Social Biases and Solutions for Procedural Objectivity

CAROLE J. LEE AND CHRISTIAN D. SCHUNN

An empirically sensitive formulation of the norms of transformative criticism must recognize that even public and shared standards of evaluation can be implemented in ways that unintentionally perpetuate and reproduce forms of social bias that are epistemically detrimental. Helen Longino's theory can explain and redress such social bias by treating peer evaluations as hypotheses based on data and by requiring a kind of perspectival diversity that bears, not on the content of the community's knowledge claims, but on the beliefs and norms of the culture of the knowledge community itself. To illustrate how socializing cognition can bias evaluations, we focus on peer-review practices, with some discussion of peer-review practices in philosophy. Data include responses to surveys by editors from general philosophy journals, as well as analyses of reviews and editorial decisions for the 2007 Cognitive Science Society Conference.

Proponents of feminist science have argued that theories are underdetermined by the evidence in the sense that there is nothing intrinsic to the data or its relationship to a hypothesis that establishes an evidential relationship between them: there remains a gap between statements describing data and hypotheses. To bridge this gap, background assumptions are needed to identify the evidential connections between them. Since there may be no observational or experimental data available, either in fact or in principle, to provide reasons for preferring one set of background beliefs over another (Longino 1990; Nelson 1990), theorists with different background assumptions can come to support different hypotheses despite a shared domain of data.

Relativizing hypothesis acceptance to individual theorists' background assumptions challenges the impartiality of science—that is, the ability for any observer to recognize evidence as evidence and to see the bearing of evidence on theory in the same way (Lacey 1999). For those who seek to avoid a

radically individualist epistemology, the puzzle becomes one of articulating other tools internal to scientific practice that can protect against illegitimate forms of bias in the evaluation of scientific hypotheses.

To address this problem, Helen Longino argues for a form of social empiricism, in which a scientific community, organized and structured in appropriate ways, can make use of the diversity of members' perspectives to identify and critique biasing background assumptions. Background assumptions "bias" hypothesis evaluation in the sense that they modulate how an observer sees and weighs the evidence bearing on it. Bias, under this account, can be negative when idiosyncratic background assumptions are left unchecked; it can be productive in its ability to open new lines of critique and inquiry, with innovations in methods, concepts, and modes of explanation. Background assumptions are not granted immunity from the demand for reasons and empirical support. Moreover, so long as the community is responsive, criticism of background assumptions can transform scientific theorizing by driving the collection of new data, reliance on new methods, and revision or innovation in argumentation. This approach attributes degrees of objectivity, not to individuals and their particular practices, but to knowledge communities and their procedures.

Longino's conditions for *transformative criticism* articulate ideal standards for procedurally objective communities. *Public venues* for criticism—such as journals and conferences—allow for the evaluation and reevaluation of evidence and background assumptions. *Community uptake* requires that community members be responsive to critical dialogue through the development of new data, reasons, and arguments, as well as the modification of beliefs. *Tempered equality of intellectual authority* requires that attributions of intellectual authority be tempered according to intellectual capacity, education, and other opportunities. This criterion "ensure[s] the exposure of hypotheses to the broadest range of criticism" (Longino 2002, 132), which makes it less likely that accepted hypotheses will result from "idiosyncratic subjective preferences of community members" (Longino 1990, 80).

Finally, *public standards* provide a shared set of principles, values, and aims needed to identify points of agreement and disagreement among community members. As such, they help arbitrate which ideas and critiques get included and responded to, and justify attributions of expertise and competence. Such dialogue is internalist through and through: justifying reasons and the standards of justification themselves are accessible to individual community members and to the community as a whole.

However, an empirically sensitive formulation of the norms of transformative criticism must recognize that even public and shared standards of evaluation can be implemented in ways that unintentionally perpetuate and reproduce forms of social bias that are epistemically detrimental. Such phenomena belong to social psychology's larger oeuvre of studies demonstrating

that the practice of giving reasons does not necessarily provide insight into or protection against the causal influence of social factors in our judgments and decisions more generally (Nisbett and Wilson 1977; Peters and Ceci 1982; Steinpreis, Anders, and Ritzke 1999). If Longino's account is to realize its full potential, it must find ways to address this kind of social-cum-epistemic problem. In this paper, we seek to do just this.

In the first part, we will illustrate how social factors can modulate epistemic evaluation in the context of peer review. We analyze peer review as a sequence of individual judgments and decisions made under uncertainty by authors, editors, and reviewers; briefly articulate normative aims of crucial stages; and foreground how social biases can distort the proper functioning of these stages. We focus on peer-review practices because they serve as gatekeepers in the broad propagation of claims for community-wide consideration, elaboration, and critique. The model is designed to be consistent with peer-review processes in philosophy for two reasons.[1] First, in the vast literature on peer-review practices in the humanities—as well as the social, physical, and human sciences—philosophy has been notably excluded.[2] Second, we focus on philosophy out of a reflexive, practical interest in understanding the ways in which philosophy might be vulnerable to forms of social bias.

Throughout, we interweave empirical generalizations about philosophy's peer-review practices. For our first source, we sent a short survey to editors of twenty-five general philosophy journals.[3] Of these, we received seventeen responses, with responses distributed across the range of prestige.[4] We supplemented these data with information from journal websites about acceptance rates and anonymous review practices. For our second source, two coders analyzed 423 reviews of 221 philosophy and psychology manuscripts submitted for the Cognitive Science Society's annual conference. These analyses allowed us to compare evaluative styles of philosophy and psychology—the dominant discipline in that setting (Schunn, Crowley, and Okada 1998)—in the context of a shared intellectual community. Note that this context is not "interdisciplinary" in the sense used in recent social-scientific research on interdisciplinary peer review that examines how research is reviewed by individuals whose disciplines and evaluative standards differ from those of the author (Boltanski and Thévenot 2006; Lederman 2006; Mallard, Lamont, and Guetzkow 2009). Nor is it interdisciplinary in the sense that it involves the reflexive creation of conventions, protocols, and systems by which different disciplines self-regulate the collective production and evaluation of evidence to address shared research questions (Cambrosio et al. 2006). Instead, our analysis focuses on within-discipline reviewers and action editors in the context of shared interdisciplinary resources (in this case, conference inclusion and publication pages).

In the second part of the paper, we move away from philosophy's particular peer-review practices to discuss how Longino's account can address forms of social bias more generally. Rather than turn the debate into one about how

optimistic or pessimistic we should be about internalist approaches in light of the empirical evidence (Solomon 2006; Tollefsen 2006), we will argue that Longino's theory can explain and redress forms of social bias by taking the following empiricist turn: like scientific hypotheses, evaluative hypotheses about the cognitive merit of others' research are underdetermined by the data. Here, the data include the research under evaluation as well as its relationship to the community's body of knowledge; the hypotheses are evaluative claims about the research under consideration.

On the face of it, evaluative hypotheses are unlike scientific hypotheses in the sense that they are not causal or unifying descriptive claims. However, from an epistemic point of view, the scientific and evaluative cases share important features in common. For both, the evidential relationship between data and hypotheses can be mediated by background beliefs. Just as background beliefs bridge the gap between data and scientific hypotheses, background beliefs can bridge the gap between data and evaluative hypotheses to make salient evidential relationships between them.[5] When these background beliefs are implicit and shared broadly within the community, they become less visible as objects of criticism. Moreover, in cases of social bias, these background assumptions can include cultural beliefs and norms of the community itself.

To address culturally driven forms of unchecked bias, Longino's account has at its disposal a familiar social tool: namely, perspectival diversity. We argue that, in order for diversity to address this problem, the nature of the diversity must bear not on the content of the community's knowledge claims, but on the beliefs and norms of the culture of the knowledge community itself.

Peer Review and Social Bias

Peer review is the standard avenue for criticism and for determining what research and claims get admitted into community-wide spaces such as journals and conferences. According to Longino's account of transformative criticism, the function of peer review is not simply "to check that the data seem right and the conclusions well-reasoned but to bring to bear another point of view on the phenomena, whose expression might lead the original author(s) to revise the way they think about and present their observations and conclusions" (Longino 1990, 68–69).

Peer-review practices implicitly acknowledge that selection processes should not be sensitive to some social features. Anonymous review aims to avoid some forms of social bias, including the Matthew effect ("For unto everyone that hath shall be given") and the halo effect (where individuals gain prestige by association) (Merton 1968; Cole 1992). By making the identity of the author anonymous, the system avoids a circular prestige bias in which the most prestigious publish by virtue of their status, become more prestigious as a result,

which enables them to publish even more. Masking the identity of the author also avoids the reversed version of the prestige cycle where the less prestigious fail to publish due to their status, fail to become prestigious as a result, which reinforces their decreased publication ability. By masking the identity of reviewers, the system avoids having reviews be used to court professional favor or ire from authors.

Peer review involves the following sequence of decisions on the part of authors, editors, and reviewers:

> *Step 1.* Author(s) submit work to public venue for review.
> *Step 2.* Editor(s) decide whether to reject work outright or send to reviewer(s).
> *Step 3.* If the work is to be sent to reviewers, editor(s) choose reviewer(s).
> *Step 4.* Reviewer(s) comment on work and make recommendations about whether to accept, reject, and/or recommend revisions.
> *Step 5.* Editor(s) evaluate whether to accept, reject, and/or recommend revisions.

These decisions are made under uncertainty. Authors must make decisions about submitting work to a particular venue without knowing whether it will be accepted or not. Reviewers and editors must make decisions about the quality of the work without knowing what the work's reception in and impact on the field would be. We will now examine these stages and discuss how they may be vulnerable to forms of social bias.

Step 1. The author(s) makes the determination to submit the paper to a conference or journal. This decision requires the author(s) to weigh considerations such as the paper's fit with the conference or journal's audience and editors, as well as the author's expectations about the quality and timing of feedback. Since rejection rates are high, authors have an interest in taking the time to improve the quality of the work before submitting.

Step 2. Once editors receive the paper, they determine whether to reject the paper outright (otherwise known as a "desk rejection") or send it on to reviewers.[6] A higher desk-rejection rate filters out the weakest papers, freeing editors from soliciting and reviewers from writing reviews for less compelling work. From the survey data received, 93% of philosophy journal editors reject papers at this stage. Among these editors, the rate of desk rejection varies widely, from 2%–65%, with a mean desk rejection rate of 22%.

Rejection at this stage can have socially and epistemically less innocuous consequences when coupled with other editorial practices. For about 81% of journals, the identities of authors are known to editors. For these journals, desk-rejection rates were slightly lower (mean of 20%, range 2–65%) than for journals in which author identities were anonymous to editors (mean of 30%, range 10–60%).

Transparency about the author's identity at this stage makes room for social stereotypes—about prestige, gender, and/or race—to play an undue role. For example, a classic study found that when articles already published in highly prestigious psychology journals were resubmitted to the same journals, but under fictitious names affiliated with low-prestige institutions, nearly 90% were rejected—a rejection rate higher than the 80% rejection rate of the journals in question (Peters and Ceci 1982). Research demonstrates that male job applicants are deemed more hirable than female applicants, despite their having identical curricula vitae (Steinpreis, Anders, and Ritzke 1999). Similar effects occur for white versus African American applicants (Bertrand and Mullainathan 2004).

In these studies, evaluators are unaware of these effects. Instead, they are under the mistaken impression that their judgments are grounded on normatively appropriate considerations. In the study on papers by individuals from high- versus low-prestige institutions, evaluators rejected papers from low-prestige institutions due to "serious methodological flaws" (Peters and Ceci 1982). In the study on male versus female job candidates, evaluators judged male applicants as having more adequate teaching, research, and service experience (Steinpreis, Anders, and Ritzke 1999). These findings demonstrate that public standards of evaluation and internalist practices of giving and responding to reasons cannot protect against socially determined ways of implementing those standards.

Step 3. If the paper is deemed sufficiently worthy of review, the next decision for editors is to select reviewers. Fifty percent reported giving consideration to choosing reviewer(s) with opposing viewpoints to the author. Fifty percent reported giving consideration to choosing reviewer(s) on the basis of whether they had published in the journal before. Thirty-eight percent reported giving some consideration to choosing reviewers who had different viewpoints from one another (though the majority of these editors also reported sometimes relying on a single review). In cases where editors took into consideration both disagreement with the author and with one another, the majority rated inter-reviewer disagreement as the more important consideration.

About 63% of general philosophy editors reported sometimes making determinations on the basis of a single review. Approximately 25% of editors reported sometimes relying on three or more reviewers (though at least half of these editors remarked that this was not normally the case). In contrast, journals published by the American Psychological Association moved from a two-reviewer system in the 1950s to a three- to five-reviewer system in the 1990s (Zanna 1992).

Step 4. Reviewers who have agreed to evaluate a particular piece provide the kind of critical feedback to ground their own and editors' judgments about whether to reject, accept, or recommend revisions on an article. In our sample, 90% of philosophy journals report making author identities anonymous to reviewers, while 10% report not adopting anonymizing practices at this stage.[7]

In order for reviewer critiques to be relevant to the research being evaluated, reviewers must appeal to public standards of evaluation. However, even when evaluative standards are stable and enjoy community-wide endorsement, their application remains underdetermined by their definitions or characterizations. Because this is so, social features from the context of evaluation can modulate how these criteria are invoked and applied in the reviewing process.

For example, reviewers can appeal to shared standards in order to provide disproportionately more negative as opposed to positive comments (or vice versa). Although the negative and aggressive style of philosophical evaluation has been remarked upon through anecdote and personal experience, there has not been a more systematic study of philosophy's relative negativity (Moulton 1983; Lachs 2004; Haslanger 2008; Lamont 2009). To examine this, we turned to reviews written by philosophers and psychologists for papers submitted to the annual Cognitive Science Society conference.

Two raters coded 423 reviews for 221 conference paper submissions from the 2007 Cognitive Science Society conference, including action-editor reviews (action editors choose reviewers, read their reviews, and then write summary reviews with recommendations that the conference chair uses to make final decisions). There were 53 submissions by philosophers and 168 by psychologists; and, there were 112 philosophy reviews and 311 psychology reviews. The review database was obtained by the second author as conference historian (someone who reports to the society regarding conference submission and reviewing trends). The reviews were stripped of paper name, author identity, reviewer identity, and action-editor identity. Retained for data analysis but hidden from the coders was the status of the reviewer (external reviewer or action editor), discipline of the reviewer, and the recommended decision for the paper (accept as talk, accept only as poster, or reject submission entirely). The raters were also not informed of the hypotheses of the analyses. Inter-rater reliability assessed on a random 100 reviews was over 90%.

We use two scales: a 1-to-5 scale on the nature of positive comments (5 = several highly positive comments given, 4 = one highly positive comment given, 3 = several mildly positive comments given, 2 = one mildly positive comment given, 1 = no positive comments) and 1-to-5 scale on the nature of negative comments (5 = makes several inflammatory comments, 4 = makes one inflammatory comment, 3 = notes several negative issues but not in an inflammatory tone, 2 = notes one negative issue but not in an inflammatory tone, 1 = no negative issues noted). We understood discipline of the submission to be as it was indicated by the reviewer. Because program committee members assign reviewers, the discipline of the reviewer tends to be a close match to the discipline of the program committee member, which in turn is a close match to the discipline of the paper submission. Thus, we used discipline of the submission as a proxy for discipline of the reviewer.

Twenty-five of the reviews indicated that philosophy was the only discipline of the submission, and 125 reviews indicated that philosophy was among multiple disciplines in the review. We conducted analyses using the more strict 25 and again using the more liberal 125. The basic pattern of results was the same in either case; here we report only the statistically more powerful larger set. By comparison, we used the 311 submissions indicating psychology as the only discipline. The imbalance in sample sizes reflects the disproportionate representation of psychologists relative to philosophers at the conference (Schunn, Crowley, and Okada 1998). All pattern differences described here were statistically significant differences at $p < .05$ using t-tests.

Overall, there are no significant differences in the amount of negative or positive content in reviewer comments (negativity and positivity ratings 2 and 3). However, there is a difference in the extreme end of the negative scale having to do with the inflammatory tone/expression of negative comments (negativity ratings 4 and 5). The percentage of philosophy papers with inflammatory negative comments is twice that of psychology papers (13% vs. 7%). In terms of reviews with multiple inflammatory comments, the rate is almost triple (4.5% vs. 1.6%). The difference in rate of inflammatory negative comments is especially salient when focusing on just the reviews in which the reviewer recommended rejection of the paper (see Figure 1)—here, almost a third of the philosophy paper reviews include inflammatory negative comments. About
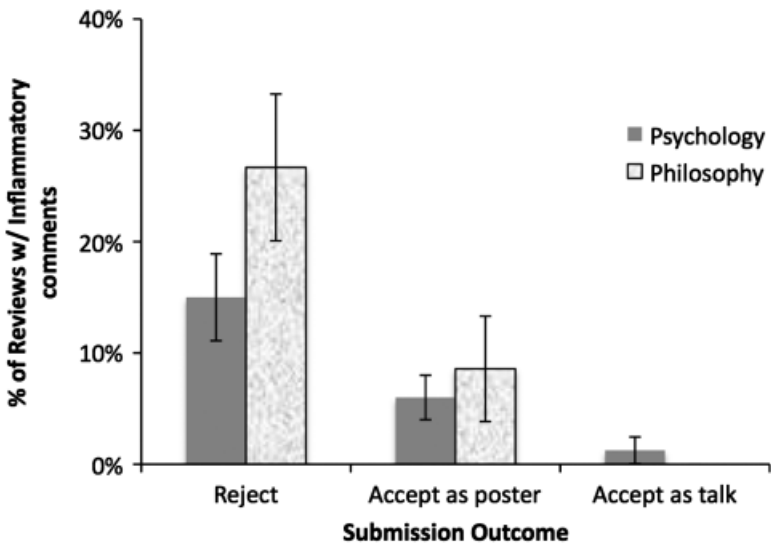

Figure 1. *Mean proportion of reviews (with standard error bars) with inflammatory comments (rating 4 or 5) as a function of paper discipline and reviewer recommendation.*

41% of philosophy submissions were ultimately rejected, while only 20% of psychology submissions were rejected.[8] On the positive side of the review comments, there are no consistent differences by discipline, overall or at the extreme end of positivity. Thus, philosophy reviews noted just as many positive features of submitted papers, but then added more inflammatory comments in the rejection recommendations.

The data suggest that philosophers, as reviewers, more frequently engage in inflammatory commenting styles than do psychologists in this shared intellectual world. Note that our claim is not that philosophers alone engage in inflammatory styles. Our claim is about the relative frequency of such behavior in this academic context. This finding is consistent with the suggestion that the culture of a discipline, when internalized as background beliefs of individual evaluators, can determine how public standards are implemented in the evaluation of manuscripts.

Step 5. Reviewer comments and evaluations help guide editorial judgments about whether an article is sufficiently viable for publication, in its current or revised form. Some editors solicit an additional review when a single positive review is received (and no other reviews have been sought) or when discrepant evaluations are received (both involve a return to Step 4).

The stability of editorial standards can help to protect the review process from being an arbitrary function of individual reviewer standards, style, and intellectual orientation. Ideally, editors should not tabulate referee recommendations, but adjust reviewer standards to achieve consistency in standards across submissions (Glenn 1982). Along these lines, we found that 80% of philosophy journal editors report exercising editorial judgment by making occasional recommendations that go against the recommendations of at least a single reviewer.

However, on average, a single negative review has more power than a single positive review on editorial determinations. About 40% of editors "never" or "rarely" accept a paper receiving a single negative review (of these editors, 80.0% report sometimes relying on a single review). In contrast, only 20% of editors claim that they do not reject a paper that has received a single positive review (of these editors, all report sometimes relying on a single review).

The power of negative reviews might be driven, in part, by the relatively high rejection rates at philosophy journals.[9] On the basis of a follow-up questionnaire sent to editors at philosophy journals, we discovered that philosophy rejection rates averaged about 92%, with reported rates ranging from 80% to 95%.[10] In this calculation, "revise and resubmit" decisions were counted as rejections, except in unusual cases where nearly all revise and resubmit papers were accepted upon resubmission (these were counted as accepted papers). In contrast, we found that top psychology journals enjoy a much lower rejection rate (mean of 78%, with a range from 68% to 86%).[11] The rate with which philosophy papers received revise and resubmit decisions and the percentage

with which resubmitted papers were accepted did not make up for the difference in rejection rates. Note that the physical sciences enjoy an even lower rejection rate (20–40%) (Zuckerman and Merton 1971; Hargens 1988). Sociologists disagree about whether these differences are driven by market factors such as the number of journal pages available (as dictated by publishers) versus the length and number of manuscripts submitted (Zuckerman and Merton 1971; Cole 1983; Hargens 1988).

Our analysis of action-editor reviews at the Cognitive Science Society suggests that, even when such market factors are removed, negative reviews continue to have more power than positive reviews in philosophy. The Cognitive Science Society conference produces electronic-only (web and CD-ROM) proceedings, so the typical cost pressures are much lower. There are no formal quotas for accepted or rejected papers. Our analysis demonstrates that action editors, who write summary reviews and make action recommendations to the conference chair, almost never made inflammatory comments, either reflecting some understanding among more senior participants that inflammatory tone/expression was not appropriate in reviews or reflecting differences in the genres of action letters versus reviews. However, action editors' reviews of philosophy papers were more likely to list multiple problems (negativity rating = 3) than were the action editors' reviews for psychology papers (88% vs. 67%). The philosophy editors' recommendations were in line with the negative ratings and comments received from reviewers. About 41% of philosophy submissions were ultimately rejected, whereas the rejection rate for psychology submissions was only 20%. This large difference in rejection rates suggests the possibility that a culture of negative evaluation in philosophy might detrimentally affect acceptance rates for philosophy papers in this interdisciplinary context. That is, evaluative styles ingrained in the culture of the discipline may drive the implementation of evaluative standards in ways that decrease the community's inclusion, representation, and success.

Some have suggested that philosophers receive fewer awards by interdisciplinary fellowship/grant agencies because of philosophy's negative evaluative style (Lachs 2004). Michéle Lamont's research on interdisciplinary fellowship panels corroborates this claim. She describes program officers' warnings to stay "open-minded" toward philosophy proposals to be "as close to a plea for 'affirmative action' toward a discipline as I witnessed during my study of funding panels" (Lamont 2009, 64). These pleas were a response not just to the negative and aggressive disciplinary reviews for philosophy proposals, but to interdisciplinary reviewers' conceptions of philosophers as members of the broader academic community. She notes that "[s]everal panelists expressed at least one of the following views: (1) philosophers live in a world apart from other humanists, (2) nonphilosophers have problems evaluating philosophical work, and they are often perceived by philosophers as not qualified to do so,

(3) philosophers do not explain the significance of their work, and (4) increasingly, what philosophers do is irrelevant, sterile, and self-indulgent" (Lamont 2009, 64).

Negative evaluative styles are not the only possible source of social bias in the final evaluations of manuscripts and proposals. An author's (perceived) membership in a social category may also play a role by means of casuistry—the (conscious or subconscious) practice of justifying one's evaluations by invoking standards in ways that mask or overlook the causal impact of normatively inappropriate considerations. For example, in evaluating a manuscript, editors are not dealing with a single criterion, but a combination of criteria that can be weighed and combined in different ways. Research suggests that evaluators subconsciously tailor how they weigh different criteria in ways that favor individuals belonging to preferred social groups. For example, when asked to hire a construction company manager—a position requiring both experience and a strong background in engineering—participants preferred the male candidate when he was more educated but less experienced than the female candidate. However, when the gender identities were reversed, only a minority of participants picked the more educated, but less experienced female candidate. Overwhelmingly, evaluators justified their choices by citing the qualification that favored the male candidate (Norton, Vandello, and Darley 2004). Ironically, this effect is more exaggerated for evaluators high in self-perceived objectivity (Uhlmann and Cohen 2005). Although evaluators are able to provide reasons that comport with shared and agreed upon criteria, these criteria can be weighed in ways that favor qualities belonging to individuals of a preferred social group.

In peer evaluations, social and cultural beliefs impact how evidence gets perceived and weighed despite shared standards of evaluation. This challenge to the impartiality of peer evaluation suggests that communities meeting Longino's conditions for transformative criticism can nevertheless reproduce and perpetuate forms of social bias in their judgments about which individuals are more expert/competent and which ideas and critiques deserve inclusion and response. The theoretical problem goes beyond local concerns about the peer review of manuscripts/proposals in philosophy. So, in what follows, we will broaden our focus and discuss social biases in peer evaluations and solutions for knowledge communities more generally.[12]

SOCIAL SOLUTIONS

Longino can account for these forms of bias by arguing that evaluations about the quality of research are themselves hypotheses underdetermined by the data. This theoretical move creates room for background beliefs to make salient some aspects of the data as evidentially relevant to evaluative hypotheses over

others. It also allows evaluators who agree about the data and standards of evaluation to disagree about the best evaluative hypothesis. This revised account can accommodate and predict for socially driven peer evaluations despite shared evaluative standards.

How can Longino's account address epistemically harmful forms of social bias in peer evaluations? According to her account, problems arise when a community's background beliefs are homogeneous, rendering them less visible and available for critique. Her solution is to require that multiple points of view be cultivated and included so that hypotheses and background beliefs are subjected to broad, critical scrutiny. This move turns objectivity into a form of "social value management rather than absence of social values" (Longino 2002, 50). We propose that, to address social bias in peer evaluations, Longino should require a specific form of perspectival diversity: namely, a diversity of perspectives bearing on cultural norms and beliefs within and about the knowledge community. In what follows, we will discuss how this solution addresses the kinds of social biases discussed earlier.

COGNITIVE SCHEMAS AND CULTURAL DIVERSITY

Like stereotypes, schemas are abbreviated beliefs we hold about the main characteristics (and their relationships) of individuals, groups, and events (Fiske and Taylor 1991; Valian 1998). Often implicit, schemas serve the cognitive function of helping us quickly interpret, explain, and predict actions and events. They do this by categorizing events and people according to already held social concepts and generalizations, which foregrounds category-consistent features.

As such, schemas work as implicit background beliefs do in evaluations of scientific hypotheses: "states of affairs are taken as evidence in light of regularities discovered, believed, or assumed to hold" (Longino 1990, 41). Like other any other kind of background belief, schemas may be accurate. However, when they are not or fail to apply to a particular case, schemas are problematic for their resilience to counterevidence: because they structure our cognition so as to make some features salient while overlooking others, it is difficult for us to attend to and see the relevance of evidence that would challenge the schema's accuracy.[13]

Specifying the kind of diversity required is important here, as traditional conceptions of "diversity" do not work to counterbalance cognitive schema effects. Women and minorities characteristically fall prey to cognitive schema effects. So, we would not expect straightforward affirmative action measures ensuring the education, cultivation, and hiring of women and minorities by themselves to address the background beliefs of concern. Other features of the liberal democratic ideal of public deliberation fail to address these forms of bias. The inclusion of those who sincerely espouse norms of equality is not sufficient since

such beliefs—formed by conscious, deliberate reasoning processes—do not displace the workings of implicit, automatic processes involved in implicit bias (Chaiken and Trope 1999; Huebner 2009). Neither are public avowals of egalitarian beliefs sufficient: research demonstrates that the opportunity to affirm one's non-prejudicial views can actually exaggerate these biases (Monin and Miller 2001). Injunctions instructing evaluators to be more "objective" seem unhelpful as well: ironically, gender biases increased for evaluators primed with questions about how objective they perceive themselves to be, where self-perceptions of objectivity correlated with degree of bias (Uhlmann and Cohen 2007).

One might understand such empirical research as identifying a set of implicit background beliefs that are homogeneous in the community's or society's culture. To address this, Longino should require a specific kind of diversity: a diversity of individuals who hold different implicit assumptions about the cognitive authority of individuals associated with different stereotypes. According to this refinement of "tempered equality," communities characterized by a diversity of background beliefs about social identities and epistemic authority might be in a better position to see and critique evaluative hypotheses and the background beliefs upon which they rest. Furthermore, community-wide practices of critique and response would be empowered to transform community-wide evaluations of the work and expertise of its members.[14]

This refinement would require that procedurally objective communities make efforts to foster this kind of diversity by means of education programs that raise awareness about the existence of such implicit biases as well as fostering diverse implicit attitudes. This refinement further suggests that procedurally objective communities be characterized not just by a diversity of perspectives that bear directly on the content of theories, but on implicit attitudes about fellow community members as (perceived) members of various social categories.

NEGATIVE BIAS AND CULTURAL DIVERSITY

Negative disciplinary styles of evaluation are also social in the sense that they are instilled and reinforced through educational and professional experiences involving evaluations of one's own work and the work of others. Disciplinary training inculcates a sensibility about which features of disciplinary research are more significant than others.[15] Once ingrained, these intuitions structure cognition so as to make some aspects salient while overlooking others. In so doing, these intuitions can serve as implicit background assumptions identifying evidential relationships between hypotheses (about the quality of some research) and the data.

When evaluating hypotheses about the value of a piece of research, antecedent judgments of significance modulate the respects in which a piece of research can be said to be (or fail to be) cogent, well-argued, well-grounded, innovative, and so on. Disciplinary training can be understood as developing

both an understanding of these evaluative standards and the significant features to which they should be applied. Shared evaluative standards can be used to make negative evaluative claims (negative in content alone, or in content plus tone), just as shared standards might be used to make positive evaluative claims (as can happen in recommendation letters). In both cases, features of the social context drive antecedent judgments of which features are significant—features to which shared evaluative standards are then applied.

Knowledge communities characterized by a homogeneous set of background beliefs about the significance of negative features of research in peer evaluations might be understood as needing a special kind of diversity: namely, a diversity of individuals who hold different background beliefs about the significance of negative versus positive evaluative claims. According to this refinement of *tempered equality*, individuals with different background beliefs might be in a better position to render visible and to critique the negativity of others' evaluative styles. This would require that procedurally objective communities make efforts to foster forms of diversity bearing not on the *content* of theories but on the cultural beliefs and norms of the community itself.

OVERVIEW: SOCIAL BIAS

Understanding how background beliefs can modulate evaluations explains a *prima facie* paradox posed by empirical results in the social sciences: namely, even though reviewers agree about the criteria to be used in evaluating new research (Sternberg and Gordeeva 1996), inter-reviewer agreement about whether a manuscript meets these criteria is so low as to be considered "poor" by psychometric standards (Scott 1974; Marsh and Ball 1989). According to this account, the variability of reviewers' evaluations along various criteria can be attributed to differences in background beliefs.[16]

It is important to note that, although cognitive schemas and evaluative styles are both biases that originate in the culture of a knowledge community, this does not imply that they are, therefore, of a kind. As Miranda Fricker observes, those who are objects of biased evaluations of credibility/authority as the result of their social identities tend to be susceptible to broader, systematic forms of social injustice along economic, educational, professional, and political lines (Fricker 2007). In contrast, members of an academic community who are susceptible to negative evaluations by their peers because of the cultural norms fostered within the community are not vulnerable to broader forms of social injustice, though they may enjoy fewer publication pages, fewer grant/fellowship awards, and fewer conference slots in the narrower social world of academia. This observation is not meant to belittle the impact that negative evaluative styles can have on a knowledge community: as these rejections add up, members of the knowledge community can accumulate disadvantage, just

as women and minorities can accumulate disadvantage as a result of cognitive schemas (Merton 1968; Valian 1998).

Conclusion

Procedural objectivity is achieved when communities cultivate and maintain social structures that promote attention and responsiveness to the background beliefs licensing inferences from data to hypotheses. Decisions about which critiques are given public voice, which are responded to, and the relative authority of their authors are decided on the basis of evaluative judgments about the quality or value of those critiques. These evaluative judgments—a form of judgment under uncertainty—are themselves hypotheses based on data, including the content of the arguments and claims made, and their relation to the community's larger body of knowledge.

However, like any hypothesis, these evaluations are, in principle, underdetermined by the data. Such underdetermination creates space for background beliefs to influence judgments about the evidential relationship between data and evaluative hypotheses about community members' arguments and claims. In order to create social structures that promote attention and responsiveness to these background beliefs, Longino's proposal for obtaining procedural objectivity must require a form of diversity that bears not on the content of the community's knowledge claims but on the culture of the community itself. In the spirit of Longino's rejection of the cognitive/social dichotomy, this way of addressing social bias is cognitive insofar as it deals with evidential or justifying reasons for evaluative judgments; it is social insofar as it relies on social features of knowledge communities to support the well-functioning of their deliberations.

If this argument is sound, then the theoretical and reflexive projects in this paper are importantly connected to one another. As Longino observes, "[a]scertaining in greater detail the practices and institutional arrangements that facilitate or undermine objectivity in any particular era or current field, and thus the degree to which the ideal of objectivity is realized, requires both historical and sociological investigation" (Longino 1990, 80–81). In this vein, this paper engages in social-scientific investigation of philosophy's current peer-review practices to assess how the processes governing access to its public venues might be vulnerable to epistemically detrimental forms of social bias. By illustration, this project's refinement of the tempered equality of intellectual authority underscores the more general meta-philosophical point that empirical work can and should inform our conceptualization of the conditions for transformative criticism.

Although our analysis provides a template for this naturalized genre of research, it is only a start. More work is needed to address further questions about how peer review governs access to public venues. At a minimum, supporting or

rejecting hypotheses about the role of cognitive schemas in peer review would require gathering information from journals about the rates at which papers by women/men and minorities/whites are submitted and rejected (with or without reviews). Even better data would involve comparisons of these rates split by whether author gender or ethnicity is known to the reviewers or action editors. Supporting or rejecting hypotheses about the more general role of negative evaluative styles in philosophy would require gathering and analyzing reviews and editorial decisions for a broader range of publications and conferences. Such research could explore whether differences in review style (for example, tone, content, and recommendations) also vary by reviewer factors other than disciplinary affiliation, such as gender, race/ethnicity, subdiscipline, and seniority.[17]

Future work should focus on the agency of authors and how peer-review norms and processes interact with their judgments under uncertainty. How do authors choose publication venues, make decisions about whether to undertake suggested revisions versus submitting elsewhere, or decide to shelve a project? How do authors come to learn what the norms of peer review themselves are, especially vis-à-vis the more problematic practices? What do answers to these questions tell us about how to conceptualize and systematize public venues, public standards, the tempered equality of intellectual authority, and critical uptake?

Procedural changes—such as making author identities anonymous to editors—can help ameliorate local worries about the role cognitive schemas might play in peer review (Lee and Schunn 2010). However, anonymizing practices are not appropriate for other domains of academic achievement, such as career prizes and invitations to write for special issues or edited volumes. Empirical and theoretical inquiry into the social-cognitive factors responsible for how attributions of expertise are "tempered" as well as the factors contributing to the development of such expertise can inform how we should conceptualize the tempered equality of intellectual authority.

The explicit use and acknowledgment of one's work in peer research confers an even more basic form of scholarly recognition (Merton 1996). Evaluating the proper functioning of critical uptake requires empirical and theoretical examination of how research becomes integrated into knowledge communities. For example, the degree to which citation patterns track author gender and ethnicity, contextualized within seniority, eminence, and subfield, can provide perspective on community responsiveness to diverse perspectives and motivate new lines of inquiry about the kinds of critical uptake that matter most.

The norms involved in evaluations of expertise and research quality are themselves open to empirical and theoretical inquiry. Future work should explore whether public standards are really shared across knowledge communities or become sequestered into somewhat insulated subdisciplines and sub-communities. Such research would be in a position to evaluate how autonomous or well integrated these areas are within a knowledge community. Research on

whether sub-community norms (such as reviewing styles) track demographic factors would provide perspective on the kind of evaluative differences that can make integration of diverse perspectives, access to public venues, and community responsiveness even more challenging.

Finally, future research should identify how to promote the proper epistemic functioning of perspectival diversity in light of empirical research suggesting that straightforward affirmative action measures are not sufficient. Are measures of implicit beliefs by individual community members needed (by means of something like an Implicit Association Test) despite invasions to privacy? If proxies for measuring implicit bias are used instead, would the assumptions needed to justify the validity of such proxies be objectionable from epistemic, social, and/or political points of view? Such questions illustrate the intricate interconnections among empirical, philosophical, methodological, and social research in theorizing about objectivity conceived as the management of social values.

NOTES

1. Longino is not averse to generalizing her account to include non-scientific knowledge communities. She suggests that "regard for logic and for the evidence of sensory experience are universal features of knowledge constructive communities, even partially constitutive of such communities" (Longino 2002, 162).

2. Lamont's research on interdisciplinary peer review (based on interviews of program officers, chairpersons, and panelists for interdisciplinary fellowships) included philosophers; however, there were only two philosophers out of the total of eighty-four panelists interviewed (Lamont 2009). Decades ago, sociologists of science surveyed five philosophy journals in a comparison of rejection rates across disciplines (Zuckerman and Merton 1971). The peer-review practices of other disciplines have enjoyed more sustained and recent attention (Peters and Ceci 1982; Lock 1985; Hargens 1988; Cicchetti 1991; Starbuck 2005; Hargens and Herting 2006).

3. To protect anonymity, the authors promised not to discuss, in writing or conversation, the details of individual editors' responses or the identities of those who did or did not respond. The following journals were invited to respond to the original survey: *American Philosophical Quarterly*, *Analysis*, *Australasian Journal of Philosophy*, *Canadian Journal of Philosophy*, *Continental Philosophy Review*, *European Journal of Philosophy*, *Inquiry*, *Journal of Philosophy*, *Mind*, *The Monist*, *Noûs*, *Pacific Philosophical Quarterly*, *Philosophers' Imprint*, *Philosophia*, *Philosophical Perspectives*, *Philosophical Quarterly*, *Philosophical Topics*, *Philosophical Review*, *Philosophical Studies*, *Philosophy and Phenomenological Research*, *Proceedings of the Aristotelian Society*, *Ratio*, *Review of Metaphysics*, *Southern Journal of Philosophy*, and *Theoria*. Editors' responses were collected from April through August 2009.

4. The prestige of journals is a difficult thing to measure in the humanities, where traditional bibliometric methods tracking patterns of citation are not appropriate or available. In assessing whether editor responses were received from general philosophy journals across a range of "prestige," we undertook imperfect but available methods: we referred to Brian Leiter's recent poll (of more than 500 philosophers) on their opinions on this matter—this poll resulted in a ranked list of nineteen journals (Leiter 2009). We chose to supplement this list with a few additional journals identified by Mark Colyvan, Professor of Philosophy at the University of Sydney and Director of the Sydney Centre for the Foundations of Science, on his website (Colyvan n.d.).

5. For a discussion about the sensitivity of evaluative claims to evidence, see Anderson 2004.

6. There is at least one exception. According to *Analysis*'s online description of its review practices, the editor accepts some manuscripts without sending them to reviewers.

7. Informal reports from reviewers suggest that some journals with formal anonymous review policies occasionally violate this policy.

8. There are no formal quotas or ideal rejection rates conveyed to reviewers or action editors.

9. As Longino notes, page limitations "contribute to the marginalization of critical discourse" (Longino 2002, 129). In order to allow for the possibility of transformative criticism, communities relying on publication as their primary public venue should move toward increasing publication space within their knowledge communities.

10. This second survey was sent to philosophy journals that responded to the original survey. We received eleven responses from journals across a range of prestige with figures on their acceptance and rejection rates for submitted and resubmitted manuscripts, as well as figures on the rate of revise and resubmit decisions. These figures were supplemented with data from journal websites when available.

11. We obtained data about thirteen of twenty journals in the psychology category of ISI's Journal Citation Report with the highest impact factor. These twenty include: *Advances in Experimental Social Psychology, American Psychologist, Annual Review of Psychology, Behavioral and Brain Sciences, Clinical Psychology Review, Developmental Psychopathology, The Journal of Child Psychology and Psychiatry, Journal of Clinical Psychiatry, Journal of Cognitive Neuroscience, Journal of Consulting and Clinical Psychology, Journal of Experimental Psychology: General, Journal of Personality and Social Psychology, Journal of the American Academy of Child & Adolescent Psychiatry, Personality and Social Psychology Review, Psychological Bulletin, Psychological Medicine, Psychological Methods, Psychological Review, Psychological Science,* and *Trends in Cognitive Science.* Rejection rates were obtained from the American Psychological Association's annual report on their journals' statistics as well as from email responses from journal editors.

12. Note that procedural changes designed to ameliorate social bias in peer review would not address the impact of social bias in attributions of cognitive authority and community responsiveness (or the lack thereof).

13. Cognitive schemas are not completely insensitive to evidence. Research suggests that very strong evidence can counteract the influence of stereotype-based judgments. For example, in the hiring study above, a woman was as likely to be hired

as a man in cases involving exceptional curricula vitae of those who went up for early tenure (Steinpreis, Anders, and Ritzke 1999). Other research corroborates the suggestion that cognitive schemas affect not the ability, but the degree to which evaluators see data as evidence for competence and ability (Biernat and Kobrynowicz 1997; Trope and Thompson 1997; Foschi 2000; Biernat and Ma 2005).

14. Some have argued that tempered intellectual authority should be extended to individuals beyond scientific communities who have access to relevant evidence, methodological strategies, and conceptual resources (Solomon 2008; Wylie 2008).

15. In addition to disciplinary training, researchers working in the sociology of ideas have argued that an author's intellectual self-concept, values, and commitments can also shape judgments about significance in disciplinary orientation and approach (Gross 2002; 2003).

16. Lamont's analysis cashes this difference out in slightly different terms: she argues that panelists on interdisciplinary fellowship panels attached different "meanings" to abstract evaluative criteria (such as "originality") and that these differences originated in disciplinary, subdisciplinary, and personal differences (Lamont 2009).

17. Thanks to Louise Fortmann for suggesting this line of inquiry.

REFERENCES

Anderson, Elizabeth. 2004. Uses of value judgments in science: A general argument, with lessons from a case study of feminist research on divorce. *Hypatia* 19 (1): 1–24.

Bertrand, Marianne, and Sendhil Mullainathan. 2004. Are Emily and Greg more employable than Lakisha and Jamal? *American Economic Review* 94 (4): 991–1013.

Biernat, Monica, and Diane Kobrynowicz. 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 (3): 544–57.

Biernat, Monica, and Jennifer E. Ma. 2005. Stereotypes and the confirmability of trait concepts. *Personality and Social Psychology Bulletin* 31 (4): 483–95.

Boltanski, Luc, and Laurent Thévenot. 2006. *On justification: Economies of worth.* Trans. Catherine Porter. Princeton: Princeton University Press.

Cambrosio, Alberto, Peter Keating, Thomas Schlich, and George Weisz. 2006. Regulatory objectivity and the generation and management of evidence in medicine. *Social Science & Medicine* 63:189–99.

Chaiken, Shelly, and Yaacov Trope. 1999. *Dual-process theories in social psychology.* New York: The Guilford Press.

Cicchetti, Domenic V. 1991. The reliability of peer review for manuscript and grant submissions: A cross-disciplinary investigation. *Behavioral and Brain Sciences* 14:119–86.

Cole, Stephen. 1983. The hierarchy of science? *American Journal of Sociology* 89:111–39.

———. 1992. *Making science: Between nature and society.* Cambridge, Mass.: Harvard University Press.

Colyvan, Mark. n.d. The best philosophy journals. http://homepage.mac.com/mcolyvan/journals.html (accessed March 27, 2009).

Fiske, Susan T., and Shelley E. Taylor. 1991. *Social cognition*. 2nd ed. New York: McGraw Hill.

Foschi, Martha. 2000. Double standards of competence: Theory and research. *Annual Review of Sociology* 26:21–42.

Fricker, Miranda. 2007. *Epistemic injustice*. Oxford: Oxford University Press.

Glenn, Norvall D. 1982. The journal article review process as a game of chance. *Brain and Behavioral Sciences* 5:211–12.

Gross, Neil. 2002. Becoming a pragmatist philosopher: Status, self-concept, and intellectual choice. *American Sociological Review* 67:52–76.

———. 2003. Richard Rorty's pragmatism: A case study in the sociology of ideas. *Theory and Society* 32 (1): 93–148.

Hargens, Lowell L. 1988. Cognitive consensus and journal rejection rates. *American Sociological Review* 53:139–51.

Hargens, Lowell L., and Jerald R. Herting. 2006. Analyzing the association between referees' recommendations and editors' decisions. *Scientometrics* 67 (1): 15–26.

Haslanger, Sally. 2008. Changing the ideology and culture of philosophy: Not by reason (alone). *Hypatia* 23 (2): 210–23.

Huebner, Bryce. 2009. Troubles with stereotypes for Spinozan minds. *Philosophy of the Social Sciences* 39 (1): 63–92.

Lacey, Hugh. 1999. *Is science value free? Values and scientific understanding*. New York: Routledge.

Lachs, John. 2004. The future of philosophy. *Proceedings and Addresses of the American Philosophical Association* 78 (2): 5–14.

Lamont, Michèle. 2009. *How professors think: Inside the curious world of academic judgment*. Cambridge, Mass.: Harvard University Press.

Lederman, Rena. 2006. The perils of working at home: IRB "mission creep" as context and content for an ethnography of disciplinary knowledges. *American Ethnologist* 33 (4): 482–91.

Lee, Carole J., and Christian D. Schunn. 2010. Philosophy journal practices and opportunities for bias. *APA Newsletter on Feminism and Philosophy* 10 (1): 5–10.

Leiter, Brian. 2009. The highest quality "general" philosophy journals in English. http://leiterreports.typepad.com/blog/2009/03/the-highest-quality-general-philosophy-journals-in-english.html (accessed March 27, 2010).

Lock, Stephen. 1985. *A difficult balance: Editorial peer review in medicine*. London: Nuffield Provincials Hospital Trust.

Longino, Helen. 1990. *Science as social knowledge*. Princeton: Princeton University Press.

———. 2002. *The fate of knowledge*. Princeton: Princeton University Press.

Mallard, Grégoire, Michèle Lamont, and Joshua Guetzkow. 2009. Fairness as appropriateness: Negotiating epistemological differences in peer review. *Science, Technology & Human Values* 34 (5): 573–606.

Marsh, Herbert W., and Samuel Ball. 1989. The peer review process used to evaluate manuscripts submitted to academic journals: Interjudgmental reliability. *Journal of Experimental Education* 57 (2): 151–69.

Merton, Robert K. 1968. The Matthew effect in science. *Science* 159 (3810): 56–63.

———. 1996. The Matthew effect, II (1988). In *On social structure and science*, ed. Piotr Sztompka. Chicago: University of Chicago Press.

Monin, Benoit, and Dale T. Miller. 2001. Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology* 81 (1): 33–43.

Moulton, Janice. 1983. A paradigm of philosophy: The adversary method. In *Discovering reality*, ed. Sandra Harding and Merrill B. Hintikka. Hingham: D. Reidel.

Nelson, Lynn Hankinson. 1990. *Who knows: From Quine to a feminist empiricism*. Philadelphia: Temple University Press.

Nisbett, Richard E., and Timothy DeCamp Wilson. 1977. Telling more than we can know: Verbal reports on mental models. *Psychological Review* 84 (3): 231–59.

Norton, Michael I., Joseph A. Vandello, and John M. Darley. 2004. Casuistry and social category bias. *Journal of Personality and Social Psychology* 87 (6): 817–31.

Peters, Douglas P., and Stephen J. Ceci. 1982. Peer-review practices of psychological journals: The fate of published articles, submitted again. *Behavioral and Brain Sciences* 5:187–255.

Schunn, Christian D., Keven Crowley, and Takeshi Okada. 1998. The growth of multidisciplinarity in the Cognitive Science Society. *Cognitive Science* 22 (1): 107–30.

Scott, William A. 1974. Interreferee agreement on some characteristics submitted to the *Journal of Personality and Social Psychology*. *American Psychologist* 29 (9): 698–702.

Solomon, Miriam. 2006. *Groupthink* versus *The wisdom of crowds*: The social epistemology of deliberation and dissent. *The Southern Journal of Philosophy* 44 (S1): 28–42.

Solomon, Stephanie R. 2008. Through a glass darkly: The hidden perspectives that challenge and redeem science's self-conception. PhD diss. Department of Philosophy, Emory University, Atlanta, Ga.

Starbuck, William H. 2005. How much better are the most-prestigious journals? The statistics of academic publication. *Organization Science* 16 (2): 180–200.

Steinpreis, Rhea E., Katie A. Anders, and Dawn Ritzke. 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): 509–28.

Sternberg, Robert, and Tamara Gordeeva. 1996. The anatomy of impact: What makes an article influential? *Psychological Science* 7 (2): 69–75.

Tollefsen, Deborah Perron. 2006. Group deliberation, social cohesion, and scientific teamwork: Is there room for dissent? *Episteme* 3 (1): 37–51.

Trope, Yaacov, and Erik P. Thompson. 1997. Looking for truth in all the wrong places? Asymmetric search of individuating information about stereotyped group members. *Journal of Personality and Social Psychology* 73 (2): 229–41.

Uhlmann, Eric Luis, and Geoffrey L. Cohen. 2005. Constructed criteria: Redefining merit to justify discrimination. *Psychological Science* 16 (6): 474–80.

———. 2007. "I think it, therefore it's true": Effects of self-perceived objectivity on hiring discrimination. *Organizational Behavior and Human Decision Processes* 104: 207–23.

Valian, Virginia. 1998. *Why so slow? The advancement of women*. Cambridge, Mass.: MIT Press.

Wylie, Alison. 2008. Legacies of collaboration: Transformative criticism in archaeology. Patty Jo Watson Distinguished Lecture. *American Anthropological Association*. San Francisco.

Zanna, Mark. 1992. My life as a dog (I mean editor). *Personality and Social Psychology Bulletin* 18:485–88.

Zuckerman, Harriet, and Robert K. Merton. 1971. Patterns of evaluation in science: Institutionalisation, structure and functions of the referee system. *Minerva* 9 (1): 66–100.