Peer Evaluation in the Social Sciences and the Humanities Compared:
The United States, the United Kingdom, and France

Report prepared for the
Social Sciences and Humanities Research Council of Canada

March 24, 2005

Michèle Lamont
Department of Sociology, Harvard University
William James Hall, 33 Kirkland Street, Room 510
Cambridge, MA 02138
mlamont@wjh.harvard.edu

Grégoire Mallard
Department of Sociology, Princeton University
Wallace Hall, Princeton NJ 08540
gmallard@princeton.edu
This study reviews the literature on peer evaluation in the Social Sciences and the Humanities (SSH) in the United States, the United Kingdom, and France. While the literature has not yet attained sufficient maturity to provide robust prescriptions, this paper identifies different findings that can help the SSHRC in its efforts to help knowledge producers reach broader publics. It adopts a critical stance against authors for whom evaluations based on the quality of the research—as opposed to the social characteristics of the researcher, like gender, race or institutional affiliation—are automatically fair, while not specifying how quality is defined. In contrast, we contend that biases in evaluation against certain categories of people can enter the process of evaluation through the favoring of epistemological styles and other cognitive categories, which might serve as proxies for social characteristics. Some studies support this idea, but their results remain too impressionistic, and more research is needed to assess whether these forms of discrimination are found systematically in peer evaluation. Social-blind procedures of evaluation should be critically assessed. The review then moves to questions about using measures of past evaluations to guide the funding of future research. It discusses the debates about the use of measures of researchers’ impact by different research councils, which reinforces the centrality of peer evaluation in academic life. By aggregating past peer decisions to fund, publish, and cite someone else’s research, these indicators strengthen the weight of past evaluations on the decision to fund future research. They highlight the total output of past evaluations while black-boxing the process by which these decisions were made. This section shows what pitfalls the Council should avoid when conducting efforts to broaden the impact of research in the SSH. We recommend that the Council give serious thought to what makes the vitality and impact of a field on society. Then, the review discusses ways by which the Council could engage the complex ecosystem of academia to better mobilize everyone’s specific resources. Rather than impose one model of scholarship for all scholars, we argue that the Council should identify the strengths characteristic of different types of scholars, and should encourage initiatives that build on these strengths.
Peer Evaluation in the Social Sciences and the Humanities Compared: The United States, The United Kingdom, and France

1. INTRODUCTION

Peer evaluation in the humanities and the social sciences has received scant analytic attention despite its centrality to research and higher education. Students of peer evaluation are motivated to study the topic for two opposite reasons. For some authors, most of whom are North American, understanding evaluation is necessary to ensure the fairness and efficiency of the process. They take the merits of peer evaluation for granted and aim to perfect it. For others, peer review represents an outdated model of scientific evaluation that artificially separates knowledge from its context of production and use (Gibbons et al. 1994). For these authors, most of whom are Europeans, peer evaluation might be an obstacle to a more democratic science. They believe that a wider range of actors should participate in the evaluation process: scientific administrators, politicians, business managers, and representatives of professional associations (Callon, Lascoumes and Barthe 2001). They believe that scientists, like other actors, should be accountable (Power 1997).

The Canadian Social Science and Humanities Research Council is well aware that a delicate balance must be found between these two alternative views of peer evaluation (SSHRC 2005). Not unlike American agencies funding research in the social sciences, it has become particularly concerned with the intellectual and social impact of the research it funds. Indeed, in the United States, agencies such as the National Institutes of Health have been asked to develop performance measures since the passage of the 1992 Government Performance and Result Acts (GPRA) and the establishment of R&D investment criteria by the White House’s Office of Management and Budget. This new concern has favored the use of quantitative indexes of research performance (primarily publications) and impact (primarily citation counts), which makes it possible to create cost-benefit ratios (financial inputs/knowledge created) for all scientific research, including the humanities and the social sciences.¹ But the implementation of such

¹ This concern for impact is relatively new: between the start of the Cold War and the fall of the Berlin Wall, research and development activities were not subjected to studies of impact and performance (Godin
measures of impact might reinforce the biases that exist in peer evaluation (Archambault and Gagné 2004). Bibliometric measures of productivity and impact are usually aggregates of past decisions to publish and to fund research. If they are used indiscriminately, these measures can strengthen biases inherent to peer evaluation: they can favor men, who tend to be cited more frequently than women. They also favor researchers working in elite research universities, benefit those working within well-established paradigms, and strengthen the status quo, as defined by reviewers and panelists involved in the peer review process. Finally, they can also benefit the impact of research on the progress of knowledge and downplay its other uses. Thus, it is particularly important to better understand how the production and impact of new knowledge is evaluated. We will focus on peer evaluation in the context of research funding. On this topic, the literature produced in the United States offers the most complete and useful answers for the Council.

The Council has been particularly concerned with changing the relationship between scientists and their publics, as demonstrated by the Council’s initiatives to foster new modes of collaboration between researchers and users of scientific research (SSHRC 2005). It is committed to going beyond broad diffusion of pure scientific research in society by rewarding researchers whose work is likely to have the greatest impact. This entails redefining the process of peer review away from the previous model, where peers are the only legitimate evaluators of knowledge. On this topic, the literature on peer evaluation that has emerged from the UK and France can offer the most useful answers concerning how to accomplish this goal.

To help the Council in its new orientation, this review identifies a) different conceptual lines of reasoning at work in the academic literature on peer evaluation; b) key results that illuminate obstacles inherent to the peer review system that might slow down the transition from a Granting Council (concerned with financial inputs for research) to a Knowledge Council (concerned with a broad range of outputs of research).
Based on this review, we explore strategies that could help the Council achieve its goals. Our international comparison of the literature is limited to three countries: the United States, the United Kingdom, and France. It shows that the relationship between academic professions, funding agencies, and the university system varies across national contexts: in the United States, we identify a professional model of evaluation, given that discussions about the fairness of evaluation have emanated from professional associations working together with Congress. In the UK, we find a managerial model of evaluation, as reforms concerning evaluation emanate solely from the government and were opposed by academics. In France, we find what we call a post-corporatist model of evaluation given the weak control of professional associations (and professionals) on scholarly evaluation. These models point to trends within each national context, as opposed to clear-cut differences. Understanding the relationship between these institutions—professions, government, and universities—in the Canadian context will be crucial for the Council’s capacity to anticipate external factors that may affect its ability to implement change. Therefore, we not only present the research on peer review that is of immediate relevance to our brief, but also describe external factors pertaining to the institutional context that are most likely to affect the implementation of the Council’s policies.

We should note at the outset the paucity of research on evaluation in the social sciences and the humanities (SSH), which is the main object of this review. The literature on academic evaluation is almost exclusively concerned with the natural and physical sciences. Institutional factors explain this trend. Research on scientific evaluation has generally been initiated in response to demands for greater transparency and accountability of the use of public funds by professional associations and/or governments.² Since the research funds directed to the natural and physical sciences far surpass those dedicated to SSH, evaluation processes in the former fields were of greatest concern. It is only with what we call the “second wave of evaluation studies” that authors became preoccupied with the diversity of “epistemic cultures” (Knorr-Cetina 1999)

---

² Understanding how the evaluation of research projects was conducted became an important concern in the United States and Europe at the end of the 1970s, due to a growing economic crisis and a stronger concern for the accountability of public institutions in the context of a stronger neo-liberalism.
across the disciplines and began analyzing the cognitive dimension of evaluation and its singularity among and between the various disciplines in the SSH.

This review of the literature proceeds as follows. It begins with a discussion of the literature on evaluation in the United States, and as requested by the Council, of the published and unpublished results from the study conducted by Michèle Lamont in collaboration with Grégoire Mallard and Joshua Guetzkow. Then, it turns to the literature on evaluation in the United Kingdom and in France. For the United States and the United Kingdom, we differentiate between two waves of research. The first wave, conducted between the late 1970s and late 1980s, assessed the extent to which indicators (of quality, output, impact, utility, or other factors) influenced evaluation by experts, as compared to other extra-intellectual, and less legitimate, factors. These studies aimed at increasing the accountability of public institutions and at determining whether they favored specific groups (e.g., researchers that have received their Ph.D. from elite institutions). Whereas American research on the topic strengthened the legitimacy of peer evaluation, British researchers concluded that peer evaluation should be better monitored and controlled by society. Compared to these two literatures, French research was not concerned with increased transparency. Instead, it focused on describing the diversity of local cultures in the academic field. Moreover, given that research on the topic remained focused on the same themes after the 1980s, we are hard-pressed to find evidence of a second wave of research on peer review in the French literature on peer review.

2. THE PROFESSIONAL MODEL OF EVALUATION IN THE UNITED STATES

The American literature on scientific evaluation can be easily divided chronologically and thematically (see Braxton and Hargens 1996 for a review). The first wave of studies on peer evaluation has been overwhelmingly concerned with the roadblocks to “fairness caused by nonscientific influences such as politics, friendship networks, or common institutional positions,” and this, even though “many scientists are most concerned about the cognitive process of evaluation” (Travis and Collins 1991:

---

3 This research has led and will lead to diverse publications, first and foremost, a book (Lamont forthcoming), and also different articles (Guetzkow, Lamont, Mallard 2004; Mallard, Lamont, Guetzkow Under Review), and notes in different newsletters (Lamont 2004).
Highly restricted access to accounts of collective deliberations, difficulties inherent in measuring meanings, and concerns with uncovering the impact of personal connections at work in the evaluation of science account for the fact that cognitive aspects of peer review have received so little attention to date (Mitroff and Chubin 1979). But as a reaction, a second wave of studies has emerged that focuses on the cognitive dimension of peer evaluation.

Institutional context helps explain why the first wave of American research on evaluation concerned fairness. The main study on evaluation by Cole, Rubin, and Cole (1978) analyzed the influence of social and network factors on who received funding. It came in response to the growing concern expressed by Congress and large professional associations that the spending of public money unfairly advantaged certain universities (hence certain states) and certain types of scholars. Cole, Rubin, and Cole’s (1978) study of the allocation of the National Science Foundation’s (NSF) funds was commissioned in response to growing demands for greater transparency from congress and the academic community. Hence, the concern with fairness was directly tied to pressures coming from elected officials and autonomous professional associations. As we will see, these groups did not play a parallel role in the UK and France, with the result that questions of fairness were largely absent from the first wave of research on evaluation that emerged in the 1980s.

A) The First Wave of Research on Evaluation in the United States

We will now detail the results from the first wave that are of relevance for the Council’s objectives. The literature posited scientific norms about fairness, and deviance from these norms in practice, primarily in the natural sciences (Armstrong 1999; G.A.O. 1994; Chubin and Hackett 1990; Bakanic, McPhail and Simon 1987; Roy 1985; Liebert 1982; Cole and Cole 1981; Cole, Rubin and Cole 1978; Cole 1978; Zuckerman and Merton 1971). Of prime concern was the norm of universalism (defined in opposition to particularism), which, along with the norms of disinterestedness, communalism of results, and organized scepticism, were intended to ensure that no one be excluded from scientific debates due to purely subjective and social factors (Merton 1973[1942]). This normative assumption appealed intuitively to the notion that funding (or publishing) decisions in
scientific fields should be based on the evaluation of the research projects (or the research results), independent of the social characteristics of the researchers. “Subjectivism,” cooptation, and in-group favoritism stood in opposition to open scientific debate, free inquiry, and unbiased discussion of results.

The empirical literature emanating from the first wave of research largely confirmed these normative assumptions. Studies found that reviewers follow universalistic norms more often than not, meaning that decisions are not based on affiliation with a common network or deference to status but on a real appreciation of the scientific quality of proposals (Merton 1996; G.A.O. 1994; Cole and Cole 1981; Cole, Rubin and Cole 1978; Cole 1978; Zuckerman and Merton 1971). Moreover, research revealed the absence of a linear relationship between success in obtaining funding and the social characteristics of the principal investigator (e.g., his/her rank; rank of his/her department; rank of his/her Ph.D. program; and having recently received NSF funding). Only the “quality” of the principal investigators’ past work (measured by an index combining the number of their publications with the number of citations to these publications) had predictive value. Hence Cole (1978) concluded his study of peer evaluation on an optimistic note. For him, the successful proposals are those that are of the highest “quality.” A more recent study of the peer review process at the National Science Foundation, the National Endowment for the Humanities, and the National Institutes of Health supported Cole’s conclusion, but it also stated that “the intrinsic qualities of a proposal (such as the research design and the importance of the question it addressed) were important factors in reviewers’ scoring” (GAO 1994: 2). Sonnert (1995) found that productivity (measured in terms of annual number of publications) explains 40 percent of the variance of final scores received by proposals in the natural sciences. However, he also found that the prestige of Ph.D.-granting institutions mattered.5

4 This result applies for funding decisions where applications are not considered anonymous, as well as for reviewing decisions, which follow the usual standards of double-blind review.

5 The prestige of the Ph.D.-granting institution, sole-authorship contributions, and annual publication record together explain 60 percent of the variance in final rankings (Sonnert 1995), which was taken to confirm the “Mathew effect” concerning the cumulation of reputational prestige (Merton 1973).
The usefulness of these results for understanding peer evaluation in the SSH might be questioned. Observing that evaluators stress the “quality” of research proposals or articles provides little analytic leverage when many conflicting criteria are used to define quality, as is the case in the social sciences and humanities disciplines. Indeed, we found that epistemological and other cognitive differences between reviewers are an intrinsic feature of peer evaluation in SSH (Mallard, Lamont, Guetzkow 2005). Authors of the first wave are not concerned with such differences. Yet, they write that the particularism and unfairness of an evaluation “can refer a) to scientists with a common view of their fields who will appraise work only by others with similar views; it could refer b) to social networks of friendship […] ; it could refer to c) social styles; that is, those scientists who achieve eminence tend to favor the proposals of others who are similarly situated in the hierarchy of science” (Cole, Rubin, and Cole 1978: 33-34). If evaluators privilege research that presents specific epistemological and methodological characteristics, the competition will be considered unfair (especially if they privilege research styles that resemble their own). To understand this dimension of particularism, one has to pay attention to the competing standards of quality that prevail within and across fields, and by which proposals are assessed.

On this latter subject, past research on peer evaluation has not offered satisfying answers. Experts on peer review have proposed indexes of cognitive consensus. Social scientists have concluded that consensus among reviewers is the exception rather than the norm, but they have not explored precisely what reviewers disagree about (Cole 1978, 1983; Cole, Rubin and Cole 1979; Cole, Cole, and Simon 1981; Hargens 1987; Zuckerman and Merton 1971). This literature has used quantitative ratings of proposal and rejection rates to measure cognitive consensus, as opposed to analyzing qualitatively how evaluators assess proposals with the use of various cognitive standards. Thus, it cannot speak to the extent of diversity or consensus among evaluators, particularly as it concerns the epistemological positions favored by reviewers. On matters of cognitive diversity, first-wave authors could “only speculate that the great bulk of reviewer disagreement observed is probably a result of real and legitimate differences of opinion among experts about what good science is or should be” (Cole, Cole, and Simon 1981: 885). They acknowledge that substantive criteria of evaluation (which remain undefined) are likely to be diverse and not
agreed upon by evaluators, but that peer reviewers routinely produce many decisions by consensus.\textsuperscript{6} We believe that it is imperative to look beyond rating and rejection rates to assess fairness and homophily in evaluation (see Mallard, Lamont and Guetzkow 2005).

To conclude, the first-wave literature legitimized peer evaluation by suggesting that it generally operates according the norm of universalism. Ranking by peers was found to be correlated most highly with research output, i.e., numbers of international publications,\textsuperscript{7} and with impact, measured by numbers of citations in scholarly journals. This literature presumed a peer review process oriented exclusively toward academic outputs, and directed strictly toward disciplinary and scientific publics, as opposed to broader public relevance and non-academic publics. If peer review functions as it is posited by this literature, it is certainly is at odd with the Council’s stated objective to function as a Knowledge Council and foster mechanisms of evaluation of knowledge that go beyond academia. The next section will show results that challenge this understanding of peer review.

\textbf{B) The Second Wave of Research on Evaluation in the United States}

The literature described so far will be of little assistance to the Council as it seeks to understand what substantial criteria reviewers use, and how to create mechanisms for broadening these criteria. In contrast, second-wave authors have focused primarily on the cognitive dimension of peer evaluation. This was the main concern that inspired Michèle Lamont’s study of five SSH fellowship competitions. Along with other authors, she has called for a sharper assessment of the cognitive categories that peer evaluators use when they negotiate their rankings and preferences (Callon 1994; Gumport 2000; Collins and Evans 2002; Guetzkow, Lamont, and Mallard 2004). Thus, Guetzkow, Lamont, and Mallard (2004) criticized the existing research for overlooking the cognitive content of

\textsuperscript{6} See Cole (1992) on economics; Collins (1998) on philosophy; Connell (1997) on sociology; Lustick (1997) on political science; and Novick (1988) and Poovey (1998) on history for studies that emphasize the presence of epistemological and cognitive diversity in the SSH.

\textsuperscript{7} Indicators of output are usually weighted by the impact of the journal in which they are published (Garfield 1979).
“originality” and disciplinary differences in how this crucial criterion is conceived.8 Calls for a better understanding of the cognitive criteria of evaluation are associated with wider concerns for unraveling the diversity of “epistemic cultures” (Knorr-Cetina 1999) across disciplines, and particularly in the SSH. These two concerns (understanding the cognitive content of evaluation and their disciplinary salience) have been operationalized in the studies produced by Lamont and her collaborators (Lamont in progress; Guetzkow, Lamont, and Mallard 2004; Mallard, Lamont, and Guetzkow 2005).

Based on a study of twelve panels sponsored by five funding organizations (the American Council for Learned Societies, the Social Sciences Research Council, the Woodrow Wilson National Fellowship Foundation, an anonymous Society of Fellows, and an anonymous agency funding research in the social sciences), Guetzkow, Lamont and Mallard (2004) first expand the literature by exploring the cognitive contents of “originality” used by panelists in the social sciences and humanities. They show that peer reviewers in the SSH define originality as: using a new approach, a new theory, a new method, or new data, studying a new topic, doing research in an understudied area, and producing new findings. Whereas the literature has not considered the salience of the various dimensions of originality across disciplines, Guetzkow, Lamont, and Mallard (2004: 191) identified important variations: “humanists and historians clearly privilege originality in approach, with humanists also emphasizing originality in the data used [for instance, looking at new manuscripts, letters, or new authors]. For their part, social scientists value most originality in method, but they also have more of an appreciation for diverse types of originality, stressing the use of an original approach, an original theory, or the study of an original topic” (see Appendix A for details).

Whereas the literature tends to equate originality with substantive innovation and to consider the personal attributes of the researcher as irrelevant to evaluation, Guetzkow, Lamont, and Mallard (2004: 191) show that panelists often equate originality with the moral character of the researcher (especially one’s “integrity” and “risk-taking”), and that these moral judgments play a central role in peer evaluation. This result is compatible with previous findings by Lamont, Kaufman, and Moody (2000) that the construction of

---

8 This criticism was directed at authors such as Merton (1973[1942]), Gaston (1973), Hagstrom (1974), Mitroff (1974) and Wessely (1996).
the self and the display of moral character were crucial in the rhetoric of excellence used by applicants to a prestigious fellowship competition. These studies of evaluation show (contra studies of the first wave) that peer evaluators do not refer only to the scientific productivity and impact of research in abstract terms. Rather, peer reviewers combine substantive and extra-intellectual criteria, including criteria pertaining to the broader moral qualities of a researcher. They have competing, and sometimes incompatible, understandings of the types of research and researchers they seek to fund (opposing “narrow-minded” and “ivory-tower” researchers to “public intellectuals” or “engaged scholars” for instance). These findings suggest that peer evaluation may very well accommodate the Council’s agenda of using criteria of evaluation pertaining to social impact. The instructions given to reviewers by the Council can easily tip the balance in favor of specific types of cognitive and moral criteria.

The other paper stemming from the research led by Lamont and her collaborators focuses on epistemological criteria that foster disagreements among peer reviewers (Mallard, Lamont, and Guetzkow 2005). This paper shows that evaluators draw on four epistemological styles to make arguments in favor of and against proposals: a) the constructivist style (used by scholars who seek to give voice to minorities or to embrace their indigenous concepts); b) the comprehensive style (used by scholars who adopt an interpretive or theoretical approach to understand social phenomena); c) the positivist style (used by social scientists who adopt a hypothetico-deductive approach and quantitative techniques to solve an empirical puzzle); and d) the utilitarian style (used by social scientists who adopt hypothetico-deductive and quantitative approaches to analyze social problems and generate policy-oriented knowledge). Among these diverse styles, the comprehensive and constructivist styles dominate in the competitions closest to the humanities and history, whereas the comprehensive and positivist styles dominate in competitions closest to the social sciences (see Appendix A).

Illuminating the epistemological styles valued by panelists complements first-wave studies concerned with levels of consensus as measured by rejection rates. A surprisingly high percentage of the panelists use the comprehensive style to make arguments against and in favor of proposals. Moreover, the preferred styles are less diverse than first-wave studies suggested which contradicts findings concerning the low
level of cognitive consensus in peer review in the natural and social sciences (Cole, Cole, and Simon 1979). Evaluators favor scholarship that addresses purely academic concerns (viz. the overwhelming popularity of the comprehensive style) over scholarship oriented toward social impact, e.g., scholarship that aims to give voice to minority groups or to find solutions to “real-world” problems. However, the preferred style varies across competitions: the constructivist style was most favored by panelists serving on one competition studied, while the utilitarian style was much favored in another (Mallard, Lamont, and Guetzkow 2005; see Table 2 in Appendix A). These outcomes were strongly influenced by the directives the funding agencies gave panelists concerning the criteria to be used in evaluating. We found that panelists were generally very amenable to serving the priorities of the funding agency.

We believe that our typology of epistemological styles may be of help to the Council as it seeks to operationalize “impact” as a criterion of evaluation. Our inductive analysis suggests that, in the context of American funding panels at least, humanists more than social scientists consider “giving voice” to groups that are underprivileged or stigmatized to be a legitimate contribution. At the same time, humanists tend to scorn applied knowledge while social scientists are more supportive of the utilitarian style and more eager to support projects that have immediate policy relevance. This finding suggests that different types of social impact should be emphasized for the humanities and the social sciences. It will be important to favor criteria that have intrinsic meaning for each category of scholars as well as to encourage each category to broaden its criteria regarding impact. Lamont (in progress) also tackles the question of homophily in peer evaluation. She shows in her larger study that peer reviewers are likely to value research that best corresponds to the criteria of evaluation which make their own work shine. Despite this preference for “interesting work” that corresponds to the panelists’ specific tastes, she also finds that evaluators believe they are able to identify the best of the proposals (that “cream rises”) if panel members follow a number of unspoken rules concerning, for instance, deference to disciplinary expertise, strategic voting, horse-trading, and the bracketing of personal ties.

These results suggest that if the Council wishes to broaden the criteria of evaluation used by peer reviewers, it should recruit evaluators whose scholarship
embodies the criteria it seeks to implement. Panel officers interviewed in the five competitions under study acknowledge that they use such strategies. For instance, program officers affiliated with the most interdisciplinary of the five competitions studied acknowledged that they selected “broad-minded” evaluators associated with interdisciplinary fields such as gender studies or film studies, with the hope that they would be in the best position to appreciate the virtues of strong interdisciplinary proposals.

The results of Lamont’s study echo others that describe the cognitive contents of peer evaluation and the process aspect of deliberations. Together, they form what we label the “second wave” of studies on evaluation. For instance, Brenneis (1994, 1999) has described how different characteristics of applicants matter at different stages in the evaluation of applicants to the Graduate Fellowship competition of the National Science Foundation. Quantitative indicators (SAT and GRE scores as well as average referee scores) have most weight at the beginning of the process of decision-making. The use of such indicators, which black-box many social attributes of the applicants, creates the feeling of impartiality among panelists. Brenneis finds that the social characteristics of applicants (e.g., gender and geographical location) only come to play a direct role when panelists seek to differentiate between applicants of similar scientific quality. These studies show that Merton’s (1973[1942]) norms of fairness, which prescribes that the social characteristics of candidates never be discussed in peer evaluation, conflicts with another norm, that of compensating for implicit biases while maintaining ultimate concern for the quality of the work.

The need for such compensation is emphasized by authors who show that biases exist in standard measures of “quality.” While reports commissioned by the U.S. Congress (G.A.O. 1994; Cole, Rubin and Cole 1978) measured the “quality” of a researcher by the number of publications and citations in scientific reviews indexed by the Institute of Scientific Information (ISI), this indicator discriminates against authors who publish books or book chapters, which is relatively frequent in the humanities and some of the social sciences: their literature of reference is excluded from the ISI data.

---

9 This criticism applies to the Social Science Citation Index and the Science Citation Index. Authors who define quality in terms of publications in scientific journals are Campanario (1998a, 1998b); Feagin 1999;
bank. This indicator also discriminates against those who use qualitative, as opposed to quantitative data, and those who wish to speak to broader audiences, as opposed to their disciplinary peers. Indeed, based on an analysis of more than fifty books published during 1987 and 1988, and of all the articles published in the *American Journal of Sociology (AJS)* and the *American Sociological Review (ASR)* for the years 1987-1988, Clements et al. (1995: 460) show that compared to authors who publish books, authors who publish articles are more likely to use quantitative data and statistics and to cite references within their discipline. Therefore, measuring “quality” by number of citations and publications in scientific reviews favors researchers who do quantitative work (they are more likely to cite from a quantitative literature) and favors those who do not speak to a broad academic and non-academic audience. Clements et al. (1995) also find that books are more cited from outside the discipline of their authors than articles, i.e., authors of books have a wider impact, even though their work is not included in ISI. These results suggest that by funding research oriented toward the publication of books (rather than journal articles), the Council could favor scholars who wish to speak beyond the limits of their discipline and area of specialty.

Other studies have shown that a definition of quality based on number of publications in top scientific journals or on number of citations also discriminates against women. A study by Lutz (1990) suggests for instance that, in the field of anthropology, the gender of authors does not so much affect the capacity to present papers at meetings or to publish papers in journals. It only affects the likelihood to publish in top journals, which are likely to receive the most citations: in the 1980s, 45 percent of the scholars who presented papers at anthropology meetings were women; 30 percent of the authors of articles published in the top journals of the discipline were women; and only 20 percent of the citations went to women. Lutz (1990: 620) also observes that “men cite women authors only half as often as women do.” This difference decreases when self-citations are eliminated, but it remains significant. This difference might be accounted for by the fact that women anthropologists are more numerous in research specialties that have not traditionally defined the core of the field (e.g., gender) – for instance, see Ward, McPhail and Simon (1987, 1989); Hargens (1987); Cole, Simon, and Cole (1987); and Champion and Morris (1973).
and Grant (1985) on feminism’s relative invisibility in sociology. Or it might be explained by gender differences in epistemological approach, if an epistemological style found in minority (like constructivism) is more likely to be endorsed by women than men.\footnote{Mallard et al. (2005) found that the constructivist style (which aims to “give voice”) is favored by feminist scholars. On the relationship between constructivism and feminism, see Harding (1986) and Smith (1990).} It might also be explained by the general discounting of the voice of women by men. Lutz (1990) cites experimental tests that confirm this latter explanation. In an experiment, the same paper was circulated with either a male or a female author – see Golberd (1975) cited in Lutz; and Friend, Kalin and Giles (1979) for a replication of the experiment. The paper with the male author was rated significantly higher by peer reviewers on a number of criteria including value persuasiveness and professional competence. The existence of these biases suggests (contra Merton (1973[1942])) that some considerations about the gender of applicants in the last stages of an evaluation process may compensate for these biases.

Overall, these studies suggest that peer evaluation introduces gender and methodological biases, but that these biases can be corrected by funding agencies if they have been fully understood. They also provide a more empirically grounded and adequate understanding of fairness and quality than the first-wave literature. Supporting further research on deliberation processes in peer review may assist the Council in ensuring that it creates the conditions for maximum fairness and that it does not unduly support a definition of quality that favors some types of research over others.

3. THE MANAGERIAL MODEL OF EVALUATION IN THE UNITED KINGDOM

If issues of fairness were prevalent in the first wave of studies on peer evaluation in the United States, they remained largely absent from the studies on peer evaluation conducted in the UK. This absence is largely explained by the initial push that drove studies of evaluation in the UK. Since the beginning of the 1980s, quantitative studies of research “quality” have increasingly been used to assess research performance of university departments and individual researchers in the UK. These studies were initiated...
by the Thatcher administration without the support of professional associations. Among these large-scale projects of evaluation of research performance, the University Grants Committee (UGC) initiated a large cross-disciplinary survey of research performance across disciplines (CVCP/UGC 1985, 1986) that was later adopted as a guide for evaluation by the Advisory Board of Research Councils (ABRC) in the UK (ABRC 1987, 1988). Initiated in 1985, the survey was used as a guide to allocate resources for all university research (Phillimore 1989). In contrast to the United States, the study of evaluation was not responding to demands of openness and fairness expressed by professional associations. Rather, Fulton (2003) shows that the use of these surveys of research performance was part of the tool kit of “New Public Management,” which aimed at weakening the long-standing peaceful coexistence between state bureaucracy and self-regulated professional associations that had prevailed from the 1940s to the end of the 1970s.

The case of the UK is of particular interest for the Council in that the British policy reform was guided by the same values of openness, increased impact, and social relevance of research for users that are at the core of a Knowledge Council. However, the policies sustained by these values failed to meet the approval of the communities of scientists, particularly in the SSH. Overall, this example shows that without coordination and cooperation with communities of scientists, a funding agency can hardly reorient the research it funds toward a more broadly defined base of users. Because these reforms were not fully discussed by communities of researchers, they strengthened the biases inherent in their system of peer evaluation. We will detail how the choice of indicators and the process by which they were implemented have led to such opposition. The British experience clearly demonstrates the importance of consulting with the scientific community in implementing new directions, which is how the Council has proceeded thus far.

A) The First Wave of Studies in the UK

The tools used in the UGC survey to measure quality (CVCP/UGC 1985, 1986) drew upon previous studies of quality in the natural sciences within individual scientific institutions (Martin and Irvine 1984; Irvine and Martin 1984; Vinkler 1986; Irvine et al.
The UGC survey represented a first attempt to compare and rank all departments and universities within a complex national ecosystem of academic disciplines (Carpenter et al. 1987). Only research activities were taken into consideration to produce this ranking, which created a bias against disciplines (particularly interpretive disciplines) in which teaching is considered to be an important source of prestige. Research performance was measured in the UGC survey in terms of: a) “output” (defined as number of peer-reviewed publications); b) “impact” (defined as number of citations); c) “quality” (measured by prior research grants allocated by the Research Councils); and d) “utility” (measured by the external income received for graduate students and other laboratory personnel from private firms and by the number of patents and licenses obtained by the laboratory). The level of “research performance” was measured and accounted by a “cost centre” (an administrative unit in charge of balancing academic budgets), which made sense from the administrative, but not from a scientific, perspective (CVPC/UGC 1986).

This method of measurement worked to the detriment of various fields. For instance, articles published in disciplinary journals were given more weight than articles published in interdisciplinary journals. Also, the system favored the natural and applied sciences to the detriment of SSH, to the extent that “utility” was measured by private funding received, as well as by income generated by patents and contracts with private actors. Community building, participation in public debates, teaching to college students, and other forms of public contributions that bear direct “utility” for the nation as a whole were excluded from measures of “utility.” As a result, the use of the UGC’s survey was strongly criticized by social scientists and humanists. The lack of a culture of audit in academia in the 1970s was a major obstacle against legitimating the use of standard measures of quality (Strathern 2000). Moreover, the secrecy that surrounded the first UGC survey of 1985-86 helped convince scholars that the latter was to be used against the interests of academics and was part of a hostile campaign against public service (Phillimore 1989). That this survey was first conducted at a time when the Thatcher administration was questioning the very existence of SSH research and shortly after discussions about the possible closing of the Social Science Research Council (SSRC) did not help matters; SSRC survived under the name Economic and Social Research Council (ESRC), but with a much reduced funding capacity (Fulton 2003). Had the
reform of evaluation criteria been implemented at a time of budgetary growth for SSH research, it would certainly have received more support. Social scientists and humanists became united in their opposition to the new measurement regimen (Phillimore 1989: 259). The new market-based focus on input/output ratio seemed particularly unacceptable given that the government was not in a position to improve the lot of academics. This case shows that the context in which such measures are implemented can have a strong impact on whether they will be implemented.

B) The Second Wave of Studies on Evaluation in the UK

The implementation of UGC surveys generated a second wave of studies of peer evaluation that criticized attempts to systematize and standardize criteria of evaluations. The criticisms took at least three different directions:

First, some authors were critical that the British Government was introducing new criteria of evaluation without consultation. For instance, Phillimore (1989) argued for the importance of building a consensus among researchers before institutionalizing new criteria, given that UGC standards became public only after the results of the survey were made available. For Phillimore, “by its failure in 1985-86 to be more explicit and open about both the conduct and the objectives of its evaluation, the UGC made the acceptance of its results almost impossible” (269). UGC’s strategy was especially faulty because the stated goal of the reform was to improve transparency, accountability, and openness. The importance of consultation is also stress by Callon, Lascoumes, and Barthe (2001). Going even further than others in their demand for consultation, these authors advocate the Danish model of “citizens’ conferences,” which consists of gathering a group of researchers and users to discuss what criteria of evaluation should be used to assess the social “usefulness” of a research project, its potential risks, and how to assess the latter. The British office of Higher Education and Research Opportunities (HERO) (which administers the national Research Assessment Exercise (RAE) conducted every four years) grew critical of, and attempted to reform, its own system of evaluation using this Danish approach. Rather than measuring scientific performance using secret criteria, the national RAE (2001) first spoke to relevant parties about pertinent criteria of
evaluation, which were defined in consultation with academic communities during the summer and autumn of 1999.11

Second, the Higher Education and Research Opportunities (HERO) reformed its methods of measurement of science research in consultation with bibliometry experts. Along with the Economic and Social Research Council (ESRC), it hired Danish experts who had perfected citation-based bibliometric techniques to evaluate the quality of British research in the 1990s (e.g., Moed et al. 1985; Nederhof and VanRaan 1989, 1993; Nederhof et al. 1989; VanRaan et al. 2001; VanRaan and Van Leeuwen 2003).12 These Danish researchers were very concerned with issues of reliability: they compared the rankings obtained from bibliometric measurement to qualitative evaluation by peers to assess the reliability of rankings of research groups produced by the Economic and Social Research Council (Nederhof and VanRann (1993). Instead of considering bibliometric measures as the ultimate arbiter of quality, they viewed such measures as complementary to other approaches. Because the SSHRC-commissioned report on bibliometric indicators for measuring quality discusses the work of Danish experts at some length (Archambault and Gagné 2004), we will only note that: a) Unlike UGC, which was concerned with “utility” measured in part by support from the private sector and revenue-generating patents, this Danish literature focused only on the least controversial aspects of quality, namely outputs and citation counts (Moed et al. 1985). b) Authors associated with this approach are careful to spell out the limits of their work. They compared data on publications and citation counts identified through the Institute for Scientific Information (ISI) database with information obtained directly from different research groups at one university. They found that in most cases, the ISI provides very incomplete data on publications and citations: for instance, only 24 percent of the publications of a group of mathematicians included in their study were part of the ISI databank (Moed et al. 1985). They also suggested that two additional potential sources of bias are that the UGC’s focus on publications in the last five years only, which may discriminate against long-term

---

11 For a typical timetable of the UK’s Research Assessment Exercise (RAE), see <http://www.hero.ac.uk/rae/AboutUs/time.htm>.

12 In contrast, the National Teaching Quality Assessment does not determine allocation of funds, which is still calculated based on numbers of students.
projects (Moed et al. 1985), and a focus on publications in English-language and international publications, which favors researchers from English-speaking countries (Archambault and Gagné 2004). c) They acknowledged that citation practices vary across fields and between national traditions, which affects the reliability of bibliometric measures of quality (Moed et al. 1985). They find that in the more humanistic fields, citations are very likely to be found in books (not covered in the ISI) than in articles (the least favored format of publication in the humanities) (Archambault and Gagné 2004). Therefore, what appears to be an unproblematic measure of scientific quality in different fields of the natural sciences might raise extensive problems if applied to the humanities and social sciences (Nederhof et al. 1989a). Such criticism seems to resonate with the skepticism of Canadian researchers concerning the implementation of simple indicators of research performance (SSHRC 2005: 3).

A third critical direction taken by studies of evaluation in the UK concerned the underlying philosophy behind the new criteria of evaluation, as opposed to the undemocratic implementation of UGC’s review or the limitations in measurement (Neiland and Woolgar 2002). Inspired by Foucault, Strathern (2000) remarked that the diffusion of an audit culture in the academic world is difficult to criticize insofar as it promotes values of openness, transparency and democracy. In her path-breaking work, this British anthropologist criticizes UGC for downplaying the fact that its evaluative work results in greater standardization and normalization of research practices: audits and performance assessment posit commensuration, i.e. the need to compare different units by using a single standard (e.g., productivity and impact) (Espeland and Mitchell 1998). Yet audits cannot take into consideration the diversity of epistemological cultures (and criteria) that exist within and across disciplines. Other effects include the strengthening of inequalities between “good” and “bad” departments, which may push faculty members in “bad” departments to disengage from research (Fulton and Holland 2001). Also building on Foucault, Shattock (1994) argues that the departments which have a low rank are more often put under “surveillance” of the central administration of universities. More monitoring and accounting of activities can fall on the shoulders of those faculty members who already face decreased budgets for research. These two factors have had a negative impact on these departments’ capacity to engage in research activities. In the
“best” departments, a governmental discourse of “New Public Management” might very well coexist with the traditional professional ethos characteristic of researchers, since funding has not affected the capacity and motivation of researchers to engage in research activities (Fulton 2003). In contrast, academics working in the “worst” departments might be affected by de-professionalization (brought about by a decrease in research funding), routinization of the labor process (due to concentration on teaching and administrative tasks in lieu of research), and overall proletarization (Halsey 1992).

These criticisms formulated against USG evaluation have been echoed by normative studies of evaluation. For Travis and Collins (1991), who studied funding committees of the British Science and Engineering Research Council, evaluators should be careful not to create “lock-in effects,” which consists in institutionalising a reality while measuring it (as in the case of racial categories used in national censuses), hindering the emergence of new poles of excellence. Some believe that quantitative and standardized assessments of research performance are more likely to create lock-in effects than qualitative and fine-grained peer review (Collins and Evans 2002; Callon 1994). The importance of adopting an incremental, collaborative and reflexive approach to the development and implementation of new criteria of evaluation seems to be the main lesson that SSHRC should take from the British experience with peer review. Far from being an obstacle to the introduction of new criteria, context-sensitive peer review may be considerably more conducive to change than the indiscriminate application of standardized quantitative measures of impact that were privileged by the British Government.

4. THE POST-CORPORATIST MODEL OF EVALUATION IN FRANCE

French research on criteria of evaluation largely focuses on hiring decisions, and it pays limited attention to funding decisions and the peer review process, especially as it concerns SSH. Unlike their North American counterparts, French scholars have not been concerned with fairness, which can be largely explained by the institutional context of research in France. Leading journals, such as the Revue Française de Sociologie and the Revue Française de Science Politique, do not have the same impact on French academic careers as their American counterparts, such as the American Sociological Review or the
Indeed, while the appointment of members to the editorial board of the latter journal follows a strict process supervised by the national professional associations, French professional associations are far less institutionalized and constitutive of the organization of the profession. In the case of sociology for instance, it is only in 2004 that a large professional association was created with the goal of promoting the profession and the discipline. This cross-national difference is due in part to the fact that SSH disciplines in the United States bring together a considerably larger number of scholars than their French counterparts, and scholars are much more dispersed geographically in the U.S. than in France (Lamont 2000). Perhaps the fact that tenure is granted shortly after the hiring of new doctors might account for the lesser concern for the legitimacy and quality of peer evaluation among French experts, and for “publish or perish” more generally (Musselin 1996, forthcoming). The situation may change with the emergence of European associations and the possibility that the Centre National de Research Scientifique, which now hosts research centers and laboratories, may be turned into a granting agency (Laredo 2003).

The limited centrality, standardization, and legitimacy of professional and disciplinary evaluation is reinforced by the presence of a relatively lively public market for ideas, which acts as an alternative sphere of legitimation of scientific products, and whose existence makes it possible for academics to bypass strict evaluation by peers. Lamont (1987) shows that humanists such as Jacques Derrida may rise in French academia by appealing to a broad public, by publishing books with publishing houses that market to wide audiences rather than publishing in professional journals, and by seeking recognition from the media more than from experts in the field. Based on an ethnographic study of literary scholars, Mallard (under review) has also observed that creating critical or digital editions of the complete work of writers is another way for scholars to

---

13 In the SSH, most research associations are organized around a topic of research and not by discipline. Generally, in the humanities, these associations are devoted to the study of an author (like Balzac, Kant, Hegel, etc.) and in the social sciences, around a research specialty (like immigration, urbanity, science, etc). But they are generally not part of a larger disciplinary association as in the U.S. Only the “Fondation Nationale des Sciences Politiques” is a powerful professional association.

14 For instance, one of the most important journals in the SHS, Actes de la Recherche en Sciences Sociales is not a peer-reviewed journal.
accumulate prestige and authority in academia while bypassing peer experts. Contracts with marketing-oriented publication firms are based more on personal contacts than on scholarly quality. These studies confirm Boudon’s (1981, 1990) analysis according to which French academia has at least two markets for intellectual products (academic and media-driven), two sets of criteria of evaluation, and thus an alternative to the judgement of peers. While the existence of these two markets has been the subject of research in the French context, Posner (2001) proposes that we are developing a parallel phenomenon in the United States.

This diversity of career paths means that peer evaluation is largely contested in the French academic field, and that compared to their American counterparts, many French scholars are skeptical of the ability of their colleagues to form informed and disinterested opinions of their colleagues’ work. Thus, the literature has focused not on double-blind peer review, but on the evaluation of job candidates, candidates for full professorship, and candidates for the chairmanship of departments. The two most systematic of these studies have been conducted by Bourdieu (1988) on the 1970s and by Musselin (1996, forthcoming) on the 1990s. The first contrasts evaluation in governmental agencies concerned with SSH research (the Centre National de Recherche Scientifique, the Institut National de la Statistique et des Etudes Economiques, the Institut National d’Etudes Démographiques, etc.) with evaluation in academic departments, both of which provide full-time and lifetime employment for researchers. Bourdieu (1988) studied the criteria of evaluation operating in university departments and research institutes. Based on a large-scale survey of scholars in different universities and research institutes, this study maps the conflicting criteria of evaluation at work across disciplines and “positions” in what he calls the “academic field.” Bourdieu (1988) finds that new understandings of academic worth were introduced in academia by the creation of CNRS

15 Since 1945, three waves of institution-building have occurred in France, to which correspond three different levels of peer review (Laredo 1995): 1) national institutes like CNRS, INSERM, etc.; 2) national programs of evaluation aimed at fostering innovation through the creation of “Technological Programs” in energy, biotechnology and information technology, like the Agence Nationale de VAlorisation de la Recherche (ANVAR); 3) regional and European funding panels aimed at fostering networks of excellence in local regions. Only recently did these committees began funding research in the SSH, which explains why the evaluation of the SSH by independent research councils has been under-developed in France.
laboratories in SSH areas in the 1960s and 1970s. Researchers in the SSH departments of the CNRS, in research institutions like the “Collège de France,” and in departments in new Parisian universities (such as “Paris VIII- Vincennes” or “Paris VII- Jussieu”) are most likely to favor interdisciplinary scholarship (associating the three pillars of anthropology/sociology, philosophy, and linguistics) as well as pure and disinterested specialized research. These very terms are in direct opposition with the criteria of evaluation prevalent in the old academic departments, particularly identified with “la Sorbonne” and the “Ecole Normale Superieure d’Ulm” (Bourdieu 1988). These latter institutions privilege criteria of evaluation such as the fact of mastering the philosophical foundation of a discipline or of having a general disciplinary culture, that is, a solid knowledge of the work of the founding fathers of a discipline -- as opposed to mastering recent developments within substantive subfields. This is in line with the emphasis on “culture générale,” a form of cultural capital valued by the French bourgeoisie, but relatively absent in other classes (Bourdieu and Passeron 1964). Bourdieu (1988) has shown that the mastery of “culture générale” and linguistic codes are essential to pass the state exams used to select the elite of the state bureaucracy (e.g., entrance exams to the Ecole Normale Superieure or agrégation). But for Bourdieu (1988) the segmentation of the academic market created by the rise of CNRS is an indication of the end of academic corporatism, and of the development of a more autonomous academic field.

Bourdieu’s (1988) mapping of the French academic field has been complemented by several studies which criticized certain of his generalizations. Musselin (forthcoming) argues that Bourdieu (1988) exaggerated the link between academic institutions and the criteria of evaluation used by scholars; institutional location accounts for less than 15 percent of the variance in criteria used by academics. She also criticizes Bourdieu for offering an overly dichotomic analysis of the academic field by emphasizing differences between the avant-garde of the academic field (found at the CNRS and the “Collège de France”) and the old guard (found at the Sorbonne). Moreover, for Dubar (2005), Bourdieu’s account, published in the late 1980s, posits that the academic field has remained unchanged since the mid-1970s and that criteria of evaluation have remained stable and have not been affected by the growing number of applicants to academic positions. Different reports on career trends produced by the CNRS show that Bourdieu
(1988) analyzed pre-1978 patterns and prematurely generalized from them. This is illustrated by the fact that from 1978 to 1982, the sociology section of the CNRS hired 100 new researchers who had received governmental contracts in various applied fields such as health and welfare, demonstrating a clear applied (as opposed to “avant-garde”) orientation (Dubar 2005). Dubar (2005) shows that after 1982, the CNRS was greatly limited in hiring theoretically promising (or “avant-garde”) researchers. Thus, the picture is far more complex than that provided by Bourdieu if one takes into consideration change over time.

Other studies have suggested that that the opposition between the avant-garde of interdisciplinary theorists in the CNRS and in new universities (created in the 1960s and 1970s) and the old guard of professors in more traditional universities (like the Sorbonne) is weakening. First, many predict that soon all CNRS researchers and laboratories will be integrated into the university system (either in newer or older universities). Thus, the CNRS is likely to become a granting agency (akin to SSHRC), as opposed to a research organization in which scholars spend their entire careers (Laredo 2003). Second, since the 1970s the criteria of evaluation favored by university professors in old institutions (like the Sorbonne) and in new ones (like Paris VIII -Vincennes) have been converging.

Duell’s (2000) study of criteria of evaluation in the humanities reveals patterns far different from those described by Bourdieu (1988). On the basis of forty interviews with French and American scholars he shows that, whereas American literary studies scholars emphasize the role of theory and political engagement in their scholarship, “in the process of entering the mainstream of French literary studies, modernist scholars appear to have […] increasingly adopted the traditional criteria of the discipline” (109). This analysis suggests a convergence of all literary scholars, independently of the type of institutions to which they are affiliated. The new era features fewer politicized and theoretical debates. Controversies have become fewer, being no longer sustained by the

---

16 Since 1982, after the massive hiring of contractual researchers, the CNRS has hardly hired more than 4 researchers per year. As a result, Dubar (2005) predicts that from 2005 to 2010, half of the CNRS population will retire, without being replaced, from which it will result that the CNRS researchers will represent less than 25 percent of the population of sociologists in France. Therefore, in 2010, 75 percent of the sociologists will be university-based enseignants-chercheurs.
deep political cleavages that characterized French academia until the 1980s. However, more studies are needed on this topic, since other authors (Mallard under review) find that in literature, boundaries have shifted rather than disappeared. Mallard (under review) finds that controversies grounded on theoretical and political differences in the 1970s are now increasingly grounded on the use of different media used to publish authors and researchers.

The French literature on evaluation is also less concerned with issues of fairness (as in the U.S.) or control (as in the UK) than with career paths, as seen in Musselin’s (1996, 2000, forthcoming) mapping of the criteria of evaluation at work in hiring decisions in France, Germany, and the United States. She argues that in France, modes of recruitment that are market-driven coexist with non-market mechanisms within a “post-corporatist academic field.” Different criteria are mobilized for different types of candidates: If recruiters are looking for a candidate charged with fulfilling local administrative and teaching needs, they value “good citizens” who have experience participating in administrative tasks, who invest in teaching more than research, and who cultivate ties with local communities. In contrast, when looking for a promising researcher (a “brilliant scholar”), they will put more emphasis on having a strong publication record and participating in international conferences (Musselin 1996, forthcoming). A market logic (which requires establishing commensuration among candidates, downplaying interpersonal connections, and avoiding salary negotiations) is mobilized differently and with different levels of intensity when recruiting these two types of candidates: the evaluation of promising researchers is more guided by a market logic than is the evaluation of candidates who are expected to fulfil local administrative and teaching needs. Personal ties with the latter case are helpful in assessing the candidate’s potential as a teacher and administrator. The prime concern of hiring commissions is not ensuring fairness but reducing uncertainty concerning the different types of hires.

These studies all point to the diversity of career paths in French academia, which appears to be a segmented and complex ecosystem. They are very useful given the Council’s recommendation to foster dialogue between research institutions, liberal arts colleges, and communities (SSHRC 2005: 5). Rather than opposing the “citizen scholars”
to “brilliant researchers” (to use Musselin’s typology), a policy that seeks to maximize the vitality of a field should mobilize the resources specific to each group and should devise ways to create synergies between them. Instead of defining exclusively the first group as knowledge diffusers and the second as knowledge producers, it would be advisable to consider both types of scholars as having specific resources to produce and diffuse knowledge. For instance, the “good-citizen scholars” may produce research oriented toward social problems and social change and may diffuse their work in a range of outlets including local newspapers, videos, and websites, while the “brilliant researchers” may produce disciplinary-oriented research and publish in scientific journals. While a Granting Council usually grants funds to “brilliant researchers” to assist them with their research, the Knowledge Council that SSHRC hopes to become could also contribute to the work of “citizen scholars” by giving them the means to create a website, build a data bank, produce a video documentary on a social issue, or develop a CD-ROM on a writer. Thus, it could foster the greater involvement of academics with their community. The “citizen scholars” may be in a much better position than elite researchers to produce such alternative forms of knowledge. The Community-University Research Alliance and the creation of Strategic Clusters seem to be two organizational designs able to foster such programs (SSHRC 2005: 9). Moreover, a Knowledge Council could create incentives for top researchers to put more energy into making their work known to wider audiences. It could include publication in non-professional outlets as a criterion of evaluation of grant proposals and offer workshop where top professionals could learn how to write for wider audiences.

5. CONCLUSION

In this review, we have described the various perspectives that emerge from the literature on peer evaluation in the United States, the United Kingdom, and France, what results each literature has produced, and what lessons SSHRC can draw from each cases. Overall, our report suggests that while traditional models of peer evaluation might constitute an obstacle to the Council’s objective to fund research that reaches broader publics, the literature also suggests ways to overcome these obstacles.
1) The American literature on professional peer evaluation suggests that the Council can foster *internally* a greater diversity of criteria of evaluation by appointing panelists who themselves are less exclusively wedded to traditional criteria. For instance, it can select peer reviewers who show a strong interdisciplinary orientation, as is often the case for individuals emerging from new fields such as women’s studies or science studies. Moreover, the goal of appealing to a broader public could and should be translated into the language of each discipline, without bias against certain methods and disciplinary traditions.

2) The literature on the managerial evaluation of research in the UK shows the pitfalls that the Council should be careful to avoid when implementing new standards of evaluation: it should be very cautious with bibliometric measures of “impact,” which tend to overlook the diversity of criteria of evaluation present in a field, introduce a bias in favor of those who publish primarily in disciplinary journals, and introduce a bias against women. We recommend that the Council give serious thought to what makes the vitality and impact of a field on society. For instance, serving on a committee of the American National Academy of Sciences charged with evaluating which areas of the research on aging show most vitality, Lamont (2004) has identified several dimensions of vitality, namely: a) Generativity: the extent to which specific contributions lead to new theories and “discoveries” and the extent to which scholars in a field come up with new ideas; b) Growth (intellectual and institutional): the extent to which a field attracts energy and produces activity, via the creation of researchers, publication outlets, sources of funding, and other resources; c) Diversity: the range of problems being tackled by those active in a field; d) Interdisciplinarity: the extent to which a field engages questions raised by other fields; e) Attraction: the extent to which researchers outside a field follow it and engage it; f) Diffusion: the extent to which knowledge from a field diffuses outside of it; g) Applicability: the extent to which knowledge generated by a field has practical applications in policy, business, law, education, and elsewhere. Perhaps the Council should consider these dimensions to define even more precisely what it means by impact and how to foster it.
3) The literature on the diversity of criteria of evaluation in France shows that academia is a complex ecosystem, where different careers and academic positions put emphasis on diverse values and skills. To better engage this ecosystem with the research it funds, the Council could design programs targeted at different populations of knowledge-producers. Instead of privileging only one type of academic, the Council should identify what specific skills each type of scholar and their specific communities (international peers, experts in think tanks, students, local communities, etc.) can bring to the table. Creating specific competitions oriented toward generating various outcomes aiming at broader diffusion (such as documentaries, CD-ROMs, websites, etc.) could make room for greater diversity in producers of knowledge. Also, it might be recommended that the Council act with the active support of professional associations, to define rules of fairness for the support of alternative forms of scholarship. Many academics are eager to broaden the impact of their work. The Council needs to tap their energies as it works toward implementing new criteria of evaluation.
# Appendix A: Tables

## Table 1:
*Generic Definitions of Originality by Disciplinary Clusters*

<table>
<thead>
<tr>
<th>Originality Type</th>
<th>Humanities</th>
<th>History</th>
<th>Social Sciences</th>
<th>All Disciplines</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>Approach</td>
<td>29</td>
<td>33%</td>
<td>26</td>
<td>43%</td>
</tr>
<tr>
<td>Data</td>
<td>19</td>
<td>21%</td>
<td>6</td>
<td>10%</td>
</tr>
<tr>
<td>Theory</td>
<td>16</td>
<td>18%</td>
<td>11</td>
<td>18%</td>
</tr>
<tr>
<td>Topic</td>
<td>13</td>
<td>15%</td>
<td>6</td>
<td>10%</td>
</tr>
<tr>
<td>Method</td>
<td>4</td>
<td>4%</td>
<td>5</td>
<td>8%</td>
</tr>
<tr>
<td>Outcome</td>
<td>3</td>
<td>3%</td>
<td>4</td>
<td>7%</td>
</tr>
<tr>
<td>Understudied Area</td>
<td>5</td>
<td>6%</td>
<td>3</td>
<td>5%</td>
</tr>
<tr>
<td>All Generic Types</td>
<td>89</td>
<td>100%</td>
<td>61</td>
<td>100%</td>
</tr>
</tbody>
</table>

Note: some columns may not sum to 100% due to rounding.

Source: Guetzkow, Lamont, Mallard (2004:201)

## Table 2:
The Distribution of Epistemological Styles Used by Each Panelist by Disciplinary Clusters and Disciplines*

<table>
<thead>
<tr>
<th>DISCIPLINARY CLUSTERS</th>
<th>Humanities</th>
<th>History</th>
<th>Social Sciences</th>
<th>All Disciplines</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>English</td>
<td>Musicology</td>
<td>Art History</td>
<td>Philosophy</td>
</tr>
<tr>
<td>Constructivist</td>
<td>11</td>
<td>14</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Comprehensive</td>
<td>38</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Positivist</td>
<td></td>
<td></td>
<td>14</td>
<td>6</td>
</tr>
<tr>
<td>Utilitarian</td>
<td></td>
<td></td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>72</td>
<td>22</td>
<td>14</td>
<td>6</td>
</tr>
</tbody>
</table>

* Each column represents the epistemological style(s) that each panelist uses. A black cell represents an epistemological style used by a panelist. While some panelists used consistently only one style while accounting for his/her evaluation of proposals, others used two.

(Source: Mallard, Lamont, Guetzkow Under review)


