Social Knowledge in the Making

Edited by
CHARLES CAMIC, NEIL GROSS, AND MICHELE LAMONT

The University of Chicago Press
Chicago and London
Comparing Customary Rules of Fairness

Evaluative Practices in Various Types of Peer Review Panels

MICHELE LAMONT AND KATRI HUITONEN

Introduction

In this chapter we offer an analysis of intersubjective understandings concerning how to identify quality in the evaluation of scholarship. We focus on peer review, the practice by which the worth of research is evaluated by those with demonstrated competence to make a judgment. We analyze aspects of epistemic cultures, defined by Karin Knorr Cetina (1999) as machineries of knowledge—what we understand to be the social and cultural structures that channel, constrain, define, and enable the production and evaluation of knowledge. Of all forms of academic evaluation, peer review is the most widely used. While other methods of evaluation (such as bibliometric measures) are gaining in popularity, they indirectly rely on peer evaluations and are typically considered less credible than peer review, especially in the humanities and the social sciences.

Evaluation is a major aspect of the knowledge-making process. It has the function of gatekeeping, filtering, and legitimating knowledge. It is also a process where standards of excellence are set and maintained, contested, and reshaped. Evaluations are used not only for separating the qualified from the less qualified but also for distinguishing between competing types of high-quality research. Evaluation practices in general, and the practices of peer review panels in particular, are both informative of how standards are intersubjectively constructed and determinant of what is prized in research. Through various conventions, peer review panels enable certain types of knowledge while constraining others. Better understanding of these conventions is thus crucial for gaining insight into one particular link between practices and politics in knowledge production.

Most of the research on research assessment has focused on issues raised by Robert K. Merton’s influential work in the sociology of knowledge:
consensus in science; issues of universalistic and particularistic criteria of evaluation relating to the ethos of science; and the variously construed "Matthew" and "Halo" effects of reputation and prestige (Cole 1992; Cole and Cole 1988; Cole, Rubin, and Cole 1979; Liebert 1976; Menton 1996; Mulkay 1991). The question for most researchers is whether judgments about "irrelevant," particularistic characteristics, like the age and reputation of the author, affect the evaluation of his or her work. Other studies (Bell 1992; GAO 1994; Roy 1985) are also concerned with the fairness of the peer review process. The questions posed by these researchers imply that a unified and fair process of evaluating knowledge can be put in place once particularistic considerations are eliminated. The empirical literature on the topic largely confirms these normative assumptions by showing that peer evaluations are not highly correlated with factors other than scientific "quality," as measured by different quantitative indicators. Nevertheless, we contest the notion that one can separate cognitive from noncognitive aspects of evaluation, as we understand the evaluative process as deeply interactive, emotional, and cognitive and as mobilizing the self-concept of evaluators as much as their expertise (Lamont 2009).

More recent studies have revealed various "intrinsic biases" in peer review such as "cognitive particularism" (Travis and Collins 1991), "favoritism for the familiar" (Porter and Rossini 1985), and "peer bias" (Chubin and Hackett 1990; Fuller 2002). These effects show that peer review is not a socially disembedded quality-assessing process in which a set of objective criteria is applied consistently by various reviewers. In fact, the particular cognitive and professional lenses through which evaluators understand proposals shape evaluation. It is in this context that the informal rules that peer reviewers follow become important, as do the lenses through which they understand proposals and the emotions they invest in particular topics and research styles. Thus, instead of contrasting "biased" and "unbiased" evaluation, we aim to capture how evaluation unfolds, as it is carried on and understood by emotional, cognitive, and social beings who necessarily interact with the world through specific frames, narratives, and conventions, but who nevertheless develop expert views concerning what defines legitimate and illegitimate assessments, as well as excellent and less stellar research.

We are interested in how scholars serving on peer review panels construct the customary rules they follow in making funding decisions. More specifically, we aim to analyze the social conditions that lead panelists to an understanding of their choices as fair and legitimate and to a belief that they are able to identify the best and the less good proposals. As in Lamont (2009; see also Guetzkow, Lamont, and Mallard 2004; Lamont, Mallard, and Guetzkow 2006; and Mallard, Lamont, and Guetzkow 2009), our analysis puts much more emphasis on meaning in the evaluation process, including evaluation criteria, than did earlier studies. We interviewed individuals serving on various types of peer review panels that evaluate fellowship or grant proposals. We found that almost without exception, these individuals consider their deliberations fair and believe that meritocracy guides the process of selection—that is, that they are able to identify the top proposals. Their investment in a "culture of academic excellence" precludes them from framing the outcome of the deliberations as an expression of cronynism. 1

The distinctive focus of this chapter is on comparing conditions for legitimacy across various types of evaluation settings. In the emerging literature on evaluation practices, there has been too little comparative work on meaning making and evaluation, leaving a gap in our understanding of variations in evaluation processes across national, organizational, and scholarly contexts (for a comparison of evaluation, dependency, and risk, see Whitley 1984). It is our aim to start filling this gap. We expect evaluative practices qua practices to vary to some extent across settings. This is so for at least two reasons. First, there are discipline-specific practices that researchers are socialized into early on (e.g., as we will see, a mode of evaluation that appears to be modeled more on deliberative democracy for humanists/social scientists versus one that is modeled on a "court of law" for natural scientists). Second, practices emerge from the dynamics and exigencies of particular intersubjective contexts (e.g., whether a panel happens to be uni- or multidisciplinary). In other words, definitions of scientific worth are interaction and context dependent. This is the case even if these practices are (1) grounded in connoisseurship, expertise, and knowledge that are largely stabilized (i.e., no longer controversial) and (2) part of much broader academic evaluation cultures that are institutionalized (to a varying extent at the national and international levels).

The analysis is based on two parallel but interconnected empirical studies, conducted in the United States and in Finland. In the United States, we

---

1. The conditions for their belief in the fairness of their evaluation include but are not limited to the customary rules of evaluation. Other factors include the process of selection and recruitment of evaluators, whether funding organizations construe this selection as determined by the expertise and status of reviewers, the habitual participation of the latter in elite academic circles, their academic training and professional status, etc. The Social Science and Humanities Research Council of Canada has gathered data on academics who do not typically apply for research funds or serve on selection committees. These data will soon allow social scientists to better understand how such academics preserve the legitimacy of evaluative mechanisms.
studied five different multidisciplinary funding panels in the social sciences and the humanities. This study was concerned, not with differences across panels, but with documenting the customary rules that panelists use.2 We also studied four panels organized by the Academy of Finland. From the outset this study was explicitly concerned with the effects of the mix of expertise on panels and on how customary rules were enacted. The idea was to compare panels with varying degrees of specialization (unidisciplinary panels and multidisciplinary panels) and with different kinds of expertise (specialist experts and generalists). However, in the course of comparing results from the two studies, other points of comparison beyond expert composition emerged: whether panelists rate or rank proposals, have an advisory or decisional role, come from the social sciences and humanities fields or from more scientific fields, and so on. Our exploratory analysis points to important similarities and differences in the internal dynamics of evaluative practices that have gone unnoticed to date and that shed light on how evaluative settings enable and constrain various types of evaluative conventions. Note that we are not concerned with national differences in cultures of evaluation (e.g., with contrasting how the American and Finnish “cultures” would enable and constrain different customary rules of evaluation). This topic will be taken up in future research.

In the United States, we studied five different multidisciplinary funding competitions: those of the Social Science Research Council (SSRC), the American Council of Learned Societies (ACLS), the Woodrow Wilson National Fellowship Foundation (WWNFF), a Society of Fellows at a top research university, and an anonymous foundation in the social sciences. As is often the case in American grant peer review (except in federal agencies such as the National Institute of Health and the National Science Foundation), evaluators involved in these competitions ranked proposals and made funding decisions (although in some cases their decisions had to be formally approved by the research board of the funding institution). In Finland, we studied four evaluation panels organized by the Academy of Finland. We considered panels in the areas of the social sciences; environment and society; environmental sciences; and environmental ecology. Unlike the American competitions under consideration, the Finnish evaluation procedure decouples peer review and funding decisions. A similar procedure has been adopted by a consortium of fifteen national research councils that have joined forces to fund research in the social sciences in Europe (see www.norface.org). It includes countries with large and smaller research communities, such as France, Germany, and the United Kingdom, on the one hand, and Austria, Denmark, Iceland, and Slovenia, on the other. Thus, our comparison contrasts two significant and widely used models of peer review.

Among the most salient customary rules of evaluation, deferring to expertise and respecting disciplinary sovereignty manifest themselves differently based on the degree of specialization of panel members: we find that there is less deference in unidisciplinary panels where the specialties of panelists more often overlap. There is also less respect of disciplinary sovereignty in panels concerned with topics such as “the environment and society” that are of interest to wider audiences. And there is more explicit reference to the role of intuition in grounding decision making in less specialized panels. While there is a rule against the conspicuous display of alliances across all panels, strategic voting and so-called “horse trading” appear to be less frequent in panels that “rate,” as opposed to “rank,” proposals and in those that have an advisory, as opposed to a decisional, role. Moreover, the customary rules of methodological pluralism and cognitive contextualism (evaluating proposals according to the standards of the discipline of the applicant) are more salient in the humanities and social science panels than they are in the pure and applied science panels, where disciplinary identities may be unified around the notion of scientific consensus, including the definition of shared indicators of quality. Finally, while the analogy of democratic deliberation appears to describe well the work of the American panels, the Finnish science panels may be best described as functioning as a court of justice, where panel members present a case to a jury.

Our argument unfolds through a description of the customary rules followed by panelists in the United States and Finland. These are intersubjective rules that guide panel deliberations without being formally spelled out. Panelists cannot always articulate these rules, as they often take them for granted. However, they make them apparent when they describe the appropriate and inappropriate behaviors of fellow panelists, as their praise and criticisms articulate the rules that are to be respected. Academics are never formally taught these rules but learn them throughout their professional socialization, so that these rules inform how they go about shaping and presenting their work. Interviews suggest that by adhering to such rules.

2. Not focusing on organizations themselves was a condition for gaining access to these panels.

3. Our analysis is inspired by the ethnographic approach to examining the rules of social order. See in particular Garfinkel 1967.
evaluators are able to bridge their epistemological differences and perform the task of evaluating while maintaining their belief that their evaluation is legitimate (Mallard, Lamont, and Guetzkow 2009). Customary rules are thus the social conditions that lead panelists to build consensus with other evaluators and to perceive the process as fair.

In our analysis, the term "fairness" refers to the collective outcome of following the rules—that is, to the shared belief among panelists that meritocracy guides the process, while corrupting forces, self-interest, and, in particular, politics are kept at bay. Fairness is crucial for producing legitimacy in peer evaluations (Lamont 2009). As argued elsewhere (Lamont, Mallard, and Guetzkow 2006), the rules for legitimacy that we have identified are in line with Max Weber’s analysis of the role of expertise in providing legal-rational legitimacy (Weber [1956] 1978, 99–100)—it requires the use of impersonal, abstract, and consistent rules.

Methods

In the United States, we studied five different multidisciplinary funding competitions over a two-year period, for a total of twelve panels. Each of the American competitions under consideration has a different focus: the CRDF competition funds social science dissertations that require qualitative research abroad; the ACLS competition targets all the humanities as well as the interpretive social sciences and funds faculty members at all ranks; the WWNFF competition supports graduate dissertations in gender studies; the Society of Fellows supports postdoctoral scholars across the social sciences and the humanities; and the anonymous foundation funds various types of social science research. Panelists for these competitions evaluate and rank proposals and make collective funding decisions.

We conducted a total of eighty-one open-ended, semistructured interviews with individuals involved in the final deliberations of these competitions, including fifteen interviews with program officers. The interviewees selected originated from a wide range of disciplines—anthropology, economics, English, history, philosophy, political science, sociology, and so forth—reflecting the multidisciplinary nature of the competitions under study. These panelists typically teach at American research universities and do not know one another prior to deliberations.

Again, we also studied four evaluation panels organized by the Academy of Finland. Their evaluation procedure is as follows. Program officers put together expert panels organized loosely around disciplines or themes, defined by the range of submitted applications in a given round. International experts mostly from Europe are invited to serve. They meet in person to collectively rate proposals (from 1 to 5) and write consensus reviews of them. Each panel focuses on a subset of proposals concerned with a topic or research area. Before the meeting, each proposal is reviewed by two panel members who draft reviews and are charged with presenting the case to the rest of the panel. All the members of the panel can consult these preliminary reviews, as well as the applications, online prior to the deliberations. After the meeting, the collective ratings and written evaluations produced by the panel are forwarded to one of the four Research Councils of the Academy of Finland, which make funding decisions for all proposals under consideration. The Research Councils are composed of Finnish scholars who are nominated for a three-year period by the Finnish government. Our study was not concerned with how the councils make funding decisions but focused solely on the workings of evaluation panels.

Data collection involved phone interviews with eighteen (out of twenty-seven) panel members who served on the selected four panels. We also conducted face-to-face interviews with ten funding officers. The selection of panels was determined in conversation with the academy, with a view to including panels with varying degrees of specialization in the panels focus we are familiar with. The Social Sciences panel was multidisciplinary, considering proposals from sociology, social psychology, social policy, social theory, social work, and cultural studies. It was composed of experts from these various fields. The Environment and Society panel was also multidisciplinary, but differently so: the panelists were not specialists in one discipline but often had degrees in multiple disciplines and were knowledgeable about a wide range of interdisciplinary topics. They typically considered interdisciplinary proposals that dealt with environmental issues or with social-environmental interactions from a social, political, economic, technological, or other perspective outside the sphere of the natural sciences. The Environmental Sciences panel evaluated proposals that dealt with natural processes in various environments, including forests, soils, peatlands, and vegetation. This panel was also multidisciplinary, since both the proposals and the experts spanned across fields. The Environmental-Ecology panel was unidisciplinary. It operated with a thematically and epistemologically coherent set of proposals emanating from one field, the ecology of aquatic environments. All the panelists were ecologists of some sort. Table 6.1 illustrates how the Finnish panels differ from each other and from the American panels (which all have a similar composition).
### Table 6.1 Comparison of the evaluative settings of the American and the Finnish panels

<table>
<thead>
<tr>
<th>American panels</th>
<th>Social Sciences</th>
<th>Environment and Society</th>
<th>Environmental Sciences</th>
<th>Environmental Ecology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specialization of the panel</td>
<td>Multidisciplinary</td>
<td>Multidisciplinary</td>
<td>Multidisciplinary</td>
<td>Multidisciplinary</td>
</tr>
<tr>
<td>Specialization of experts</td>
<td>Experts</td>
<td>Experts</td>
<td>Generalists* and interdisciplinary experts</td>
<td>Experts</td>
</tr>
<tr>
<td>Technology of evaluation</td>
<td>Ranking</td>
<td>Rating</td>
<td>Rating</td>
<td>Rating</td>
</tr>
<tr>
<td>Role in decision-making process</td>
<td>Distribute awards</td>
<td>Inform award-making process</td>
<td>Inform award-making process</td>
<td>Inform award-making process</td>
</tr>
<tr>
<td>Field of research</td>
<td>Social sciences and humanities</td>
<td>Social sciences</td>
<td>Interdisciplinary</td>
<td>Applied science</td>
</tr>
<tr>
<td></td>
<td>Pure science</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Our term "generalist" refers to a person whose expertise is characterized by broad knowledge beyond any one academic field. While we sometimes use the term "expert" as the opposite of "generalist," we acknowledge that a "generalist" is a kind of expert, too. Generalists have special expertise precisely because they have a broader view than most other experts.

---

In both studies, we interviewed journalists shortly after the panel deliberations. During the interviews, journalists stated that while they made about a range of proposals, we asked them to describe the arguments they made about a range of proposals. We asked them to describe the arguments they made about a range of proposals. We asked them to describe the arguments they made about a range of proposals. We asked them to describe the arguments they made about a range of proposals.
terpretive frames and attempt to convince one another that theirs is the most adequate. It is this context that gives rise to "deferring to expertise," a foundational rule for sustaining collective belief in the fairness of peer review.

When panelists want to advocate a position regarding a proposal, they invest energy in staking their rightful claim to evaluate it. That is, they mark their territory. In other cases, they draw on previously established proofs of competence. In this context panelists give more weight to the opinion of experts; this is essential when panelists are comparing proposals that speak to a wide range of unfamiliar topics. A historian noted that a proposal "looks good until somebody says there's a whole literature that you cannot reasonably be expected to know." Particularly when listening to someone who "comes in extremely expert and careful and [is] a person I respect a lot," this historian finds it prudent to defer. "If this expert says, '...this is really a fair bal proposal,' then I just sort of say that must be true."

The most common form of deference involves what we call the customary rule of "respecting disciplinary sovereignty." Panelists' opinions generally are accorded more weight with regard to proposals emanating from "their" field. Violating this rule creates major conflicts, as panelists often feel slighted if their competence is not respected.

Unsurprisingly, we find that this rule is more widely respected in the American panels, which are all multidisciplinary, than in the one Finnish panel that can be defined as unidisciplinary, the Environmental Ecology panel. Overlapping expertise makes it more difficult for any one panelist to convince others of the value of a proposal when opinions differ; insisting on sovereignty would result in intense conflict for scientific authority. While distance lends authority to the view of others, the toughest rivals are those who are closest—this was anticipated by Emile Durkheim, who, in The Division of Labor in Society, pointed out that "the closer the functions are to one another, the more points of contact are between them, and as a result, the more they are in conflict." And indeed, in the Environmental Ecology panel we observed that panelists working within the same field or on closely related topics tended to have the strongest disagreements.

This is acknowledged by the panelists themselves. As one of them put it: "When it was clear that the first person was a real expert on this particular field which the second person hasn't known, obviously they deferred to the first person's opinion.... But it changes the climate of the discussion if you're both huge experts on this field; then you can argue about it."

The rule of deferring to expertise also appears to play out differently depending on the substantive issues with which the panel is concerned. Indeed, we found that less weight was put on specialized expertise and more was put on general arguments having to do with the common good in the panel Environment and Society, which concerns the social aspects of environmental changes—a topic that is broadly debated by the wider public, the media, and activists who claim the right of nonexperts to participate in decision making about issues that affect their lives. In the eyes of one evaluator serving on this panel, the combination of expert opinion and broader considerations is essential for reaching optimal decisions:

"I think you need to have the experts in the field to comment, particularly if there's a proper methodology and if there's a proper question, because only they really know the literature. However, I do think we need a wider group to ask bigger questions, like: Is this particular research of sufficient interest for public funding? Also I think often nonexperts can ask sort of idiot questions like "Why do you do this?" which can often be a shock to a specialist."

We also find that the relevance of the customary rule of deferring to expertise varies with the co-presence of generalists and experts on a panel. This was particularly evident again in the Finnish panel Environment and Society, which had a mixed membership. Individuals serving on this panel tended to consider all proposals "fair game" and viewed them as located in a disciplinary no-man's-land. Since the legitimacy of the process was not entirely based on the use of specialized expertise, consensus was more often reached through mutual learning, compromising, or simply by relying on panelists' integrity or intuition. As one of the panelists described the situation: "You could put your hands on your heart and then say to each other, 'Do you really, honestly, think that this is a "good" proposal, or an "excellent" proposal? What do you think, really?' The persuasiveness of a colleague was often enough to convince other panel members, even in the absence of expertise or warranted arguments.

In the two Finnish panels that were composed of experts from different disciplines (the Environmental Sciences and the Social Sciences), the

---

4. To continue the quotation: "The magistrate is never in competition with the industrialist. But the brewer and the winegrower, the draper and the maker of sails, the poet and the musician often attempt to mutually supply each other" (Durkheim [1893] 1984, 267).

5. See Langfeldt 2004 for similar findings.

6. On scientific and lay expertise, see Collins and Evans 2007. On this topic, see also Stark 2006.

7. On the complementarity of expert and noneexpert opinion, see Collins and Evans 2007.
research council that made the funding decisions based on the ratings. In this latter case, the directions from the funding agency to panelists explicitly encourage them to evaluate the intrinsic strengths and weaknesses of each proposal instead of comparing and ranking. Strategic voting does not play as much of a role in such a context since each proposal under consideration could, in principle, receive the highest score. Because the proposals are not explicitly pitted against one another, there is less of a sense that one should engage in quid pro quo to ensure that one's favorite will "win."

However, there is evidence that evaluative contexts in which proposals are rated rather than ranked do not necessarily discourage strategic behavior. Some panelists were skeptical about the value of "abstract" (as opposed to comparative) rating since in any case only top-ranked proposals have a chance of receiving funding. Whereas some perceived the process as fair because they judge the proposals on their own merit, others viewed the process as "meaningless." As one of the most critical panelists explained:

> There's a problem that we are not ranking the proposals, although we know the proposals very well. . . . If you [analyze] our grades, it will be kind of normal distribution: there are lots of number three, which are useless, and very few trees and tours. I guess, and I don’t think we assigned any ones, and just a few twos. So the Finnish committee that will take over after us, they are not very much helped by our statements or grading. . . . They will have to do everything again by themselves, and do the ranking by their own criteria. I think this is kind of meaningless.

Critical voices tacitly advocated in favor of more strategic behavior, including comparative ranking, but they were deterred by the explicit instructions of the funding agency. This illustrates how the evaluative technique imposed by the funding agency influences the behavior of panelists. However, it does not determine it entirely, as the evaluators are also guided by the evaluative practices that they have developed elsewhere. The peer review culture that is part of the larger academic world is also likely to influence their behavior.

8. Fuller (2002, 237) distinguishes between one-chamber and two-chamber representation of peers by analogy with legislative bodies and notes that each type has its own special functions and problems. However, he presents no empirical observations to elaborate those findings.
Promoting the Principles of Methodological Pluralism and Cognitive Contextualism

Observations of and accounts concerning the multidisciplinary American panels suggest that these are not a forum for challenging other methodological or disciplinary traditions. Panelists abide by the rule of methodological pluralism. They are encouraged to evaluate proposals according to the epistemological and methodological standards that prevail in the discipline of the applicant—and are personally committed to doing so. This principle, which we have dubbed "cognitive contextualization" (Mallard, Lamont, and Guetzkow 2009), was summarized by an evaluator as he described the dynamics of his panel:

[There are] differences between people who work with large data sets and do quantitative research. And then the very polar opposite. I suppose, folks who are doing community-level studies in anthropology. There are such different methodologies that it's hard to say that there's a generalizable standard that applies to both of them. We were all, I think, willing and able to understand the projects in their own terms, fortunately, and not try to impose a more general standard, because it would have been extremely difficult. . . . I wouldn't mind a canadace in political science responsive to what seemed to me to be having overly instrumental or diagrammatic ways of understanding what they're going to do, because they have to have those. They have to have a certain clarity; they have to have a certain precision.

The premium put on "cognitive contextualization" pushes panelists to downplay their personal preferences and to assess proposals through the lenses that are distinctive to the applicant's field.

Maintaining consistency in criteria for judging qualitative and quantitative proposals is crucial to panel legitimacy, and it is complicated by the fact that panelists compare different subsets of proposals (defined by shared topics, comparable relative ranking, or proximity in the alphabet) at different times. The characteristics that are shared by any one batch of proposals vary and may make different criteria of evaluation more salient, as a historian pointed out:

It does sometimes happen that we get some that are very close to each other, and I always go back again and look at the ones that I thought were really the best and really the worst and see if they're really all that much differ-

Comparing Customary Rules of Fairness / 223

ent. It's like working yourself through any batch of applications or papers or whatever your standards kind of evolve as you go through it. I don't sort mechanically. . . . Until I've read the whole batch, I don't even know exactly what the standards are going to be.

While the respect of disciplinary differences is salient in most kinds of panels, the principles of methodological pluralism and cognitive contextualization appear to be most supported by the epistemology of the social sciences and especially that of the humanities. Indeed, a close examination of the Academy of Finland panels reveals that the members of the Social Sciences panel were indistinguishable from their American counterparts with respect to these rules. Their discussion of the appropriate criteria for evaluating some business school proposals illustrates the salience of cognitive contextualization. The panelists noticed that there was no expert on that field among themselves, which made them worry about imposing sociological criteria on those proposals. A sociologist pondered:

Obviously we could use a general social science expertise to evaluate the proposals, but . . . it was quite difficult for us to place them, as it were, academically, because we don't know what the norms and values of the business school kind of proposal might be. So, for instance, from a sociological point of view, we found them lacking in many ways, but it could be that within that kind of business and critical management studies those kinds of proposals are actually great some time, but we didn't have anyone with that exact area of expertise to, kind of, give us the kind of key markers.

In contrast, the more strictly scientific panels appeared to be more committed to using consistent standards for evaluating all proposals, as opposed to adjusting their judgment to what counts as "good work" across fields. This goes hand in hand with an epistemological culture where controversies between what is defined as true and false tend to be less open-ended, as scientific and other types of evidence may more strongly constrain debates and the "blackboxing" process (as described by Latour 1987). Moreover, consensus formation may be more central to the identity (and, possibly, evaluative cultures) of scientific disciplines. This concern for consistency is illustrated by an ecologist who recalled many occasions where the panelists worried about inconsistency: "Sometimes we went back to previous applications and said: 'If we evaluated that in this and this way, then we have to use the same criteria when we are looking at this one' . . . If
we say that a person hasn’t been abroad means that and that, we will have to use the same criteria for another application. I think we tried to be fair.

An important means of producing coherent evaluations among environmental scientists was a harmonization of rating scales. At the start of the meeting, these panelists had discussed “in what journals we would have outstanding, excellent, and very good papers, in this sequence, or only good papers.” Whenever panelists hesitated about giving a 5 (the highest ranking) to a proposal, they reported having discussed: “Can it, if we are lucky, [lead to findings that could] be published in Science or Nature?” Thus, they attempted to agree on shared matrices through indirect indicators of quality, which streamlined evaluation. They perceived consistency as crucial for fairness, since panelists were convinced that scientific quality could best be detected by the use of given criteria. The chair of a panel demonstrated the legitimacy of the evaluation he presided over by arguing: “The grade ‘five’ proposals would have received a similar grading in any of the national or international panels on which I have sat.” The concern for consistency was low in the American and Finnish humanities and social science panels because most experts serving on these panels believed that evaluators played a crucial role in defining the lenses through which quality can be recognized.

The comparison of different panels also reveals that the generalists in the panel of Environment and Society did not share the same concern for cognitive contextualization that we found among more specialized experts in both the American and the Finnish panels. In spite of their social science orientation and background, these panelists prided themselves on their detachment from disciplinary traditions and on their ability to locate seemingly disparate proposals within a broad matrix of evaluation. This may be because they are simultaneously involved in several different epistemic communities, which often requires an ability to see beyond particular criteria and to compare relatively smoothly proposals emanating from a range of disciplines, that is, proposals that could easily be viewed as incommensurable. Thus, they did not argue in favor of methodological pluralism. Instead, they typically favored general criteria of quality that are shared across the social and environmental sciences. As one of them put it:

What we [were] looking [for] was not particularly disciplinary attributes of the applications. We were looking at things like research design, is it going to produce useful results, would the results be useful for policy makers? These sorts of methodology [concerns] is it well-explained and good . . . ? They were more generic questions rather than is it good sociology or good economics or good this or good that? And I think we all really took that view.

The Environment and Society panelists thus encouraged each other to downplay epistemological differences between disciplines and strengthen what was shared in their conceptions of quality. However, this process was sometimes costly and required thorough discussions on methodological questions. This became evident in a series of disagreements between two panel members, whose opinions on several proposals strongly differed. Both were experts in case study methodology, but their theoretical backgrounds diverged. During a private discussion at breakfast, they came to an agreement concerning where their criteria of evaluation could overlap. One panelist explained: “I had not been as critical on [particular methodological choices], because I’ve read [the proposals] in the context that I worked from, and I didn’t have as much problem with these methodological decisions. But I concurred with his concerns when he went through them in some detail.”

The panelists came to an agreement on a new set of similar criteria about how to evaluate case study applications. Such negotiations of meaning were essential to allow panelists to “save face” and sustain the conditions necessary for continuing the work of the panel. At the same time, the emerging understanding between the two parties renewed the panelists’ belief in the legitimacy of their decisions.

Limiting Idiosyncratic Tastes and Self-Reproduction

Rational legitimacy, Weber reminds us, comes from the application of impersonal and consistent rules ([1956] 1978, 212–71). Thus, by trying to bracket their idiosyncratic tastes, panel members help sustain collective belief in the fairness of deliberations. An English professor serving on one of the American funding panels advocated distinguishing between one’s personal preferences and criteria of competence, and privileging the latter when the two are in conflict. In subordinating personal preferences to more neutral standards, this scholar explicitly protects the legitimacy of the process, but he also recognizes the role of individual subjectivities in evaluation. But this panelist is more scrupulous than many. Most reviewers uphold the legitimacy of the process when they seamlessly fold their idiosyncratic preferences and tastes into the formal criteria of evaluation. So, for example, they tend to define originality in ways that are in line with the type of originality that their own work exhibits. As one interviewee acknowledged, evaluators tend to like what speaks to their own interests: “I see scholarly excellence and excitement in this one project on food, possibly because I see resonance with my own life, my own interests, who I am,
and other people clearly don't. And that's always a bit of a problem, that excellence is in some ways what looks most like you."

During interviews with the American panelists, multiple examples of how panelists' idiosyncratic interests shape their votes emerged. Apparently, equating "what looks most like you" with "excellence" is so pervasive as to go unnoticed by some. Moreover, panelists cannot spell out what defines an "interesting" proposal in the abstract, irrespective of the kinds of problems that captivate them personally. Most behave as if they have no alternative but to use their own personal understanding of what constitutes a fascinating problem in order to do the work that is expected of them.

A close examination of the two more scientific panels of the Academy of Finland suggests that natural scientists may be more explicit in their efforts to bracket idiosyncratic tastes and avoid self-reproduction than are the social scientists in the two national settings we studied. This is suggested by evidence revealing how scientists on these panels (Environmental Sciences and Environmental Sociology) attempt to cancel out idiosyncrasies by relying even more on collective judgments. The role of the group was perceived as crucial for "judging the arguments and viewpoints" of individual experts, "trying to find balance in the discussion," and "discussing the general principles."

Whereas we argued at the outset that the American panels in the social sciences and the humanities followed principles analogous to those of deliberative democracies, the more appropriate analogy for the natural science panels may be that of a court of justice. Scientists are more concerned with consistency in evaluation and maintaining impersonal criteria than they are with sustaining full, equal, and free exchange of opinion. Accounts by panelists indicate that the reviewers in charge of presenting a proposal played the role of an expert witness, and the rest of the panel acted as a critical jury. Thus, panelists found it important that experts discussed each proposal "in front of the evaluators." The panel on Environmental Sociology had even set up a routine of assigning the task of drafting each evaluation statement to a panelist who was "a little further removed from the field of the applicant" so that the given proposal was "not so close with his personal emotions." This panelist acted "as a kind of independent judge, [who] could look more at the formal aspects, keep things equal, and judge across different cases."

The belief in the value of calibration was also present among the social scientists, but to a lesser degree and it took a somewhat different form. In the American panels and the two Finnish panels consisting of social scientists (Social Sciences and Environment and Society), the experts acknowledged personal standpoints as inevitable components of evaluation. Rather than trying to cancel out the biases that each panelist brought to the evaluation process, they did so by environmental scientists and ecologists, social scientists tended to believe that free exchange of opinion helped them become aware of their personal mind-sets and made them more open to rethinking their evaluation. A Finnish social scientist explained this social process as follows:

The panel would have to be explicit about how it understood the criteria, in relation to the application, and those discussions would be explicit and substantive. One could then detect different perspectives around the criteria. . . . I think where positions were very different, I would say, "This is my take on it, this is how I see it, but"—you know—"okay, having heard what you said, and looked at some of the other applications, where we had some similar discussions, I can see that I was possibly underestimating the importance of x, y, and z."

This quotation indicates that relativism in judgment, or awareness of how worldviews affect evaluation, is part of the social science culture of evaluation. More comparative data will be needed before we can fully ascertain whether and how scientists understand the place of tastes and "individual perspectives" in evaluation.

Conclusion

This chapter has discussed the customary rules of evaluation that panelists typically follow in making decisions and analyzed the specific applicability of those rules to panels that work in various evaluation settings. We consider our customary rules to be part of "epistemic cultures" and essential to the process of collective attribution of significance. In this context, considering reasons offered for disagreement, how disagreements are negotiated, and how panelists interpret agreement is essential to capturing fairness as a collective accomplishment.

Our interpretive analysis posits that evaluative practices are shaped and constrained by the context in which they occur, including intersubjective agreements concerning the conditions for fair and optimal evaluation. Instead of contrasting "biased" and "unbiased" evaluation, we examine how panelists constitute the evaluation process, including the role played by intersubjectivity in assessment (Lamont 2009). Contrary to what is suggested by the classical approach to peer evaluation (e.g., Cole and Cole...
Table 6.2  Conditions affecting and modifications of selected customary rules of fairness

<table>
<thead>
<tr>
<th>Customary rules</th>
<th>Conditions favorable</th>
<th>Conditions unfavorable</th>
<th>Modifications of the rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deferring to expertise and respecting disciplinary sovereignty</td>
<td>- Panels composed of experts from multiple fields</td>
<td>- Unidisciplinary panels</td>
<td>- Argumentation of alternative perspectives</td>
</tr>
<tr>
<td>Pragmatic use of alliances and strategic vetoing</td>
<td>- Panels that rank proposals and make final decisions</td>
<td>- Panels composed of generalists</td>
<td>- Taking account of both expert opinion and broader considerations</td>
</tr>
<tr>
<td>Methodological pluralism and cognitive complementation</td>
<td>- Social sciences and humanities panels composed of experts</td>
<td>- Panels that rate proposals and have an advisory role</td>
<td>- All proposals are &quot;fair game&quot;: use of mutual learning and intuition</td>
</tr>
<tr>
<td>Limiting idiosyncratic tastes and self-reproduction</td>
<td>- Social sciences and humanities panels</td>
<td>- Natural sciences panels</td>
<td>- Evaluating the intrinsic strengths and weaknesses of each proposal</td>
</tr>
<tr>
<td></td>
<td>- Natural sciences panels</td>
<td>- Panels composed of generalists</td>
<td>- Concern for consistency of criteria across cases</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Use of a general matrix for comparison to assess &quot;incomparable&quot; proposals</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Democratic deliberation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Court of law</td>
</tr>
</tbody>
</table>
(including faith in the general legitimacy of the system; for an analysis of the evaluation crisis in the French higher-education system, see Cousin and Lamont 2009).

While in this chapter we have stressed differences between the humanities and social sciences, on the one hand, and the sciences, on the other, there exist parallels that should be examined more closely. Calibration may be valued across all fields as a way to limit differences in standards due to professional affiliations and other factors. Also, while the natural scientists did not promote methodological pluralism or cognitive contextualism as such, they clearly avoided challenging each others’ standpoints explicitly. Compromises were thus created through a balance between competing criteria or by drawing on a majority opinion, rather than by imposing particular standards. We may explore whether, in fact, evaluative practices across fields have started converging. It is quite possible that in the context of an increasingly present audit culture in higher education and research (Strathern 2000), evaluative practices are becoming standardized and widely institutionalized (nationally and internationally) while disciplinary differences are declining to various extents (with economics leading the pack and the interpretive social sciences defending national distinctiveness). This in itself defines an important path for future research on the evaluative cultures of the social sciences.

Our study opens a new window through which to look at some contested effects in the peer review of research proposals. Numerous scholars have pointed out both potential and observed risks in the peer review system. It is argued that the system is conservative and suppresses innovative research. Effects such as nepotism and old-boyism in peer review are seen to hinder pioneering research (Chubin and Hackett 1990; Roy 1985), while "cognitive particularism" and "favoritism for the familiar" function to support the kind of research the reviewers themselves are conducting (see, e.g., Porter and Rossi 1985; Travis and Collins 1991). At the same time, scholars working with the organization of evaluation panels have found that evaluation by groups is less problematic than evaluation by individuals, while other flaws may arise due to group dynamics (Grigg 1999; Langfeldt 2001, 2004, Laudel 2006). Our findings on customary rules point in the same direction. But they also suggest that some of the perverse effects of peer review, such as cronyism, the pursuit of self-interest, and cognitive particularism, may be influenced by the way panels are set up. Much more work is needed on this topic before we can reach definite conclusions.

It is sometimes claimed that funding officers can manipulate the peer review system to deliver the recommendations they prefer by shrewdly choosing reviewers (Roy 1985). These claims do not pay attention to the complexities that the social dimension brings about in evaluation panels. However, it is an interesting question whether a proper understanding of the impact of various social conditions on the workings of panels would improve the evaluation process. While specific decisions are hardly predictable, it is possible that a better understanding of the impact of various types of set-ups would lead program officers to put in place optimal processes of deliberation. We believe that the question should be of interest for policy makers and for the larger academic community. Consensual standards in academic evaluation may influence what kind of research gets supported and may thus have long-term consequences for the cognitive development of the social sciences.

References


Lamont, Michèle, Grégoire Mallard, and Joshua Goetzkows. 2006. "Beyond Blind Faith:
Meetings by the Minute(s)
How Documents Create Decisions for Institutional Review Boards

LAURA STARK

Knowledge Evaluation as Statecraft

Institutional review boards (IRBs) regulate interactions between researchers and the people they study if researchers work for an organization that gets money from the U.S. federal government. Today, most universities, hospitals, and scholarly institutes do get public money in some form. Because review research is required at the sites of knowledge production, IRBs end up overseeing researchers who earn their pay not only as biomedical scientists but also as social scientists and humanists in the United States and, increasingly, abroad (Hedgcoe 2010; Heimer and Petty 2010). This has been the case since 1966, but IRBs have been more fastidious and conservative in enforcing regulations that attempt to protect human subjects since 2000. As a result, more social research has been pulled into regulatory purview (Bledsoe et al. 2007; Schrag 2009; Stark 2007).

The ideas in this chapter are part of a larger study of IRBs as examples of what I call "declarative bodies," expert groups that are empowered by governments to make decisions without consulting citizens, for example. Through public referenda (Stark, forthcoming). I am interested in how, through the process of evaluation, experts who make up declarative bodies—such as IRBs, data- and safety-monitoring boards, funding panels, editorial boards, and film-rating committees—have a hand in creating the products of art and science that they are reviewing. Part of the work of declarative bodies involves creating and then sustaining the image that members have reached a legitimate decision. Bureaucratic documents are central to this task, and I open up this phenomenon in the current chapter.