DON BRENNEIS

Anthropology in and of the academy: globalization, assessment and our field's future

In considering the challenges and opportunities likely to be faced by social anthropologists over the coming 20 years, this paper begins with a recognition of the critical role of institutional structures and processes, especially practices of evaluation and assessment, in the future trajectory of our discipline. The core of the article critically explores two general modalities of assessment and evaluation: deliberative processes, of which peer review is a classic example, and more formal techniques focused on particular quantitative indicators such as citation factors and impact analysis. The discussion draws upon ethnographic work on and from the midst of such bureaucratic sites, on tracking in some detail the conflation of descriptive and evaluative practice embedded in the forms of quantitative metrics, and on current critical examinations of both deliberative and analytical strategies. The article argues that deliberative, consultative peer review can lead to much more acute, textured and realistic outcomes for such reviews, whether of programmes or individuals, than can a reliance solely on bibliometrics. I also suggest that scholarly associations such as EASA have a particular role to play both in arguing for the value of serious collegial engagement in such work and in modelling, in ways with which social anthropologists are deeply familiar, how such qualitative reviewing might be responsibly and proactively pursued.

Key words research policy, academic assessment, audit culture, peer review, bibliometrics

Introduction

When I became Editor of *American Ethnologist* in 1989, friends and colleagues immediately began to accord me oracular powers, assuming that the editorial vantage point would give me a better chance to look into the discipline's future. While I deeply enjoyed the editorial work and learned a great deal from it, I was never very good at foretelling anthropological futures, at least those extending beyond the time frame required as publication lead time. In thinking about our next 20 years, then, I'll be making no substantive predictions concerning topics, methods, sites and practices. Rather, I want to point to the broader context of institutional structures and practices within which our future course will take shape – and which will be likely to play a consequential if not always obvious role in that shaping. More specifically, I will be exploring some aspects of the crucial relationship between practices of evaluation and assessment and decisions concerning the future trajectories of social anthropology and the interdisciplinary enterprises in which we are increasingly engaged.

doi:10.1111/j.1469-8676.2009.00077.x

The contemporary context is both complex and challenging, with already ongoing processes of centralisation, standardisation and transformation deeply inflected by the current economic situation. Several key elements of the present scene are worth highlighting. First, higher education in particular has traditionally been considered what some economists call a 'trust market' (Winston 1999: 14), one not subject to direct short-term analysis and valuation. There has been a major transformation in this respect, one especially evident in the rise of academic audit and accountability (Power 1994, 1997; Strathern 1997, 2000; Shore and Wright 1999; Brenneis *et al.* 2005). Such institutional monitoring is not limited to reviews of research and pedagogy; generalised processes for evaluating human subjects research have become increasingly constraining and are shaping knowledge production in very complex ways (Lederman 2006; Brenneis 2005). Second, as is perhaps most clearly evident in the UK and Australia, higher education has become a major 'export product' and source of foreign revenue. And, for graduates and bureaucratic planners alike, certification and the labour mobility it can afford have become increasingly important. Third, in the context of a highly touted and in some areas actually emerging 'knowledge economy', practices of knowledge production, circulation and reception once centred within universities have become increasingly, if variably, subject to privatisation (McSherry 2001; Brenneis 2004). While not all scholarly or scientific research holds the promise of direct profitable applications, styles of assessment increasingly informed by managerial and commercial perspectives have become widely consequential. League tables, rankings and other artefacts through which the value of scholarly work is represented have become commonplace features of the academic landscape. Finally, funding for interdisciplinary research has expanded dramatically, especially in the EU, and extradisciplinary engagements both with other fields and with external stakeholders have become critical.

I must note a significant difference between the European scene, where centralised funding and policy-making can and do directly influence higher educational institutions, and the United States, where such direct control is presently unlikely. The US scene is a very pluralistic one, with many different kinds of universities with multiple funding sources. The accreditation processes at the heart of institutional governance remain in the hands of voluntary associations of schools and colleges, although, during the recent Bush administration, the federal government tried to legislate more direct regulatory involvement. On the European side, a national government can mandate a process such as the Research Assessment Exercise (RAE) in the UK. In the US, on the other hand, there often is a great deal of discursive seepage, that is, the language and practices informing European policy may be picked up by particular schools or administrators, but, at present, no agency is in the position of requiring and implementing them across the board. While current discourses shaping higher education and policy discussions may indeed be not just transatlantic but globally dispersed, their effects are quite variable.

In the remainder of this article I want to draw upon a distinction central to a 2007 report from the National Academy of Sciences, *A Strategy for Assessing Science*, the work of a committee organised by the US National Institute on Aging (Feller and Stern 2007). The authors distinguish heuristically between two general modalities of assessment and evaluation: deliberative processes, of which peer review is a classic example, and analytical techniques involving more formal and quantitative indicators. I'm going to turn first to peer review, a classic deliberative process, and then consider one particular form of analytical technique, the use of bibliometrics. I'll then provide a

brief account of a large-scale peer review project in which I participated to give you a sense of a complex terrain and how we tried to map it, and then conclude with some comments about the particular perspectives and possibilities that both anthropology and anthropological societies can bring to questions of assessment.

Disciplining deliberation

Turning first to peer review, 'the practice by which the worth of research is evaluated by those with demonstrated competence to make a judgment' (British Academy 2007: ix), I should note that there is a great deal of variety in what is being reviewed, who the relevant actors are, how the review is conducted, and what the consequences might be. Collegial consultation, whether pursued long distance with individual respondents or taking place in face-to-face meetings, figures centrally in the allocation of postgraduate, postdoctoral and other fellowships (where those being reviewed are unlikely actually to be the peers of those doing the reviewing) and of research funds through the proposal evaluation process. Peer review is also central to the evaluation of manuscripts for possible journal and book publication, constitutes a critical element in individuals' personnel evaluations, and is key to departmental programme reviews. The UK RAE is perhaps the largest scale example of this last variety of peer review (Strathern 2006; Campbell 2006). Peer review combines individual judgement with collaborative deliberation, whether pursued through the mail, online or over several days in a windowless committee room. Indeed, we all are perhaps overly familiar with peer review, as we have been applicants, examinees and assessors many times over. At the same time, the very familiarity of the practice in its many forms can make it invisible, save as a source of exasperation, tedium and occasional satisfaction.

Peer review based at US government agencies has been the subject of considerable critical study. Some literature has concentrated on particularly flagrant ethical abuse, but most scholarship has been concerned with more everyday structural and procedural problems within the system (Chubin and Hackett 1990; Cole *et al.* 1978; Roy 1985). Peer review is sometimes seen as providing too much opportunity for self-interest, given the centrality of individual position taking. Over time, concerns about fairness (General Accounting Office 1994) have alternated with a sense that panellists are often not fully up to the task, as they might lack the degree of specialised knowledge and sophistication required to appraise cutting-edge science or scholarship responsibly. A perceived dichotomy between generalists and real experts has proven particularly significant in trying to explain the particular challenges of interdisciplinary panels (Feller 2006; Lamont *et al.* 2006; Lamont 2009; Brenneis 1999).

A particularly salient and consequential term in critiques of peer review is subjectivity, that is, the key role of individuals in coming to their own conclusions, even in contexts of intensive consultation and joint consideration. There is frequently a tension between this sense of individual judgement and concerns that such decisions make sense, that is, reflect appropriate outcomes reasonably reached. Reviewer subjectivities are, however, rarely in free play, as agencies very actively work to shape both written responses and panel discussions. At both the National Institutes of Health (NIH) and the National Science Foundation in the US, for instance, great attention is paid not only to the general criteria that should guide commentary and discussion

but also to the more specific rating systems in use. For example, panel meetings often begin with calibration exercises, that is, with trial scorings to see how one's ratings and scores align with those of others. One recurrent concern about subjectivity is that some individuals may always give high scores while others give low ones; calibration is one strategy to bring panellists into closer accord. On some kinds of panels even more explicit opportunities for self-monitoring and regulation are provided; on postgraduate fellowship panels, for example, participants are routinely informed about how many applications they have read and what their average scores are (for a detailed account, see Brenneis 1994). This ongoing struggle with subjectivity, to balance individual judgement with concerns for fairness, reasoned discussion, and good science and scholarship, is a hallmark of peer review in practice – and puts into play an often productive tension.

Two other elements figure in criticism of peer review. One has to do with its closed nature; only a few scholars or scientists are involved, and such small numbers in themselves might accentuate the influence of particular individuals. With the advent of electronic resources, a number of scholars have argued that larger scale peer review online would make for better results. Others (for example, Harnad forthcoming) suggest that peer review can now productively be augmented and enriched by continuing post-publication online peer commentary.

A final very significant concern is that highly innovative grant proposals are either actively stifled (Roy 1985) or for other reasons do not rise to the top. My own ethnographic research on funding panels suggests that the premium placed on amity around the committee table and a commitment to comparative discussion, one which particularly innovative and idiosyncratic proposals might make difficult, may indeed keep some of the most creative work from being supported (Brenneis 1999).

I want briefly to touch further on my own ethnographic work on peer review and note three key aspects of the funding committees on which I worked as a literal participant–observer. First is that peer review is deeply entangled with a complex sociality. 'Peer review' can actually be seen as multidimensional, as 'we (as panelists) were both reviewing the work of our peers and, in our discussions, concerned with being peers … Participation in such decision making made one, for the moment at least, an equal. In peer review we jointly constituted an ephemeral peership among ourselves as well as vis-à-vis those whom we were evaluating' (Brenneis 1999: 141). Further, panel discussions on the interdisciplinary panel on which I served for 3 years were marked by a remarkable degree of deference across disciplines; 'our discussions were generally characterised by amiability and a willingness to listen to what others had to say. Ironically, this willingness to listen often served to limit what one might say or how strongly one might be willing to say it … the highly collaborative nature of panel work makes disagreement difficult' (Brenneis 1999: 142). Such ongoing social process is likely to be no surprise to social anthropologists, but it has rarely been noted elsewhere in the literature.

A second comment has to do with the struggle with subjectivity noted above. As panellists we learned how to become disciplined readers and actors. Ewald has written of the norm as a 'principle of communication, a highly specific way of reducing the problems of intersubjectivity' (1991: 32). To varying extents – most markedly in the postgraduate fellowship panels – panellists are plunged into a normalisation process. In so doing, we 'acquire a new language in different ways: by attempting to standardize those criteria by which we evaluate proposals; by participating in a fiction of

objectifiability engendered, at least in part, by the negotiation of what constitutes [licit and relevant] "information;" and by the adoption of fungible categories of discussion and comparison' (Brenneis 1994: 32).

Finally, as panellists, applicants or referees, we are often engaged with a variety of documents, whether a recommendation form or a scoring sheet. These are, to use Richard Harper's (1998) term, 'mundane' documents, ones that engender routine responses, both from those filling them out and from their readers. Such documents are designed by staff in consultation with committees, circulate among and are given specific substance by individual scholars, and go on to play a major role in subsequent decisions. At the same time, they and their afterlives are usually, 'in large part because of their very ordinariness, analytically invisible' (Brenneis 2006: 42). My analysis of the social lives of such mundane forms has focused on several dimensions, including the forms themselves as both texts and frames and the kinds of 'doings with documents' (Harper 1998: 3) that recurrently take place within an institution and the ways in which documents both derive from and help constitute the work of that institution. These forms both require certain kinds of response and make others unlikely – or, at times, unthinkable. To borrow another term from Harper, this time from his research with Abigail Sellen (Sellen and Harper 2002: 16–18), these forms have very particular 'affordances': they enable and perhaps even require some activities and efface the possibility of others. And such forms are also instrumental in making possible the production of particular kinds of artefacts, for example, league tables, artefacts that often have significant careers of their own.

In its recent examination of peer review for assessing work in the Humanities and Social Sciences, the British Academy concludes that '(p)eer review has its critics, who allege that it is costly, time-consuming and biased against innovation. None of these criticisms is entirely without force, but the Working Group concluded that there were no better alternatives and that often the criticisms were directed at deficiencies of practice rather than the principle of peer review' (British Academy 2007: ix). Other institutions, for example the US National Institutes of Health in a major 2007 review, also argue that peer review remains the strongest option, especially if practices are reshaped to encourage more focused and consistent deliberation. And, as noted above, the US National Research Council sees combining deliberation with appropriate and multiple analytical strategies as optimal.

Analytical approaches

I begin here with a quote from Professor Eric Thomas, Chair of the Research Policy Committee, Universities UK. Universities UK released a major report (Evidence 2007) proposing a new set of evaluative practices to replace the Research Assessment Exercise (RAE), the last of which was held in 2008. In his foreword to this report Thomas wrote 'It is widely expected that the ratings will initially be derived from bibliometric-based indicators rather than peer review. These indicators will need to be linked to other metrics on research funding and on research postgraduate training. In a final stage the various indices will need to be integrated into an algorithm that drives the allocation of funding to institutions. The quality indicators would ideally be capable of not only informing funding but also providing information for higher education institutions and

stakeholders. They are also expected to be cost-effective to produce and should reduce the current assessment burden on institutions' (Thomas 2007: 2).

While this vision – and the virtues of clarity, comparability and administrative ease it is assumed to imply – is unlikely to be realised in the UK in the near future, a similar scheme is currently under very intense discussion and debate in France. Central to the Sarkozy government's plan to make the universities 'autonomous' along lines assumed to replicate North American institutions is a plan by which 'the merits of researchers will be judged by bibliometrics, with bibliometric exercises applied every four years to every university by experts in "scientific productivity." These measures will serve as a basis for the differential rewards, especially in terms of the ratio of teaching to research, which presidents of the newly autonomous universities can decide to allocate to the most deserving of their "employees" every four years' (Mallard personal communication February 2009; see also Foucaut *et al*. 2009a, 2009b). This purely metric evaluative scheme is clearly borrowed from the UK rather than the US, although it is assumed to be an American invention. It has been a key element in the university strike of the spring of 2009.

What are the key elements of the metrics schemes that Universities UK and President Sarkozy want to replace more deliberative evaluation? And why do some consider them such an attractive alternative to peer review? Quantitative data such as average years to degree and research dollars awarded often figure in such measures, but the core element is bibliometrics, 'using counts of journal articles and their citations' (Evidence 2007: 3). Such metrics measure raw productivity (without making judgements of quality) but also count instances of citation by other scholars, which are taken as proxies for or indicators of the paper's quality, and which can further be aggregated as journal impact scores as well. It is crucial that only citations within two calendar years of an article's publication are counted, a very brief window indeed, and one quite inappropriate given widely varying citational practices – and lead times – across the disciplines. Even within biology, for example, articles in molecular biology are much more rapidly cited than those in ecology; similarly, there is a much quicker turn-around for articles in biological anthropology than in social. A 5-year citation figure is now available as well, but 2 years remains the norm.

The Universities UK report claims that '(c)itations between papers are signals of intellectual relationships. They are a natural, indeed essential, part of the development of the knowledge corpus. They are therefore valuable as an external index about research because they are produced naturally as part of "what researchers do" and because they are related naturally to "impact" and "significance"' (Evidence 2007: 7). The journal impact factor, a score for individual journals 'created in the 1960s as a way to measure the value of journals by calculating the average number of citations per article over a specific period of time' (Adler *et al.* 2008: 6), is particularly important in ranking journals and thus determining the value of individual articles appearing in them. The data for these citation statistics are all drawn from the Thomson Scientific (formerly ISI, now Thompson-Reuters) indexes that cover '8,700 of the most prestigious, high impact research journals in the world' (Evidence 2007: i): 8,700 journals represent a very small proportion of world journals in the sciences and social sciences. In addition, the literature considered is overwhelmingly in English. Crucially for fields such as Social Anthropology, there is no attempt to index citations of books and other consequential genres of scholarly writing, including the high-prestige edited volumes so central to our field. A very recent debate at the 2008 Ljubljana EASA meeting highlights some of the

likely if unintended consequences of the Thompson Scientific system (Wulff *et al.* 2009). (For a very helpful review of the much broader range of 'research outputs' central in social science and humanities research see Huang and Chang 2008). It is also important to note that bibliometrics (and the broader field of scientometrics of which it is a part) 'were developed originally for exploring the working of the scientific enterprise, that is, as descriptive and analytical tools, not as evaluative or predictive ones' (Feller and Stern 2007: 100). Here a data base and approach developed for research were rapidly adapted for evaluation and management. This evokes what has come to be known as Goodhart's Law, named after the senior economist and advisor to the Bank of England who first made the observation that 'Any observed statistical regularity will tend to collapse once pressure is placed upon it for control purposes.' Or, in Marilyn Strathern's more user-friendly rephrasing, 'When a measure becomes a target, it ceases to be a good measure' (both cited in McIntyre 2000).[1]

The attractiveness of this approach, in the words of a recent critical report, lies in the fact that '(t)hose who argue for this simple objectivity believe that research is too important to rely on subjective judgments. They believe citation-based metrics bring clarity to the ranking process and eliminate ambiguities inherent in other forms of assessment' (Adler *et al.* 2008: 2). Apparently simple, clear, systematically related and ideally suited for commensurability, bibliometric assessment is also routinised, does not require the investments of time inevitable in peer review, and is relatively inexpensive. The fit with an administrative environment in which audit, transparency and accountability figure centrally is remarkably good.

Not surprisingly, such metrics have engendered considerable criticism; some of the strongest and most convincing arguments against reliance solely upon this approach have come from mathematicians, engineers and other scientists. A direct response to the Universities UK proposal is the September 2007 British Academy Report, *Peer Review: the Challenges for the Humanities and Social Sciences*. The report recommends that '(c)are should be taken to ensure that any metrics employed reflect the distinctive nature of the humanities and social sciences research and do not have an adverse effect on the quality of the work that they are seeking to measure' (British Academy 2007: xi). The report also argues that the Thomson Scientific data base is clearly inadequate for evaluating work in the humanities and social sciences and that ERIH, the European Reference Index for the Humanities currently under development, would require extensive modification to be of value. At present ERIH is still under revision; its scheme for categorising journals has met considerable critical resistance and is being reworked.

A committee organised by and representing the International Mathematical Union, the International Council for Industrial and Applied Mathematics, and the Institute of Mathematical Statistics has provided a particularly forceful response (Adler *et al.* 2008). This exceptionally cogent paper provides a rigorous, well-documented and daunting critique of the proposed move to metrics. Some of the criticisms echo those in the British Academy report; the 2-year window for citation counting, for example, proves as inappropriate for mathematics as it does for the humanities. There are also detailed analyses of gaps and errors in the statistical claims being made and of the limitations of specific measures. But most interesting for me as an anthropological reader was what the authors considered a key issue: 'Those who promote citation statistics as the

1 For a penetrating and thought-provoking examination of the increasing role of indicators in the domain of international human rights, see Rosga and Satterthwaite (2009).

predominant measure of research quality do not answer the essential question: What do citations mean? They gather large amounts of data about citation counts, process the data in order to derive statistics, and then assert the resulting assessment process is "objective." Yet it is the interpretation of the statistics that leads to assessment, and the interpretation relies on the meaning of citation, which is quite subjective' (Adler *et al.* 2008: 14). Drawing upon a particularly acute article on the rhetorical roles of citations (Cozzen 1989), they provide a brief but bracing discussion of the sociology of citations, noting that authors cite others for many reasons beyond recognising an intellectual debt. The authors conclude that 'citation data and statistics can provide some valuable information . . . (b)ut citation data provide only a limited and incomplete view of research quality, and the statistics derived from citation data are sometimes poorly understood and misused. Research is too important to measure its value with only a single coarse tool' (Adler *et al.* 2008: 3).

Despite critiques such as these, the Thomson Scientific indexes by themselves have long been a major tool for both individual and programme assessment in academia. They also play a significant role in generating a wide range of second degree artefacts, for example the kinds of tables that appear in such periodicals as *THE* (*Times Higher Education*). The 17 July 2008 issue of *THE* includes a figure entitled 'Top countries in sciences and social sciences based on impact' (*THE* 2008: 17). This league table lists in rank order 20 countries, the number of articles per country published in Thomson-indexed journals, the number of citations and the citations per paper. At the top of the list is Switzerland; while only 161,879 articles were published in the journals counted, the average citations per paper number 14.36. The US is second with an average of 13.76 citations, the UK lags at seventh and France at fifteenth. Several things are worth noting about the chart. First, it appears (as do similar representations of rankings in other, usually more limited domains, in other weeks) without any framing or qualifying commentary. In a periodical that frequently carries articles quite critical of such metrics, it is striking that such representations are just presented 'as is', strongly suggesting a taken-for-granted quality. Second, the high ranking of Switzerland might be somewhat unexpected. My own sense is that it reflects the Swiss location of CERN and related research centres. It is likely that such sites generate a fair bit of intra-institutional cross-citation. In addition, physics is a field in which citational turn-around is routinely quite rapid and where, therefore, the 2-year window actually makes sense.

A second, somewhat more complicated artefact is the 'Map of science' figuring in an article, 'Mapping the backbone of science' by Boyack *et al.* (2005: 364). This is a very complex figure in which the intensity of interdisciplinary cross- and co-citation among articles within the Thomson Scientific data base is represented through proximity. The more nearly adjacent the fields on the map, the closer their ties and more intense their interactions as measured by reciprocal citation. This map is one of several provided in an article intended to 'represent . . . the structure of all science, based on journal articles, including both the natural and social sciences, [one that] can be used to visually identify major areas of science, their size, similarity, and interconnectedness' (Boyack *et al.* 2005: 351). The authors see their project as part of a scientific study of the structure of science – clearly a research project in its own right. At the same time they highlight a 'desire to help the senior R&D [Research and Development] manager understand their enterprise and navigate their relevant environment . . . [p]otential actions on these maps (e.g., exploring new territory or reducing resources in existing territory) have a direct relationship to decisions that these managers must make. It is important that a science

map be as accurate as possible when used in a decision-making context within the S&T [Science and Technology] enterprise' (Boyack *et al.* 2005: 352). The authors do note that such maps should be used only in the context of broader gauge deliberative planning and assessment; at the same time, both the technical brilliance of their modelling and the assumed value of their data base could well make it very seductive as a free-standing instrument. This kind of mapping provides a fine example of what my colleague Melissa Cefkin (personal communication April 2008), an anthropologist at IBM, refers to as 'data base determinism' – if you have the data, use it. Here, though, the data reflect the very real limitations of the Thomson Scientific index: restricted to a relatively small range of articles of a particular sort, that is, research-based; articles that are cited in other indexed journals within a 2-year window; articles that are primarily in English; and, I would suggest, articles in publications that are online and searchable. One strategy for evaluating this particular map's accuracy is to locate anthropology and its closest disciplinary neighbours, which are neuroscience, meteorology, astronomy and palaeontology. How can we account for this siting? Here my personal hunch has to do with the fact that principal journals in biological anthropology and, to a lesser extent, in archaeology, have been available both online and searchable for considerably longer than those in social and cultural anthropology. Perhaps more significantly, citational practices in biological anthropology more closely approximate those in most natural sciences, with a relatively short time depth, while social and cultural anthropology has a longer turn-around time. The 2-year window central to the Thomson Scientific indexes effectively excludes the very real citational afterlives of our discipline's work and is key in generating such potentially consequential misrepresentations. Recent research has tried drawing upon different data sources. Bollen *et al*. (2009) acknowledge the limitations of the Thomson Scientific data base and turn rather to mapping 'clickstreams', sequences of online search link connections within such portals as JStor, Ingenta, Elsevier and the like. This may hold out some promise – at least anthropology finds itself in a somewhat more immediately compatible neighbourhood near archaeology, human geography and psychology – but thinking through these measures, possible unrecognised limitations in the web portals examined, and the affordances and limitations of the modelling strategies employed seems necessary.

Benchmarking UK social anthropology: a case study[2]

In 2005 the Economic and Social Research Council (ESRC), Association of Social Anthropologists (ASA) and Royal Anthropological Institute (RAI) collaborated in organising a Steering Committee to plan a study to 'benchmark the quality and impact of Social Anthropology research in the UK against international standards' (ESRC 2006: 3). The Steering Committee, chaired by John Gledhill, then ASA President, invited a number of non-UK anthropologists to conduct the review. Eight of us (from Australia, Norway, Mexico and the US) constituted the committee, which I chaired. The core objective of this review was to gain a clearer and empirically grounded understanding of how the quality, visibility, impact and broader contributions of British

2 The discussion in this section draws extensively upon the ESRC (2006) report, *International Benchmarking Review of UK Social Anthropology*, for which I both chaired the review panel and wrote the final assessment.

Social Anthropology figured within the field worldwide. While this benchmarking review was commissioned by the ESRC in collaboration with the two associations and was intended to be useful in ESRC's own strategic planning, both ESRC and the steering committee members clearly hoped that it would be of interest and value to other funding agencies, stakeholders, administrators and colleagues within the discipline – and that it could provide an informative public face for the field. It is important to note that, from its inception, both an institutional actor and the two primary scholarly associations were jointly involved in the project.

Two general elements of the charge to the visiting committee should be noted. First, this benchmarking was to be a review of the field as a whole within the United Kingdom, not a comparative evaluation of departments vis-à-vis each other, as in the RAE. A second key element of the benchmarking was that we were specifically asked to pursue it as a qualitative exercise. As our charge stated, 'The qualitative nature of the review is particularly important. It will help balance the UK Government's increasing use of metrics, especially citation counts, to make judgments about research impact and standing.' Or, as Ian Diamond, the Head of ESRC and a noted social statistician, commented at the Steering Committee meeting of 10 August 2005, relying 'on qualitative rather than quantitative measures [was necessary] in order to complement the assessment made by the Research Assessment Exercise (RAE) and avoid some of the problems which ESRC has already identified with existing quantitative indications such as citation indexes'.

One striking contextual feature of the UK scene was the long-term involvement of UK social anthropologists in and sophistication about government assessment and allocation institutions and procedures. It was clear that, especially when compared with the US, UK Social Anthropology had over the past 20 or so years developed an empirically rich, reflective and critical literature on the field itself within the context of higher education and research policy and practice, one deeply grounded in both social and institutional contexts. EASA has been instrumental in catalysing and developing similarly reflective and productive considerations of the state and possibilities of the field in a range of meeting sessions and publications, including the *Learning Fields* series edited by Dorle Dracklé, Iain Edgar and Thomas Schippers (Dracklé *et al.* 2003; Dracklé and Edgar 2004) and the germinal volume *Audit Cultures* (Strathern 2000).

Before the review we were provided with a very helpful, concise, yet wide-ranging set of statistical data on UK Anthropology (Mills 2006). One particularly helpful aspect of the statistical materials was that they were not presented as providing a global account. They didn't present the field in broadly comparable terms; rather they illuminated the local particularities of Social Anthropology. It was both distinctive and significant, for example, that Social Anthropology is an 'exporting field', that is, its PhDs are often hired by departments in other disciplines.

The central element of the review consisted of visits to 12 departments of Social Anthropology, departments that had been selected by the steering committee. For most of the visits we were in two smaller teams. Each of the 12 departments had provided its own background materials for the committee. In each of these visits we worked to explore the three core topics of central concern to ESRC and the steering committee: research issues, research capacity and impacts on policy and practice. We were able to spend three or four hours at each site, meeting separately with academic staff (or some portion thereof) and with postgraduate students and postdoctoral fellows.

Formats varied from school to school, with some departments making fairly elaborate presentations and others moving more directly into conversation with the committee.

As an international group, our panel was able to view the UK scene within an international comparative framework; what might be taken as routine and expectable by our UK colleagues could be – and often was – quite striking and distinctive to panel members. We also could, crucially, rely on our own expertise – years of reading in the field, conducting research, meeting with colleagues both in our home countries and internationally – in working towards a picture of the field in the UK.

We learned in the course of our various campus conversations that, when it comes to evaluations, one size does not fit all. When considered vis-à-vis the social sciences more generally, Social Anthropology is distinctive along several dimensions. The time required for design, preparation (including language learning), fieldwork, analysis and publication is considerable and likely to be longer than for other fields; expectations concerning rates of productivity and research 'turn around times' should accordingly be tailored to the field's temporalities. While the time required for anthropological research is considerable, the financial scale for funding such research is often much lower than in other social sciences (and certainly in the natural sciences). In financial terms, Social Anthropological research recurrently provides a real bargain, a factor that should be taken into account when thinking about what kinds of risks a funding agency might be willing – or eager – to take. We were not claiming here that Social Anthropology is singular in its singularity. This was rather to argue that review measures and practices take into account the local particulars of each discipline or interdisciplinary cluster – and to think through strategies appropriate for and likely to be informative about each field.

The visiting committee noted in our report that, in our view, many assessment practices were problematic not because they were quantitative per se but because they were restrictively and reductively so. Taking a much wider range of more subtle, countable evidence into account, however, can complement qualitative findings and help substantiate and refine them. As a simple example, relying on aggregate grant income as a measure of research quality makes little sense in a field in which, for multiple reasons, scholars can pursue relatively inexpensive research. As the report notes, 'output measures such as the number of proposals funded or the range and quality of resulting publications' (ESRC 2006: 9) would be more informative. Better and longer term tracking of the postgraduate careers of PhDs well beyond their first employment would effectively help speak to both training and impact. A recently released study of US PhDs in six social science fields, five and more years after their degrees (Nerad *et al.* 2008) provides both a substantively rich account of contemporary careers and of the fit – or misfit – between preparation and professional trajectory and an invaluable model of principled and illuminating quantitative review.

To quote our report at some length, several factors seemed critical for an appropriate and effective review: 'First, any measures and evaluations of a discipline should be understood vis-à-vis the specific contexts and characteristics of that discipline. One size doesn't fit all, and the meaning of any measurement is certain to be discipline-sensitive. Second, proxies and indicators can be mischievous; letting one variable stand for an entire universe often leads to real misunderstanding. And any measurable variables should be considered not only in relation to each other but also, more significantly, as a complement to broader qualitative findings. Third, comparative perspectives, whether cross-nationally within the same discipline or across disciplines within a specific national

context, are invaluable, in large part because of the unexpected and revelatory moments of recognition or, occasionally, surprise that they afford. Finally, central to the success of a large-scale qualitative review such as the UK Benchmarking exercise is that it involves a group of colleagues – reviewers and reviewed alike – in serious inquiry, reflection, and consultation' (ESRC 2006: 24–5). Finally, I want to stress the value of joint engagement of relevant administrative administrators and the scholarly societies in shaping the project all along. This kind of collaborative work made the necessary tailoring of the review to fit the discipline possible and led to a much more complex and, I think, realistic account of the field.

Accounting for anthropology

I want briefly to suggest here that we as social anthropologists can bring a particularly valuable sensibility to bear on the current moment in academia and that, jointly, we may in fact be in a position to make a difference within those institutional structures within which we live our professional lives. This moment is marked by the rise of a range of formal analytical measures but also of increasing critiques of their limitations and problematic features. Similarly, the language of accountability and audit is widespread, but so too is an increasing call for human actors to be responsibly – and collaboratively – engaged in making judgements drawing upon a complex range of qualitative and analytical information. Interventions may well be possible, and we should be able to contribute to them.

Key to these possibilities is anthropology's simultaneous commitment to and sense of the complex challenges posed by translation – across cultures, communities and languages. In large part this is because both the meanings and the social resonances of communicative practice are deeply context-dependent; meanings are rarely invariant across multiple sites, and the more abstract the concept, the less likely its stability. In the context of contemporary assessment practices, it is crucial that we as anthropologists work to account for and convey our own scholarship without losing the particularities and texture that give it distinctive value. Figuring out how we can effectively pursue mutual comprehension without formal commensurability, recognition without ranking, is a real challenge, but our field does afford the analytical and empirical capacities that can make such translation possible. Social anthropologists, especially through organisations such as EASA, are in a position to try to change the terms in which current institutional conversations are being conducted – and in so doing help shape the possibilities of the next 20 years and beyond.

Acknowledgements

This article was originally presented as the Distinguished Lecture at the 20th anniversary conference of EASA, 6 February 2009, Vienna. I want to thank then President, Shalini Randeria for the invitation to speak, Andre Gingrich and Thomas Fillitz for their – and the University of Vienna's – warm hospitality, and many EASA colleagues at those meetings for their comments and conversation. I would like to thank both EA SA and the Austrian Academy of Sciences. Thanks also to Dorle Dracklé and Helena

Wulff for the invitation to revise my lecture for publication in *Social Anthropology*. I have been particularly fortunate in the many colleagues I've been in conversation with over the course of the broader project underlying this paper, especially Marilyn Strathern, Cris Shore, Susan Wright, Andres Barrera, Bonnie Urciuoli, Oona Schmid, Melissa Cefkin, Davydd Greenwood, Marc Abeles, Susana Narotzky, Gregoire Mallard, Michele Lamont, Andre Gingrich, Rena Lederman, George Marcus, Alan Scott, Fred Turner, Mary Huber and Irene Bellier. Thanks also to Melissa Caldwell and John Bowen for comments on the original talk.

Don Brenneis
Department of Anthropology
University of California, Santa Cruz
Santa Cruz, CA 95064
USA
brenneis@ucsc.edu

References

Adler, J., J. Ewing and P. Taylor 2008. *Citation statistics: a report from the International Mathematical Union in Cooperation with the International Council of Industrial and Applied Mathematics and the Institute of Mathematical Statistics*. International Mathematical Union (www.mathunion.org/Publications/Report/CitationStatistics) Accessed 31 May 2009.

Bollen, J., H. Van de Sompel, A. Hagberg, L. Bettencourt, R. Chute, M. Rodriguez and L. Balakireva 2009. *Clickstream data yields high-resolution maps of science*. Public Library of Science (www.plosone.org/article/info:doi%2F10.1371%2Fjournal.pone.0004803) Accessed 31 May 2009.

Boyack, K., R. Klavans and K. Börner 2005. 'Mapping the backbone of science', *Scientometrics* 64: 351–74.

Brenneis, D. 1994. 'Discourse and discipline at the National Research Council: a Bureaucratic Bildungsroman', *Cultural Anthropology* 9: 23–36.

Brenneis, D. 1999. New lexicon, old language: negotiating the 'global' at the National Science Foundation, in G. Marcus (ed.), *Critical anthropology now*, 123–46. Santa Fe: School of American Research Press.

Brenneis, D. 2004. 'A partial view of contemporary anthropology', *American Anthropologist* 106: 580–8.

Brenneis, D. 2005. Documenting ethics, in L. Meskell and P. Pels (eds.), *Embedding ethics*, 239–52. Oxford: Berg.

Brenneis, D. 2006. Reforming promise, in A. Riles (ed.), *Documents: artifacts of modern knowledge*, 41–70. Ann Arbor: University of Michigan Press.

Brenneis, D., C. Shore and S. Wright (eds.) 2005. *Universities and the politics of accountability*, special issue of *Anthropology in Action* 12 (1): 1–78.

British Academy 2007. *Peer review: the challenges for the humanities and social sciences*. London: The British Academy (www.britac.ac.uk/reports/peer-review) Accessed 31 May 2009.

Campbell, D. 2006. 'The Research Assessment Exercise 2008 in the United Kingdom', *Plattform: fteval* 28: 32–47.

Chubin, D. and E. Hackett 1990. *Peerless science: peer review and US science policy*. Albany: SUNY Press.

Cole, S., L. Rubin and J. Cole 1978. *Peer review in the National Science Foundation*. Washington, DC: National Academy Press.

Cozzen, S. 1989. 'What do citations count? The rhetoric-first model', *Scientometrics* 15: 437–47.

Dracklé, D. and I. Edgar (eds.) 2004. *Current policies and practices in European social anthropology education*. Learning Fields Vol. II. Oxford: Berghahn.

Dracklé, D., I. Edgar and T. Schippers (eds.) 2003. *Educational histories of European social anthropology*. Learning Fields Vol. I. Oxford: Berghahn.

ESRC 2006. *International benchmarking review of UK social anthropology*. Swindon: Economic and Social Research Council.

Evidence 2007. *The use of bibliometrics to measure research quality in UK higher education institutions*. London: Universities UK.

Ewald, F. 1991. Norms, discipline, and the law, in R. Post (ed.), *Law and the order of culture*, 138–62. Berkeley: University of California Press.

Feller, I. 2006. 'Multiple actors, multiple settings, multiple criteria: issues in assessing interdisciplinary research', *Research Evaluation* 15: 5–15.

Feller, I. and P. Stern (eds.) 2007. *A strategy for assessing science: behavioral and social research on aging*. Washington, DC: The National Academies Press.

Foucaut, M., E. Lépinard, V. Lepinay and G. Mallard 2009a. 'Une évaluation transparente et collégiale', *Le Monde* 4 February 2009.

Foucaut, M., E. Lépinard, V. Lepinay and G. Mallard 2009b. 'Pour les universités plus juste', *La Vie des Idees* (www.laviedesidees.fr/Pour-des-universites-plus-justes.html) Accessed 31 May 2009.

General Accounting Office 1994. *Peer review: reforms needed to insure fairness in Federal Agency grant selection*. Report to the Chairman, Committee on Government Affairs, US Senate.

Harnad, S. forthcoming. The post-Gutenberg open access journal, B. Cope and A. Phillips (eds.), *The future of the academic journal*. Cambridge: Chandos.

Harper, R. 1998. *Inside the IMF: an ethnography of documents, technology, and organizational action*. San Diego: Academic Press.

Huang, M. and Y. Chang 2008. 'Characteristics of research output in social sciences and humanities: from a research evaluation perspective', *Journal of the American Society for Information Science and Technology* 59: 1819–28.

Lamont, M. 2009. *How professors think: inside the curious world of academic judgment*. Cambridge: Harvard University Press.

Lamont, M., G. Mallard and J. Guetzkow 2006. 'Beyond blind faith: overcoming the obstacles to interdisciplinary evaluation', *Research Evaluation* 15: 43–57.

Lederman, R. (ed.) 2006. 'AE forum: IRBs, bureaucratic regulation and academic freedom', *American Ethnologist* 33: 477–548.

McIntyre, M. 2000. 'Audit, education, and Goodhart's Law, or taking rigidity seriously' (www.atm.damtp.cam.ac.uk/peopme/mem) Accessed 24 February 2009.

McSherry, C. 2001. *Who owns academic work? Battling for control of intellectual property*. Cambridge: Harvard University Press.

Mills, D. 2006. 'Briefing document for the International Benchmarking Review of Social Anthropology. ESRC (www.esrcsocietytoday.ac.uk).

Nerad, M., E. Rudd, E. Morison and J. Picciano 2008. *Social science PhDs – 5+ years out: a national survey of PhDs in six fields*. Seattle: CIRGE (Center for Innovation and Research in Graduate Education), University of Washington.

Power, M. 1994. *The audit explosion*. London: Demos.

Power, M. 1997. *The audit society: rituals of verification*. Oxford: Oxford University Press.

Rosga, A. and M. Satterthwaite 2009. 'The trust in indicators: measuring human rights', *Berkeley Journal of International Law* in press.

Roy, R. 1985. 'Funding science: the real defects of peer review and an alternative to it', *Science, Technology, and Human Values* 10: 73–81.

Sellen, A. and R. Harper 2002. *The myth of the paperless office*. Cambridge: MIT Press.

Shore, C. and S. Wright 1999. 'Audit culture and anthropology: neoliberalism in British higher education', *Journal of the Royal Anthropological Institute* 5: 557–75.

Strathern, M. 1997. 'From improvement to enhancement: an anthropological comment on the audit culture', Founder's Memorial Lecture, Girton College, Cambridge University, 22 November.

Strathern, M. (ed.) 2000. *Audit cultures: anthropological studies in accountability, ethics, and the academy*. London: Routledge.

Strathern, M. 2006. 'A community of critics? Thoughts on the new knowledge', *Journal of the Royal Anthropological Institute* 12: 191–209.

THE 2008. 'Top countries in sciences and social sciences based on impact', *THE (Times Higher Education)* 17 July 2008: 19.

Thomas, E. 2007. 'Foreword', in *Evidence, the use of bibliometrics to measure research quality in UK Higher Education Institutions*, 2. London: Universities UK.

Winston, G. 1999. 'Subsidies, hierarchies, and peers: the awkward economics of higher education', *The Journal of Economic Perspectives* 13: 13–36.

Wulff, H., D. Handelman, M. Berghahn, V. Dominguez and B. Moeran 2009. 'What is happening to the anthropological monograph? Live debate at the EASA conference in Ljubljana 2008', *Social Anthropology/Anthropologie Sociale* 17: 217–34.