are a very flawed source of evidence.

Reliance on Evidence

It is usually optimal to use multiple forms of evidence for any criterion (thus, not to rely on a single indicator of research or teaching excellence). In the case of teaching, one would consider, for example, the sophistication and creativity of syllabuses and assignments, peer evaluations of teaching as well as student evaluations, the mix of courses offered and evidence of

success at different levels (e.g., first- and second-year students vs. upper level students) and in different kinds of courses (e.g., lecture vs. seminar format), attention to individual mentoring of students at various levels, and so on. The particular standard employed should be articulated within fields and units because the nature of the teaching, research, and service demands differs by field. For example, how introductory courses and laboratory or writing-intensive courses are understood and assessed is field specific.

It is worth noting again that evidence of minimal competence tends to be evaluated relatively leniently for "outgroup members" (i.e., White women and racial-ethnic minorities) while evidence of meeting a high standard is likely to need to be higher for those same groups (see Biernat, 2012; Biernat & Kobrynowicz, 1997; Biernat, Ma, & Nario-Redmond, 2008). Moreover, any sign of incompetence or failure on a stereotype-relevant dimension (e.g., female performance in a male-dominated field, minorities' performance in academic settings) is likely to be noticed and weighted heavily (Biernat, Fuegen, & Kobrynowicz, 2010).

Partly because of the subtlety and complexity of these cognitive processes, time pressure militates against fairness. Because the cognitive demands are greater in evaluations of "outgroup" members, time pressure increases reliance on stereotypes or group-based schemas rather than on individual performance (see, e.g., Biernat, Kobrynowicz, & Weber, 2003, in which even judgments of men's and women's heights were influenced by time pressure). Many occasions on which faculty evaluate their junior peers are accompanied by powerful time pressures. Only by minimizing these pressures can one reduce the probability of reliance on group-based stereotypes.

External Letters

Many—perhaps most—institutions rely heavily on external letters from field or subfield experts in the course of tenure and promotion reviews, though they are sought using a variety of procedures (e.g., allowing the candidate to nominate some but not all of the reviewers, ruling out reviewers who are not "arm's length," or alternatively requiring letters from collaborators and/or dissertation and postdoctoral advisors). In addition, some institutions seek a small number of letters, while others seek many (we are aware of a range from at least 3 to 20). Some choices are driven by a desire for objectivity, others by a desire to ensure that the candidate's work is fully understood.

Reliance on external letters from field or disciplinary experts may seem both wise and inevitable—perhaps especially when internal reviewers feel they are not knowledgeable enough about the candidate's area of research to carry out an "expert" assessment. These issues may also feel particularly complicated when individuals work at the intersections of multiple fields or in truly interdisciplinary fields. We agree that there may be special challenges in such cases and reviewers—internal and external—must be selected particularly thoughtfully in those cases.

But we also note that the tenure and promotion decision is always made in part by individuals who are not experts in the field or subfield of the candidate. How do such individuals weigh the accomplishments of an individual in a field they do not know or understand? To some extent they rely on the judgments of those in the unit who are closest in expertise to the candidate (just as they did at the time of hiring), though, of course, the degree of closeness in expertise varies in different cases. However, to a much higher degree at the time of tenure and promotion review, institutions depend on the judgments of individuals outside the institution who are demonstrably (or allegedly!) expert in the same or a related field to assess the scholarly contributions of the candidate.

Despite those good reasons for seeking external opinions, overreliance on letters is an unwise practice. Given the lack of control over the education of external letter writers and their lack of accountability, we are tempted to recommend that departments not request external letters. We understand that it is unlikely that this recommendation will be adopted soon—or perhaps ever. For that reason, we will address in detail the reasons for our skepticism. Our hope is that we will persuade some readers of letters at least to adopt a high degree of skepticism about the weight they should have in the review process. In addition, we strongly recommend adoption of a procedure requiring that at the earliest stages of the review process, individuals within the field at the institution—however incomplete their knowledge of the area of research—should first form their own judgment based on the same material provided to external reviewers, and without any reference to the external reviewers' comments. This procedure will make it more likely that any comments that are inconsistent with the views of the internal reviewers are then subjected to the higher level of scrutiny we believe they merit.

Our reservations about the value of letters in the review process arise first from the fact that these letters are prepared by individuals with unknown

education about the cognitive biases outlined in this volume and with unknown capacity for fair-minded assessment. If an institution has—as we hope it has—adopted a set of procedures that makes it likely that *internal* reviewers meet a high standard on these two dimensions, why would it lower the overall quality of the review—to an unknown degree—by adding in material that is generated by individuals who do not meet that standard?

We would not be concerned if there were little or no evidence that external letters are influenced by group-based stereotypes. Unfortunately, the evidence is substantial that they are (Dutt, Pfaff, Bernstein, Dillard, & Block, 2016; Schmader, Whitehead, & Wysocki, 2007; Trix & Psenka, 2003). We reviewed much of this evidence in the context of hiring at the entry level in chapter 6 but it is important specifically to consider the implications for individuals with a longer record of accomplishment, such as tenure candidates. In one key study (Steinpreis, Anders, & Ritzke, 1999), the process of review by external letter writers was mimicked. Thus, raters were sent one CV for an applicant for a job or for early tenure (parallel to material sent to external reviewers). Gender was varied and assigned in advance (by providing a name to the applicant) to records that were otherwise identical. This experiment was conducted with two sets of records, one for a dossier at the entry level and one for a dossier at the point of review for "early tenure."

In the first case, as we might expect, the man's accomplishments—though identical to the woman's—were more highly rated, presumably because of the kinds of schema-related language differences in the letters spelled out in chapter 6. In the second case, of an unusually strong record compatible with review for early tenure, the judges' recommendation for tenure was the same for the man and the woman. However, these recommendations were accompanied by expressions of reservations by raters *four times more often* in the case of the dossier with a female name—perhaps because of the schemas of the judges.

Even when coming to a favorable decision about the female tenure candidate, external reviewers who reviewed her record were especially likely to refer in their comments not to her accomplishments, but to concerns about their source. This finding is consistent with other evidence that, even when associated with positive overall ratings, judgments of women and minorities are influenced by negative stereotype-consistent attributions and memory distortions (see Lord & Taylor, 2009). The reservations expressed by the external reviewers (if summarized in a report) might influence other levels

of review, despite the presence of their positive overall judgment. In addition, the expressed reservations may influence internal committee members' own interactions with their colleague, creating a work environment in which in which she must "prove herself" more than her comparably accomplished male counterpart.

In addition to evidence that stereotypes are expressed in external letters, there are other reasons to believe that the letters add little of value to tenure and promotion reviews. External letters are often favorable and therefore provide little differentiation among candidates (Chance, 2012; see also Schneider, 2000). Chance points out that he "will spend much time trying to read what is said and digging deeply for what is not said" (p.3); he recommends that others also do so. We believe that this process of intensive inference about small cues in external letters is another source of unknown error and bias in the review process. Suggestions from review committee members about negative characterizations that may be hinted at or implied by silences in external letters are threats to the fairness of the review process.

The restricted range of opinion expressed in external letters may also limit their value (Thorngate et al., 2009): "With no variability, there can be no predictive validity" (p. 82). Peer reviewers of grant proposals offer views that are only moderately correlated, and disagreement between reviewers is significantly associated with lower judgments of merit regardless of the content of the arguments (Thorngate et al., 2009). Negative inputs are weighted more heavily than positive ones, even though everything we know about cognitive biases suggests that strong inferences from weak cues are not likely to be accurate (Lord & Taylor, 2009).

The Role of Committee Discussion

Group discussions do not automatically ensure fairness in judgment processes even if including more judges is valuable. Discussions in committees making decisions about recommendations to fund or not fund grant proposals (which are, of course, not tenure reviews, but provide an imperfect analogy) do not move members toward greater consensus; the correlation between the average rating given by pairs of primary assessors in advance of discussion and the average rating following discussion was extremely high in the two instances studied: 0.82 and 0.92. Thus, the discussion moved

judgments very little. Relying on averaged initial ratings without committee deliberation would have been equally fair and saved a lot of time (Thorngate et al., 2009; see also Pier, Raclaw, Nathan, Kaatz, Carnes, & Ford, 2015, re findings about inconsistency in study section review ratings).

We do not believe that faculty confidence—rightly or wrongly—would be high in a review involving no discussion and collective deliberation, but the results of Thorngate et al.'s research do point to the importance of procedural care in the deliberative process. They recommend, as we have, that planning and preparation establish "clear and concise eligibility rules, merit criteria and [in the case of grant applications] budgeting regulations before an adjudication" (p. 85). They note that committees will generate procedural rules and merit criteria "on the fly" in the absence of prior discussion. Thus, they advocate, as we do, "clearly stated rules" and "clear examples" (p.85), with ongoing feedback from both applicants and evaluators used to improve the clarity of both the stated rules and the examples.

Articulating Best Practices in Letter-Writing Requests

If committees *are* going to request outside letters, we recommend a request along the lines of the template in box 9.3. We note several features about what is and is not being requested in this template. The template acknowledges that people can adopt different approaches to a subject and requests that writers evaluate the candidate independently of whether they agree with the candidate's approach. Although we are somewhat skeptical about reviewers' ability to fully accomplish this, the letter encourages them to try. The letter also specifically asks the reviewers to assess the quality of the person's work and to do so independently of time since degree.

The letter tells reviewers the institution is not asking for a recommendation about tenure or promotion. That is the job of the review committee, based on a full consideration of all of the candidate's qualifications, one of which will be estimates of the quality of the candidate's work. The letter also does not ask reviewers to compare their candidate with any specific people. (Some schools do this: they provide a list of prominent people and ask the reviewer to compare the candidate to those people at comparable stages in their careers, or they ask the reviewer to compare the candidate with their own chosen set of people. This is an invitation to reviewers to let their subjective preferences run wild, with little or no accountability.)

Box 9.3
Template of Request for Outside Letters

Dear Prof. X:

Our department has been reviewing the work of [Dr. ABC]. We make a recommendation to [a divisional meeting] in [month & year] regarding [his or her] [tenure] [promotion to associate professor] [promotion to full professor] [some combination of the above].

I am writing to ask if you would be willing to assist us by providing an evaluation of [Dr. ABC's] scholarly work. We seek impartial evaluations from people such as you, who can comment with knowledge and authority on the scholarly competence and significance of [his or her] research, and who can place [his or her] work in the context of [his or her] particular field.

We are not asking you whether we should or should not recommend [Dr. ABC] for [tenure] [promotion]. Rather, we want you to assess [Dr. ABC's] scholarly achievements and potential and tell us what you see as [his or her] contribution to the field. We are particularly interested in your comments on the quality, rather than the quantity, of [Dr. ABC's] written work and presentations. Assessment of particular articles or chapters in [Dr. ABC's] book, evidence that Dr. ABC brings to bear [if appropriate], and theoretical arguments will be particularly valuable to us.

Finally, please note that at our institution we assess people's suitability for [tenure or promotion] without regard to time in rank or time since degree; please assume that the individual is coming up for review for tenure and promotion at the right time for our institution and refrain from any comment about time.

We recognize that within a field there may be different approaches and controversies; as far as possible, we would like you to address [Dr. ABC's] contributions independent of your own agreement or disagreement with [his or her] approach. We would welcome your comments on all of [Dr. ABC's] work, but you may focus on particular aspects of it.

Our general criteria for evaluating faculty members include their current research and potential for future research, their teaching, and their service to their department and the college. With respect to research, we take into account the fact that [X University is a very high research activity university] [or X College is a liberal arts college where faculty teach as many as three courses a semester at the undergraduate and graduate level, as well as fulfill committee responsibilities] [or some variant]. We do not expect an evaluation of [Dr. ABC's] teaching from you, though [his or her] effectiveness in giving presentations may be something you are familiar with and may wish to comment on.

[Insert here a discussion of the confidentiality and legal status of letters of this kind at your institution and in your state. It is important to provide clear information about who within the institution will have access to the letter, how it will figure in deliberations, and whether the content of the letter and/or its author's identity will or may be divulged to the candidate, and under what circumstances.]

I have enclosed a copy of [Dr. ABC's] curriculum vitae. I do hope you will be able to help us with this important decision. If so, I will send you a more complete dossier, including copies of relevant publications, in electronic or paper format or both, as well as statements by the candidate. Since the Department must review the comments of outside reviewers and submit its recommendation by [date], it would be most helpful if we could receive your signed letter and a copy of your CV in either electronic or paper form by [date].

Promotion to Full Professor

Most institutions separate the award of tenure from promotion to full professor (the final rank of tenure-track faculty). In recent years evidence has accumulated that promotion to full professor is no more insulated from equity concerns than any other process. Both analyses within fields and within institutions across fields indicate a high level of variability in the timing of promotion to full professor (including that it never happens) between women and men, and by discipline.

Natural scientists spend the shortest time in the associate professor rank, social scientists are next, and those in the humanities spend the longest time in the rank. As we have seen in chapter 4, salary differentials by field follow the same pattern. This is unsurprising if promotion is accompanied (as it often is) by a larger raise in salary compared with the size of annual raises. Given the logic of the accumulation of advantage over time (compound interest), the earlier a raise occurs in a career, the bigger the impact it will have on the final salary an individual is paid. Considering the gender distribution across disciplines (more women in the humanities and fewer in natural science), it is unsurprising that time to promotion is related to gender (Anderson et al., 2009) and may also be related to race and ethnicity (Fang et al., 2000). Additional factors beyond distribution by field probably account for the gender differences since there is also clear evidence of differences in time to promotion within a given discipline, both across and within institutions (Committee on the Status of Women in the Professions, MLA, 2009; Giesler et al., 2007).

Specification of Criteria for Promotion

One factor identified by several analyses as a potential explanation for the disciplinary differences (independent of gender or race and ethnicity) is the lack of specification of criteria for this promotion beyond those for tenure and (ordinarily) promotion to associate professor (see Anderson et al., 2009; Perna, 2001; Tierney & Bensimon, 1996). In the absence of new criteria, many faculty, trying to assess their own promotability, and many of the senior faculty they consult, assume that the amount and kind of productivity is roughly identical—there should just be more of both. This assumption may be reasonable in fields in which article publication is the primary evidence, but in fields that depend on book publication, it is likely to add

time to the period before promotion. The "first book" is normally based not only on scholarship generated during a faculty member's period as an assistant professor, but on the work the faculty member completed for his or her dissertation. The "second book" is unlikely to require less research or integration and must be produced in the context of a full-time job with other responsibilities. Therefore, it is unsurprising that individuals in book fields (the humanities and interpretive social sciences) take longer to accumulate a record for promotion to full professor.

Is it, then, simply inevitable that there must be disciplinary differences in time to promotion? No! Some institutions have adopted a time-based promotion procedure; once faculty members have been in rank some number of years (normally 5–7), their continued scholarly activity and their teaching and service contributions are assessed. If they meet a standard of continued productivity at the level required for adequate performance of a faculty member, they are promoted (see Anderson et al., 2009, pp. 14–15). This approach to the review process requires that "scholarly activity" be measured in terms that may differ from those used at the tenure review. For example, it may be demonstrated in terms of articles that will eventually form part of a monograph, in grant applications, in presentations at national and international meetings, and in partially completed manuscripts.

Recognizing Institutional Demands for Service

Many institutions rely differently on associate professors than on assistant professors for conducting the ongoing work of the institution: taking part in faculty reviews of less experienced faculty, leading committee labor associated with recruitment of faculty and students, curriculum revision, mentoring junior faculty, and so forth. If the job is different, then the weighting of the criteria may need to be adjusted, too, with somewhat greater weight applied to teaching, service, and leadership than at the tenure review (when scholarly accomplishment and promise tend to lead the other criteria and service and leadership are not weighted heavily). As there is considerable evidence that women and minorities carry a disproportionate responsibility for service and leadership, and carry it earlier in their careers (Guarino & Borden, 2017; Misra, Lundquist, Holmes, & Agromavritis, 2011; Turner et al., 2008), a shift to taking all domains into consideration might help eradicate the gender and race and ethnicity differences in time to promotion.

Disproportionate Impact of Ambiguity

Finally, the ambiguity in timing of promotion to full professor is likely to disadvantage those groups that feel the least comfort and acceptance—belonging—within the institution. Feeling like "an outsider within," as women and minority tenured faculty often do, is likely to lead an individual to have less information and less confidence in moving forward with a promotion in the absence of clear guidelines about timing. In some institutions faculty members are not reviewed for promotion to full professor on a timetable; instead, they must "apply" for promotion. Given what we know about the biases in self-evaluation that affect judgments of individuals' own performance, this process is likely to lead to inequities. One way to combat this bias is to create a routinized procedure requiring all individuals at a normative time after tenure (e.g., 3 or 4 years) to be evaluated by their department in terms of the areas that require additional performance in order for promotion to be considered. Individuals can then be encouraged to develop plans for meeting those criteria on a timeline and can review their projects with a mentor or the department chair on an annual basis. In contrast, when individuals seek advice about whether they are "ready" for promotion, that advice is subject to advisors' unexamined schemas and may lead to unnecessary and counterproductive delays. Adoption of this kind of routinized review process would reduce both disciplinary and demographic differences in time to promotion—and therefore in salary.

Similarly, differences would be reduced or eliminated if more institutions accepted more flexible evidence of continued scholarly activity and impact, engaged in more equal weighting of all elements of a tenured faculty member's job, and provided clear and explicit procedures and timing for the promotion review. Some institutions view promotion to full professor as a rare privilege, to be accorded only to the "special few"; they have to live with the resulting demoralization of those faculty not promoted. Moreover, those faculty can become poor role models for junior faculty in their fields, as they withdraw from meaningful participation in the community and may decrease their productivity. Our perspective is that the institution benefits when the individual benefits—and that entails a system in which the assumption is that all faculty can be promoted to full professor, and support is provided to all tenured faculty, enabling that achievement.

Posttenure Reviews

Because many faculty remain full professors for a long period—often more than half of an academic career—some institutions believe that a substantial posttenure review process is helpful in maintaining faculty engagement and accountability. This seems potentially true, particularly in the absence of an effective and trusted annual review process. However, it is often viewed by faculty as imposing an unproductive burden of labor for little benefit. We know of no research demonstrating the value of posttenure reviews in maintaining faculty productivity, though some have argued that the University of California "step system" (which imposes such reviews throughout the academic career) has done just that and also minimizes salary inequities. (Because of the powerful impact of the "counteroffer" culture in academia, with large raises only tied to external offers, we are dubious about the latter claim.)

In any case, we suspect that an effective annual review system that relaxed the timing of the formal reviews of full professors (so that reporting occurred annually but formal reviews less often) might accomplish the goals of a formal posttenure review. In addition, it seems to us most important that faculty throughout their careers have the opportunity to learn which of their efforts are most highly valued by their near colleagues and to benefit from advice at choice points in their careers. Thus, it is the communicative function of reviews that is most likely to be valuable to faculty members, which requires that great care be taken in both the design and the implementation of procedures for these and all other reviews.

Summary

Faculty review and promotion processes are subject to some of the same sorts of evaluation biases as is faculty hiring—and there are some additional pitfalls. Further, the problem of actually lowering faculty productivity by lowering morale is a risk that must be faced. Many of the risks of unfair review are heightened in the instance of review of faculty who bring diversity to departments and institutions. Both individual faculty and institutions are served best by review processes that mitigate the risks of bias, are transparent, and are conducted according to explicit criteria and on a normative but flexible timetable.

Recommendations for Enhancing the Fairness of Evaluation and Promotion Procedures

Senior Administrative Leaders

1. Ensure that there is institution-wide effort to articulate review criteria clearly for every form of faculty review.

2. Establish the norm that review of one's colleagues is a privilege requiring evidence of conscientiousness and fairness, as well as openness to education about how bias influences such reviews and can be mitigated.

3. Develop an educational program for faculty participating in reviews that allows them to reflect on unintended bias as well as on procedures for mitigating it.

4. Monitor review processes for evidence of inequity by discipline and demographic group. Hold units accountable for fair outcomes.

Department Chairs and Deans of Small Schools

1. A precondition for perceived fairness is that criteria and procedures are spelled out adequately and fairly. Make transparency a priority, and post information about both criteria and procedures prominently.

2. The choice of fair and unbiased reviewers—internally and externally—is a critical factor in the perception that a process has been fair to the candidate and the institution. Make these selections carefully.

3. The burden of clear, productive, and respectful communication is substantial. Develop a process for conducting reviews internally that maximizes your success at this part of the job. This may involve sharing it with a senior faculty member appointed to oversee these reviews or including senior faculty mentors in feedback processes. It should also involve seeking and responding to feedback about how well administrators are doing this aspect of the job.

Faculty Evaluators

1. Become familiar with the literature on how cognitive biases influence judgments of others. Ensure that colleague-reviewers do so as well.

2. Take responsibility for fair deliberation within the review process. If some people are adopting or using arbitrary or different standards, or if they are referring to stereotypes rather than evidence, colleagues should hold those reviewers accountable to a higher standard.

3. Try to avoid making decisions in haste. Allocate the time needed to arrive at a fair opinion. Encourage the committee to do the same in its group process.

Faculty Being Reviewed

1. Keep good records of all work-related activity in a format that is compatible with the format(s) required by the review processes in the institution. Good record keeping will reduce the burden of labor when documents are prepared for the review and will minimize the likelihood of forgotten names (of committees, students, etc.) or lack of mention of crucial, time-consuming activities. Consult with senior and just-promoted colleagues about what records are important to keep.

2. Take responsibility for reviewing timetables, guidelines, and review criteria carefully and well in advance. Ask questions about things that are unclear, and encourage revision of materials for future reviews if things were unclear in this one.

3. Make use of any relevant material that illustrates the desiderata of materials submitted for review. This may include examples of previously submitted material shared widely or requests to individual colleagues.

4. Prepare materials carefully and well in advance. Ask trusted, more senior colleagues for feedback on critical elements that may not be familiar forms of writing to some committee members (e.g., statements of research activity, teaching goals, etc.). Build sufficient time into the timeline to use that feedback to revise the final submission.

5. Always bear in mind that key faculty and administrators who review these documents will not be in the field or subfield and will need clear exposition rather than reliance on disciplinary understanding of specialized language. Ask someone outside the subfield, or even the field, to flag instances of reliance on disciplinary codes in the documents.

Note

1. Nothing prevents a reviewer from visiting a faculty member's website to see what other papers are listed and to read as many as he or she might like. But we imagine most reviewers would be delighted to have only a faculty member's top five publications to read.

References

Anderson, E., Collins, C., Fierke, C., Juster, S., Kimeldorf, H., Penner-Hahn, J., et al. (2009, June 15). *Report of the Associate Professor Rank Committee*. Ann Arbor, MI: College of Literature, Sciences and the Arts, University of Michigan. Unpublished report.

Ash, A. S., Carr, P. L., Goldstein, R., & Friedman, R. H. (2004). Compensation and advancement of women in academic medicine: Is there equity? *Annals of Internal Medicine*, *141*, 205–212.

Barbezat, D. (2002). History of pay equity studies. In R. Toutkoushian (Ed.), *Conducting salary-equity studies: Alternative approaches to research* (pp. 9–40). *New Directions for Institutional Research, 115*. San Francisco, CA: Jossey-Bass.

Bavishi, A., Madera, J. M., & Hebl, M. R. (2010). The effect of professor ethnicity and gender on student evaluations: Judged before met. *Journal of Diversity in Higher Education*, *3*(4), 245–256.

Biernat, M. M. (2012). Stereotypes and shifting standards: Forming, communicating and translating person impressions. *Advances in Experimental Social Psychology*, *45*. doi:10.1016/B978-0-12-394286-9.00001-9

Biernat, M., Fuegen, K., & Kobrynowicz, D. (2010). Shifting standards and the inference of incompetence: Effects of formal and informal evaluation tools. *Personality and Social Psychology Bulletin*, *36*, 855–868.

Biernat, M., & Kobrynowicz, D. (1997). Gender- and race-based standards of competence: Lower minimum standards but higher ability standards for devalued groups. *Journal of Personality and Social Psychology*, *72*, 544–557.

Biernat, M., Kobrynowicz, D., & Weber, D. (2003). Stereotyping and shifting standards: Some paradoxical effects of cognitive load. *Journal of Applied Social Psychology*, *33*, 2060–2079.

Biernat, M., Ma, J. E., & Nario-Redmond, M. R. (2008). Standards to suspect and diagnose stereotypical traits. *Social Cognition*, *26*, 288–313.

Boring, A., Ottoboni, K., & Stark, P. B. (2016). *Student evaluations of teaching (mostly) do not measure teaching effectiveness*. ScienceOpen Research; 10.14293/S2199-1006.1. SOR-EDU.AETBZC.v1.

Castilla, E. J. (2008). Gender, race, and meritocracy in organizational careers. *American Journal of Sociology*, *113*(6), 1479–1526.

Chabram, A. (2016). Tenure in the contact zone: Spanish is our language, too. In P. Matthew (Ed.), *Written/unwritten: Diversity and hidden truths of tenure* (pp. 95–108). Chapel Hill, NC: University of North Carolina Press.

Chance, D. M. (2012, November 14). The idiocy of promotion-and-tenure letters. *Chronicle of Higher Education* http://www.chronicle.com/article/The-Idiocy-of/135740/

Cohen, G. L., Steele, C. M., & Ross, L. D. (1999). The mentor's dilemma: Providing critical feedback across the racial divide. *Personality and Social Psychology Bulletin*, *25*, 1302–1318.

Committee on the Status of Women in the Profession, MLA. (2009, April 27). *Standing still: The associate professor survey*. Modern Language Association of America. Retrieved from https://www.mla.org/About-Us/Governance/Committees/Committee -Listings/Professional-Issues/Committee-on-the-Status-of-Women-in-the-Profes sion/Standing-Still-The-Associate-Professor-Survey/

Crocker, J., Voelkl, K., Testa, M., & Major, B. (1991). Social stigma: The affective consequences of attributional ambiguity. *Journal of Personality and Social Psychology*, *60*, 218–228.

Curry, T. H. (2006, February). Faculty performance reviews. *Effective Practices for Academic Leaders*, *1*(2). Sterling, VA: Stylus.

Dutt, K., Pfaff, D. L., Bernstein, A. F., Dillard, J. S., & Block, C. J. (2016). Gender differences in recommendation letters for postdoctoral fellowships in geoscience. *Nature Geoscience*, *9*, 805–809.

Fang, D., Moy, E., Colburn, L., & Hurley, J. (2000). Racial and ethnic disparities in faculty promotion in academic medicine. *Journal of the American Medical Association*, *284*(9), 1085–1092.

Fiske, S. T. (2011). *Envy up, scorn down: How status divides us*. New York, NY: Russell Sage.

Fiske, S. T., Xu, J., Cuddy, A. C., & Glick, P. (1999). (Dis) respecting and (dis)liking: Status and interdependence predict ambivalent stereotypes of competence and warmth. *Journal of Social Issues*, *55*(3), 473–489.

Giesler, C., Kaminski, D., & Berkley, R. (2007). The 13+ club: An index of non-promotion of full professors. *NWSA Journal*, *19*(3), 145–162.

Graves, A., Hoshino-Browne, E., & Lui, K. P. H. (2017). Swimming against the tide: Gender bias in the physics classroom. *Journal of Women and Minorities in Science and Engineering*, *23*(1), 15–36.

Greenhaus, J. H., & Parasuraman, S. (1993). Job performance attributions and career advancement prospects: An examination of gender and race effects. *Organizational Behavior and Human Decision Processes*, *55*, 273–297.

Guarino, C. M., & Borden, V. M. H. (2017). Faculty service loads and gender: Are women taking care of the academic family? *Research in Higher Education*. doi:10.1007/s11162-017-9454-2.

Hammond, P. T., Bailyn, L., Brown, E., Harris, L., Norford, L., Ortiz, C., et al. (2010). *Report on the initiative for faculty race and diversity*. MIT. Retrieved from http://orgchart .mit.edu/sites/default/files/reports/20100114_Provost_InitFacultyRaceDiversity Report.pdf

Harrison, R. L. (2016). Building a canon, creating dialogue: An interview with Cheryl A. Wall. In P. Matthew (Ed.), *Written/unwritten: Diversity and hidden truths of tenure* (pp. 46–63). Chapel Hill, NC: University of North Carolina Press.

Heilman, M. E., & Haynes, M. C. (2005). No credit where credit is due: Attributional rationalization of women's success in male-female teams. *Journal of Applied Psychology*, *90*(5), 905–916.

Heilman, M. E., & Okimoto, T. G. (2007). Why are women penalized for success at male tasks? The implied communality deficit. *Journal of Applied Psychology*, *92*(1), 81–92.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(46), 16569–16572.

Jagsi, R., Griffith, K. A., Stewart, A. J., Sambuco, D., DeCastro, R., & Ubel, P. A. (2012). Gender differences in the salaries of physician researchers. *Journal of the American Medical Association*, *307*(22), 2410–2417.

King, E. B., Madera, J. M., Hebl, M. R., Knight, J. L., & Mendoza, S. A. (2006). What's in a name? A multiracial investigation of the role of occupational stereotypes in selection decisions. *Journal of Applied Social Psychology*, *36*(5), 1145–1159.

Landy, F. J. (2008). Stereotypes, bias, and personnel decisions: Strange and stranger. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *1*, 379–392.

Livingston, R. W., Rosette, A. S., & Washington, E. F. (2012). Can an agentic Black woman get ahead? The impact of race and interpersonal dominance on perceptions of female leaders. *Psychological Science*, *23*(4), 354–358.

Lord, C. G., & Taylor, C. A. (2009). Biased assimilation: Effects of assumptions and expectations on the interpretation of new evidence. *Social and Personality Psychology Compass*, *3*(5), 827–841.

Maynard, D. C., & Brooks, M. E. (2008). The persistence of stereotypes in the context of familiarity. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *1*, 417–419.

Meho, L. I., & Yang, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of Science vs. Scopus and Google Scholar. *Journal of the American Society for Information Science and Technology*, *58*(13), 2105–2125.

Mengel, F., Sauermann, J., & Zölitz, U. (2017). Gender bias in teaching evaluations. IZA Institute of Labor Economics Discussion Paper Series 11000. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3037907/

Misra, J., Lundquist, J. H., Holmes, E., & Agromavritis, S. (2011). The ivory ceiling of service work. *Academe*, *97*(1), 22–26.

National Center for Educational Statistics. (2016). Table 316.80. Percentage of degree-granting postsecondary institutions with a tenure system and of full-time faculty with tenure at these institutions, by control and level of institution and selected characteristics of faculty: Selected years, 1993–94 through 2015–16. https://nces.ed.gov/programs/digest/d16/tables/dt16_316.80.asp?current=yes/

Perna, L. W. (2001). Sex and race differences in faculty tenure and promotion. *Research in Higher Education*, *66*(5), 541–567.

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

Phelan, J. E., & Rudman, L. A. (2010). Prejudice toward female leaders: Backlash effects and women's impression management dilemma. *Social and Personality Psychology Compass*, *4/10*, 807–820.

Pier, E. L., Raclaw, J., Nathan, M. J., Kaatz, A., Carnes, M., & Ford, C. E. (2015). Studying the study section: How group decision making in person and via videoconferencing affects the grant peer review process (WCER Working Paper 2015–6). Retrieved from University of Wisconsin—Madison, Wisconsin Center for Education Research website: http://www.wcer.wisc.edu/publications/workingPapers/papers.php

Pollack, M. (2015, April 2). E-mail re: Faculty evaluation: Recognition of entrepreneurial, creative, and outreach activities. Retrieved from http://advance.umich.edu/resources/FASTERThirdYearTenurePromotionGuidelines.pdf

Porter, S., Toutkoushian, R. K., & Moore, J. (2008). Pay inequities for recently hired faculty, 1988–2004. *Review of Higher Education*, *34*, 465–487.

Purdie-Vaughns, V., & Eibach, R. P. (2008). Intersectional invisibility: The distinctive advantages and disadvantages of multiple subordination group identities. *Sex Roles*, *59*(5–6), 377–391.

Roberts, T. (1991). Gender and the influence of evaluations and self-assessments in achievement settings. *Psychological Bulletin*, *109*, 297–308.

Roberts, T., & Nolen-Hoeksema, S. (1994). Gender comparisons in responsiveness to others' evaluation in achievement settings. *Psychology of Women Quarterly*, *18*, 221–240.

Rudman, L. A., & Glick, P. (1999). Feminized management and backlash toward agentic women: The hidden costs to women of a kinder, gentler image of middle-managers. *Journal of Personality and Social Psychology*, *77*, 1004–1010.

Ruscio, J., Seaman, F., D'Oriano, C., Stremlo, E., & Mahalchik, K. (2012). Measuring scholarly impact using modern citation-based indices. *Measurement: Interdisciplinary Research and Perspectives*, *10*(3), 123–146.

Sarsons, H. (2017). Recognition for group work: Gender differences in academia. *American Economic Review*, *107*(5), 141–145.

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.

Schneider, A. (2000). Why you can't trust letters of recommendation. *Chronicle of Higher Education*, *46*(43), A14–A16.

Sesko, A. K., & Biernat, M. (2010). Prototypes of race and gender: The invisibility of Black women. *Journal of Experimental Social Psychology*, *46*(2), 356–360.

Sohr-Preston, S. L., Boswell, S. S., McCaleb, K., & Robertson, D. (2016). Professor gender, age, and "hotness" in influencing college students' generation and interpretation of professor ratings. *Higher Learning Research Communications*, *6*(3), 3–25.

Stanley, C. A. (2006). Coloring the academic landscape: Faculty of color breaking the silence in predominantly White colleges and universities. *American Educational Research Journal*, *43*(4), 701–736.

Stanley, C. A., Porter, M. E., Simpson, N. J., & Ouellett, M. L. (2003). A case study of the teaching experiences of African American faculty at two predominantly White research universities. *Journal on Excellence in College Teaching*, *14*, 151–178.

Steinpreis, R. E., Anders, K. A., & Ritzke, D. (1999). The impact of gender on the review of the curricula vitae of job applicants and tenure candidates: A national empirical study. *Sex Roles*, *41*(7/8), 509–528.

Stewart, K. D., Dalton, M. M., Dino, G. A., & Wilkinson, S. P. (1996). The development of salary goal modeling: From regression analyses to a value-based prescriptive approach. *Journal of Higher Education*, *67*(5), 555–576.

Storage, D., Horne, Z., Cimpian, A., & Leslie, S.-J. (2016). The frequency of "brilliant" and "genius" in teaching evaluations predicts the representation of women and African Americans across fields. *PLoS One*. doi:10.1371/journal.pone.0150194

Syracuse University. (2005, November 20). *Faculty manual*. Section 2.34 on Tenure. Retrieved from http://www.syr.edu/academics/office_of_academic_admin/faculty/manual/pdf/Section2FM.pdf

Thorngate, W., Dawes, R. M., & Foddy, M. (2009). *Judging merit*. New York, NY: Psychology Press.

Tierney, W. G., & Bensimon, E. M. (1996). *Promotion and tenure: Community and socialization in academe*. Albany, NY: State University of New York Press.

Toutkoushian, R. K., Bellas, M., & Moore, J. (2007). The interaction effects of gender, race, and marital status on faculty salaries. *Journal of Higher Education*, *78*, 572–601.

Toutkoushian, R., & Conley, V. (2005). Progress for women in academe, yet inequities persist. Evidence from NSOPF:99. *Research in Higher Education*, *46*(1), 1–28.

Trix, F., & Psenka, C. (2003). Exploring the color of glass: Letters of recommendation for male and female medical faculty. *Discourse & Society*, *14*(2), 191–220.

Trower, C. A. (2002). What is current policy? In R. P. Chait (Ed.), *The questions of tenure* (pp. 32–68). Cambridge, MA: Harvard University Press.

Turner, C. S. V., Gonzalez, J. C., & Wood, L. (2008). Faculty of color in academe: What 20 years of literature tells us. *Journal of Diversity in Higher Education*, *1*(3), 139–168.

University of California. (2005). *Appointment and promotion*. Academic Personnel Manual, ADM-210. Retrieved from http://senate.universityofcalifornia.edu/_files /committees/ucaade/apm245.pdf

Valian, V. (2000). *Why so slow? The advancement of women*. Cambridge, MA: MIT Press.

Vazire, S., & Gosling, S. D. (2004). E-perceptions: Personality impressions based on personal websites. *Journal of Personality and Social Psychology*, *87*(1), 123–132.

Way, S. F., Morgan, A. C., Clauset, A., & Larremore, D. B. (2017). The misleading narrative of the canonical faculty productivity trajectory. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(44), E9216–E9223. Epub 2017 Oct. doi:10.1073/pnas.1702121114.

Xie, Y., & Shauman, K. (2003). *Women and science: Career processes and outcomes*. Cambridge, MA: MIT Press.