'The Limits of the Numerical'

Cambridge University, University of Chicago,
University of California, Santa Barbara

Strand 3: Higher Education (pp. 16-23)

University of California, Santa Barbara
Christopher Newfield
Bishnupriya Ghosh

INTRODUCTION TO THE PROJECT

This project consists of three teams at three universities exploring one of the most pressing sets of questions for the modern humanities and social sciences and their relation to policy. The questions are: What are the *effects* of the introduction of numerically-based quantification into all aspects of social and cultural evaluation? *How* do the uses of numerical calculation achieve their effects? Does quantification *work against the values it is introduced to support*, promote or capture? When this happens, *how can quantification be revised* to fit better with a system's purposes and its participants' needs?

The structure of the project offers a unique experiment in interdisciplinary research. Three strands of research will be conducted by a team of two post-docs, a PI, a co-PI, and an academic advisory board. Each team will be located in a major research university – the University of Chicago, Cambridge University, and the University of California at Santa Barbara. The three teams will cover three main strands of social policy: climate change, healthcare, and higher education. The teams will come together for a month on four occasions to run shared seminars and early career research workshops. Research will be developed and conducted in a collaborative manner, both within each strand and between the different strands. This model offers the rare possibility of bringing together detailed academic research in a broad, interdisciplinary, international framework.

UC Santa Barbara will focus on *higher education*. Its PI is Christopher Newfield, a professor of literature and American Studies in the English Department who has just completed a third book on higher education's purposes and policies and who has an international reputation as a humanities analyst of the future of public universities. The Co-PI is Bishnupriya Ghosh, a professor and chair of the English department who has an international reputation in post-colonial studies and is a founder and leader in the analysis of risk cultures.

The UCSB team will report on four areas in which quantification has changed university teaching, research, administration, and public policy. They are *(1) university rankings, (2) calculations of returns on investment in BA majors, (3) quantified learning outcomes, and (4) research bibliometrics in qualitative fields.* All reflect efforts to create indicators that measure output and productivity, and that can serve as the basis of objective evaluation. Areas (1) and (2) have come to mediate the public understanding of higher education, and their operations and variations need clarification. In areas (3) and (4), quantification has been transforming the core university missions of teaching and

research. We have four research questions that correspond to the reseach questions posed by the two other groups.

The overall project, *The Limits of the Numerical*, will synthesize findings from the three groups. In so doing the three project teams expect to produce both new methodological approaches and novel substantive conclusions on the question of the numerical and its limits in the human sciences.

1. THE RESEARCH QUESTION OF THE PROJECT

It is fundamental to the design of the project that these three strands are integrated to create a single intellectual plan for the whole project. Thus the project as a whole has a single overarching question that motivates its different strands. The **PROJECT RESEARCH QUESTION** has **three parts**:

I.   What are the effects on a system of social policy when numerical quantification and evaluation is introduced into that system?

II.   How does the use of numerical evaluation exclude, trivialize or distort other systems of political, moral and social evaluation?

III.   What are the political and moral consequences of this shift towards numerical evaluation?

2. THEORY AND METHODOLOGY (ALL THREE CENTRES)

The research questions will be explored in the three areas of higher education, climate change and healthcare, because these are the areas where social science, policy, and the gritty world of politics interact with intense urgency. There are no arenas where citizens and their behaviour are more implicated in social policy. Our central question faces a pressing and real-life problem full on.

I.   The first part of our central question starts from the fact that numerical evaluation is introduced into a system of social policy. It is introduced for good reasons to do with epistemic rigour, comparability across different jurisdictions, and the tractability of mathematical figuration for modeling and formalization. However, as the term "introduced" stresses, the introduction of any particular scheme of quantification, indeed quantification at all, is neither inevitable nor natural. Therefore, we want to explore why regulators seek to quantify, how these reasons shape the tools they use, and how these tools might misrepresent the things they purport to measure. Such work will draw both on existing analyses of the drive to quantification (Porter, 1996 Power, 1999) and on work on the uses, nature and meanings of measurement in the social sciences (Alexandrova, 2012).

But, on the other hand, we also stress 'system'. A common ground between such diverse intellectual traditions as Chicago school "law and economics" (Lucas, 1976) and "social constructivist" analyses of psychiatric categorisation

(Hacking, 1998) is that the introduction of quantitative tools can have unintended, sometimes perverse, effects on complex systems. It is crucial that the object of analysis here combines institutional frameworks, decision making bodies, normative discourse, planning agencies, authority figures and those who are in the system as patients, citizens, pupils as well as teachers, doctors, scientists, politicians. What is needed is an integrated analysis that moves beyond questions of policy or philosophy to stress the interaction of the different elements of the system. In analysing these interactions, we will draw on a wide variety of tools, most notably work in network theory, systems analysis, and the anthropology of institutions (Yearley, 2004). As the response to Latour and Woolgar's analysis of the science lab reveals, such a process is likely to reveal the interconnections that the system's smooth running needs to keep obscure (Latour and Woolgar 1986). This work is important not only because of the consequences to patients or pupils, say, of introducing numerical evaluation, but also because it reveals the consequences of quantification for policy makers and practitioners (and their vested interests) within the systems.

II.   The second part of our central question draws out an issue that is often obscured both in theory and in practice. There is, of course, a well-known and well-developed set of philosophical debates on the nature of value and evaluation, not only in ethics and political philosophy following Rawls' seminal work (Rawls, 1971), but (relevantly to this project) in epistemology (Zagzebski 1996; Williamson 2000; Steup ed 2001; Pritchard 2005; Alston 2005). While these abstract discussions will, of course, inform our analysis, we hope to focus on a less commonly debated question. In a democratic, pluralist society, there is bound to be widespread disagreement over which values we should seek to promote, how they should be balanced and so on. When we choose to use a particular quantitative metric to assess and guide policy, we risk of downplaying, trivialising or simply ignoring value considerations which the particular metric does not measure, and which, perhaps, could never be quantified at all. Numbers can be used to ignore, trivialize, and distort other systems and ideals of evaluation; and the different rhetorical and political strategies of this process within a system demand further discussion.

This part of the question involves two approaches: on the one hand, it involves some theoretical, philosophical questions about whether all values are commensuarable (Peterson, 2013) and on how to balance competing value systems in democracies (Fisch and Benbaji 2011); on the other, there are empirical test-cases to be explored of actual policy consequences of particular quantative tools. We shall look at how regulative systems use numbers rather than other systems of evaluation within a set of test-cases. This work will build on important work analysing the relationship between different forms of expertise (Wynne, 1996). A fascinating phenomenon in the history of science is "Kuhn-loss", where a new paradigm cannot answer problems or questions which a superseded paradigm could resolve (Post, 1971). Systems of quantification seem to face an analogous problem: similar to a Kuhnian paradigm, any system of quantification will soon seem "obvious" and "natural" to those who use it, but, in fact, any such system will only be an incomplete approach to messy reality. Test cases will allow us to identify the features of

such processes, while philosophical work will allow us to place these studies within a broader perspective.

III.   The first part of the project looks at the often aggressive effects of introducing quantification, including their unexpected costs; the second looks at a very particular kind of cost of quantification, the distortion or downplaying of incommensurable values; the third, which will be dominant in the last part of the project, aims at a more constructive goal: the theoretical and practical possibilities of change. In this context, we hope to address at least two issues. First, if some systems of quantification are necessary for the smooth or just functioning of modern societies, what makes for a good system (or set of systems) of quantification (Nord, 1999)? In answering this question, our aim is not so much to adjudicate disputes between competing approaches, but, drawing on sociological and philosophical work (for example, Daniels, 2008), to suggest how they might be adjudicated. Second, more broadly, given the limits to quantification, what alternative ways of decision-making might we adopt, and how might these alternative approaches draw on or relate to quantitative metrics? Again, this task involves not only analysis of the rich philosophical literature on such topics as the relationship between abstract philosophing and policy needs (Wolff, 2011), but a close analysis of the successes and failures of alternative systems of governance and policy-making.

Finally, the project needs to be integrally comparatist as well as interdisciplinary and collaborative. To this end, we have deliberately chosen three different Anglophone jurisdictions where systems of public health, public/private higher education and attitudes and institutions of climate change are different but interrelated. We believe that the three strands together tell importantly comparative stories that allow a far richer picture to emerge than any one study could reveal. So, the national politics of universities contrast with the global politics of climate change. The personal and national politics of health care contrast with the personal and national politics of education, not least in terms of visible results: an education for life is not viewed in the same light as an operation to save a life. The aim of this project is that its final outputs should have a potentially significant impact in policy arenas, both within and beyond our chosen jurisdictions.

3. INTERDISCIPLINARITY OF THE PROJECT

One of the most exciting aspects of the project is not just its novel structure but also its insistent requirement of a new methodological approach to the problems. Social science relies on the success of the numerical and we will need to turn the tools of social science towards their own processes and bring to bear on them both a rigorous philosophical and critical theoretical method. At the same time, while it would be premature to overdetermine the form or results of the research, it is clear that each strand of the project must be based on detailed empirical research.

I.   **Test cases**

Although the final aim of the project is work that is both general and has policy implications, we will also develop and explore twelve detailed test cases over

three years, four from each jurisdiction (England, USA, Canada), chosen collaboratively by the team members. The initial research will also require preliminary collection of the relevant data, and the evaluation of competing test-cases. These test-cases will be discussed in the four meetings, providing a tight and manageable focus for the group discussions.

Each test-case will require at least:

(a) a <u>historical account</u> of the introduction of numerical evaluation into a system.

(b) The contextualized <u>rationale</u> and explicit arguments for such introduction.

(c) the <u>effects of such numerical evaluation</u>, as perceived by actors and publicly, and as possible to determine by research.

(d) the <u>potential for change</u> within the schemes of evaluation.

(e) the <u>policy implications</u> of the research.

II. Analyses

Each strand must also develop a critical and historical analytical model in order to formulate and investigate the broader issues. Each strand will have its own requirements, but in coming together certain approaches must also be shared. This will require at least:

(a) a <u>critical understanding</u> of how such metrics affect such institutional and discursive systems: here recent history and sociology of science with its interest in network theory, system theory, and the role of automation in everyday life (Latour, Schaffer, Siskin and Warner, Castells) provide an initial approach to understanding the impact on a system of an introduction of numerical metrics.

(b) an <u>anthropologically-inflected analysis</u> of how numerical metrics change places of work: the analysis needs an anthropological understanding of social process to explore changes of behaviour and the relation between public discourse, normativity and behaviour within a community (Mackenzie 2008).

(c) a <u>philosophical understanding</u> of the interplay between the normative work of a normative system and the role of evaluation (feedback systems) within ethics. The question of competing value systems needs the clarifications of a philosophically trained analysis. Here, we start from the post-positivist resurgence of normative ethics, which already contains a small, but rich, literature on the relationship between meta-ethics and policy tools (Anderson, 1995). Furthermore, we hope to build on recent work in the philosophy of science which questions the possibility of value-free science (Douglas, 2009).

(d) The combination of anthropology, history of science, critical theory, philosophy and social science is an interdisciplinary goal rarely achieved. This is for good reason: members of different disciplines use different methodologies and vocabularies. However, we believe that the structure of the project and the methodology proposed make this goal achievable, by creating time and space for informed and mutually respectful discussion. Furthermore, the success of recent projects, most notably Brown's work combining sociology of science with normative political theory (Brown, 2009), and the growing field of "ordinary ethics", which combines anthropological insight with meta-ethical debate (Lambek, 2010), suggest the time is ripe for such inter-disciplinary work. Note, furthermore, that a project which investigates the importance of a plurality of tools of assessment will naturally lend itself to use of a plurality of methodological tools and approaches.

RESEARCH PROPOSAL FOR <u>PROJECT STRAND 1</u> AT THE UNIVERSITY OF CHICAGO

Climate Change is at once a problem of planetary proportions as well as one that impacts on lives of individuals and localities, making both action and inaction possible at the same time. While global negotiations may often prove intractable, many local and individual actions remain possible. This highlights the importance of scale in any intellectual and political attempt to grapple with climate change. It is a problem that can lead to discussions of individual responsibility as well as governmental and intergovernmental policies, big science as well initiatives based on local knowledge. This constant need to move between different scales of thought and action is a fundamental premise of this project. Democratic societies looking for solutions to the climate crisis need to create sites where conversation may proceed on how to move and translate between and across scales that may seem incommensurable.

Numbers, models, statistical tables, calculations of risk, probability, and uncertainty are central to the way scientists and policy makers define the very problem of climate change. But numerical imagination has to be translated into terms available in everyday life in societies where governmental or even municipal plans for action derive their legitimacy, ultimately, from public opinion. Even scientists who think of climate on a planetary scale feel the pressure to do some cultural translation into terms that will make sense to individuals who are not trained to think with abstract scientific tools. Thus the NASA scientist James Hansen calls his book on climate change *Storms for My Grandchildren* in order to bring it within the grasp of everyday heuristics available to most individuals (Hansen 2009). A similar move is made by geophysicist David Archer who opens his book, *The Long Thaw*, aimed at communicating to a general reading public the urgency of action needed on climate change on a planetary level. Archer directly confronts questions of inter-generational ethics that straddle and illustrate the problem of temporal scale. If indeed our greenhouse gas emissions are changing the climate of the planet for the next hundred thousand years, as Archer shows, how many generations beyond us should we – or even can we – really care for when we think of the planet as a whole? Our capacity to thus care, something that has evolved over a long period of time, may not be unlimited. Trying to explain as to why we "mere mortals" *should* "worry about altering climate 100,000 years from now," Archer has to step beyond the limits of his science and ask his reader to place herself in a moral-historical scenario: "How would it feel if the ancient Greeks...had take advantage of some lucrative business opportunities for a few centuries, aware of potential costs, such as, say a stormier world, or the loss of 10% of agricultural production to rising sea levels – that could persist to this day?" (Archer 2009: 9-10). This project aims to address self-consciously and theoretically – from within the traditions of humanities scholarship – three key problems the climate crisis makes visible with regard to the numerical imagination: (a) the authority of numbers, (b) the politics of numbers, and (c) the problem of translating very large and abstract numbers into ideas that can speak to human experience and thus motivate action.

We now briefly describe the kind of questions that may be raised in these three areas of investigation.

The Authority of Numbers in the context of climate change: The problem of

planetary climate change cannot be defined without the work that numbers do. From measuring and modeling for the planet's temperature, both past and present, to thinking of the effects of climate change in terms of decades, centuries, and millennia – at every stage of the definition of the problem, numbers are involved (Pierrehumbert 2011). So what is the work that numbers actually do for the general public as distinct from what they do for the specialist? Do they create and establish the "reality" of what the literary critic Timothy Morton has recently described as hyper-objects (such as planetary climate), entities that do not exist as physical objects as such but to which we ascribe object-like properties in order to conceptualize and deal with them (Morton 2013)?

The Politics of Numbers in Climate Action and Climate Skepticism:
Numbers are central to the many political positions that have evolved around climate change, from activism to skepticism. Thus, a major sense of climate injustice finds expression in the calculation of per-capita emissions across nations (Agarwal and Narain 1991). Climate skeptics, as Naomi Oreskes and Eric Conway showed, also use numbers to shore up their positions. The so-called Climategate scandal actually accused scientists
of deliberately distorting numbers (Mann 2012). Debates between economists about the cost of mitigating climate change turn on assumptions about the most suitable discount rate, once again a number (Pearson 2011).

Translating Numbers - Climate as Affect:
It is clear that climate numbers and statistics do not, by themselves, galvanize people into action. Popular interest in climate change varies over time. One could do some very interesting case studies to see where and under what circumstances scientific positions on climate change have enjoyed some traction in public life. What convinces people to be pro-active or political with regard to climate? In 2007, for instance, climate change became a "burning" political issue in the Australian national elections; it has since lost that position. What makes for these changes? Apart from such case studies, the project will also look at the role that cinema, science fiction and other forms of fiction, photography etc. can play in translating climate science (or even skepticism of that science, as in the case of Critchton 2004) into positive or negative political affect (Banerjee 2012).

The science of climate change is, ultimately, an invitation to think on a planetary level. It does not argue against local forms of environmentalism or action, but what distinguishes it from previous Green politics is the planetary scope of its imagination. Our strand in this project is about exploring how scholars in the humanities can respond to that challenge.

Specifics of Chicago Organization

Chicago has established an advisory board for the project that consists of five senior academics who have been involved in the composition of the proposal, and will be involved in the appointment and management of the project members. These are:

Professor James Chandler (formal PI): Chandler is a renowned scholar of literature and cinema who has long been interested in issues of inter-disciplinary conversation. He has organized conferences on "questions of evidence," methods of defining "cases," and has led the Center for Disciplinary Innovation at the University of Chicago.

Professor Dipesh Chakrabarty (formal PI): Chakrabarty is a leading scholar of postcolonial theory and history who has led the conversation among humanities scholars on the intellectual challenges of climate change.

Professor Fredrik Jonsson (formal Co-I): Jonsson is an environmental historian whose work on the Scottish highlands has elicited high praise. He is currently working on "climate change and history" and has co-taught courses on the subject with Chakrabarty under the aegis of the Center for Disciplinary Innovation at the University of Chicago.

Professor Kenneth Pomeranz (formal Co-I): Pomeranz is a globally renowned scholar of China, world history and environmental history in the eighteenth and nineteenth centuries. He has recently been working on issues of "big" history and at the interface of environmental history and climate studies.

Professor Joe Masco (formal Co-I): Masco is an anthropologist whose work on the politics of nuclear weapons and "controlled explosions" has been widely acclaimed and who has now begun research on geo-engineering in the context of climate change.

Other Chicago scholars associated with the project will be: Professors Kathleen Morrison (Anthropology), Mark Lycett (Global Environment Program), David Archer (Geophysics), Ray Pierrehumbert (Geophysics), Elizabeth Moyer (Geophysics), Eric Posner (Law), and David Weisbach (Law).

The Board and the larger group will meet regularly with the PIs and the Co-Is and post-docs to help direct research. They will also attend seminars and colloquia where suitable. We would like to host a series of short-term visitors. The short-term visitors typically would be climate change scholars who work with numbers but are prepared to engage scholars in the humanities and interpretive social sciences in discussions of what numbers do and do not communicate.

RESEARCH PROPOSAL FOR <u>PROJECT STRAND 2</u> AT THE UNIVERSITY OF CAMBRIDGE

I. THE RESEARCH BACKGROUND FOR STRAND 2

This strand takes a historical, anthropological and philosophical approach to the question of how quantification affects systems by looking at the example of QALYs in the British Public Health system. The historical perspective is necessary for two purposes: first, to understand and contextualize the arguments that successfully established QALYs as the prime means of evaluating resource decisions; second, to understand the overall shift in the institutional practices over time from before the adoption of QALYs to their current dominance. The anthropological perspective also has two lines of enquiry. First, it will explore the effect of the introduction of QALYs on the actors within the institution; second, it will investigate the interface between the institution and its social context, and how this is affected by the use of QALYs. Finally, there is a double angle to the philosophical enquiry, too. On the one hand, through an analysis of the principle and functioning of QALYs we wish to broach the central question of the overall project, namely, how quantification, especially numerical quantification, alters normative social systems. On the other, we wish to map out the anxious terrain between ethical judgment and quantified evaluation in health care as part of social policy with an eye to articulating broader concerns about how quantification can distort or ignore important value considerations. It should be immediately evident that these three areas of enquiry overlap significantly. It is also the case that although there is a bibliography on each of these topics, not only is there much still to be done on each question discretely, but also there is no research group that is looking at this crucial and, in our judgment, integrated set of questions, either within a single jurisdiction or transnationally. Furthermore, although there has been much collaborative work between philosophers and economists on ways of improving particular tools of quantification, there is far less work on the fundamental problem of why and whether to quantify at all. Indeed, we claim that it is only through such a collaborative, interdisciplinary combination of empirical historical analysis, with a detailed account of changing networks of actor interaction, and with a philosophical appreciation of the complex normative tensions involved, that anything like an adequate approach to the central concerns of modern society regarding quantification, as exemplified by the case of QALYS, can be articulated and addressed.

In health care, the UK has led the world in basing allocation of scarce resources on quantitative economic models. From its inception, the National Institute of Health and Clinical Excellence (NICE) has assessed novel interventions in terms of Quality-Adjusted-Life-Years (QALYs). This has allowed for a "rational" model for allocating scarce resources: broadly, if an intervention costs less than £20,000 per QALY produced, it should be funded; if the intervention costs more than £30,000 per QALY, then it should not be funded; if an intervention costs between £20,000 and £30,000 then further deliberation is required before a funding choice is made (NICE, 2013). This model for rationing resources has been extremely controversial, and, perhaps as a result, there has been a steady accumulation of exceptions to the broad framework (Steinbrook, 2008). More fundamental changes might well be afoot. The highly contested UK government reforms to the National Health Service – currently opposed by the British Medical Association – will shrink NICE's role to that of a technical advisory body, as local physicians and clinical commissioning groups come to play a

central role in deciding which health interventions get funded. At the same time, however, QALYs continue to be used in UK policy-making; for example, by the newly founded "Public Health England", which has a remit to consider determinants of health outside the provision of healthcare.

In general terms, no one would deny that health care is a foundational element in contemporary society's sense of itself and its notion of crisis as much as of well-being, and thus it is a key battleground of social policy. But the moment for a re-evaluation of the general question of the effects of quantification and specifically the role of QALYs could not be more timely. The first generation of QALYs are being re-considered from at least two directions: on the one hand, health care economists and policy makers, while insisting on their continuing usefulness, are refining and extending the models utilized; on the other hand, ethicists are beginning to put pressure on QALYs as an adequate response to the need for judgment within health care policy. This is a prime moment to make a telling contribution to an on-going debate, as well as to use this particular debate as the focus for more wide-ranging reflection.

Apart from official guidelines and reports, the academic literature on these issues can be separated into three main areas. First, there is technical literature, setting out variations of QALY-based assessment, typically in response to specific concerns (for example, Nord, 1999; Devlin, Tsuchiya, Buckingham and Tilling, 2011; Edlin, Tsuchiya and Dolan, 2012). Second, there is a philosophical literature, relating debates over QALY-based allocation to questions about distributive justice and the conceptualization of well-being (Daniels, 2008; Harris, 2005; Hausman, 2006), and a move towards reinvigorating questions of ethics with regard to the process of evaluation (Tabuteau and Morelle 2010, Leonard 2008). Third, there is a more diffuse literature in the history and sociology of medicine, studying the emergence and use of QALY-based measures as a part of more general trends in medical research and practice, most notably the rise of evidence-based medicine, and broader trends towards quantification, and the subsequent devaluing of other forms of knowledge (Asmore, Mulkay and Pinch, 1989; Coulter, 2002; Timmermanns and Berg, 2003). We believe that by combining historical, anthropological and philosophical perspective we can ask a set of interrelated questions which will significantly illuminate, integrate and advance the current debates, not so much at the level of specific detail (should we prefer QALYS or DALYs), but at the more fundamental level of articulating the costs and benefits of quantifying at all.

The Cambridge team will be led by Stephen John who is one of very few academics in the country to have a position specifically in the philosophy of public health, and who has already published on the uses and limits of Cost-Benefit-Analysis in environmental and healthcare contexts (John, 2010; John, forthcoming). He will be joined by Anna Alexandrova, a philosopher whose work is dedicated to integrating perspectives from philosophy of science and moral philosophy on the notion of happiness as an element in public social policy (Alexandrova 2012). They will be assisted by two newly appointed post-doctoral fellows, one in history/history of science and the other in anthropology/social science. This team of four cutting edge researchers will be advised by a board of directors made up of Huw Price (philosophy); Tony Lawson (economics); Ash Amin (social science); Simon Goldhill (director of CRASSH, historian, and expert in interdisciplinary project management). The CRASSH management board also takes an overview of each of the centre's major projects, and this includes the sociologist of medicine, Sarah Franklin, the anthropologist, Caroline Humphrey, and the philosopher, Tim Crane. Cambridge University has recently established a Strategic Network in Public Health specifically to interlink the university's different medical research areas with local, national and international public health professionals: the network provides a forum to help us bring relevant scholars and professionals together

for our project's seminars and colloquia. Together with the Centre for Science and Policy, with which CRASSH collaborates closely, we will be working to integrate policy makers into the project from the start.

II. THE RESEARCH QUESTIONS FOR STRAND 2

There are three areas of questions for strand 2, which correspond roughly to the project's historical, anthropological and philosophical bases. Each area will be conducted primarily by one of the research team, but the development of the research questions, the interrelation of the research questions, and the production of at least some of the final documents will be an actively shared collaboration, under the direction of the advisory board.

Research Question [1]

<u>What is the history of the introduction of QALYs?</u>

This element of the project will be conducted primarily by the first appointed post-doctoral fellow who will have an expertise in history, history of science, history of medicine, under the direction of Stephen John, Ash Amin, Tony Lawson and Simon Goldhill, and within the framework of Cambridge's world-class Department of History and Philosophy of Science. It is a striking fact that despite an ongoing technical literature produced largely within the health care economics community, which evaluates the modeling of QALYs (e.g. Steinbrook, Nord, Devlin et al, Edlin et al.), there is as yet no adequate, contextualized, broad history of how QALYs were introduced and what the impact of their introduction has been on health care systems, either for Britain or in a wider comparative perspective. This history will provide a necessary empirical basis for the project's work. It must be stressed how novel such a history should prove to be. The interplay between institutions, ideology and science has been tellingly analysed by a string of scholars from Foucault's work on psychology and hospitalization, to Palombo, Pelliteri, Verdeber and Fine who have shown the connection between hospital architecture and social process. Yet it is extremely rare to find a significant account of the introduction of a system of evaluation that views it fully as a historical event, with its social, political and philosophical implications, as well as a huge impact on policy and decision making. To this end, the researcher will need to draw on the history of science, history of institutions and history of public health, and will need to investigate (1) what QALYs are and what they actually measure, and what the arguments were that were successfully made for their adoption and what the opposition was; (2) how QALYS have been used and what the processes of qualification have been, and how this has changed as a result of political and other pressures across time; (3) how the introduction of QALYs should be understood within a broad history of medicine as social policy; (4) what the impact of decentralizing such decision making may be. The history of the introduction of QALYs opens a new perspective on what sort of history is needed for modern institutional processes and will provide a portrayal of a major process in modern social policy that will be of great interest and use across a series of fields. For our project, it is crucial to have a fully articulated account of how a particular form of numerical quantification became the dominant model of evaluation in healthcare decisions within a particular jurisdictions, and how it changed the social and medical practices of the institution, particularly given that NICE's work has been used both as a model for other nations and as a warning of the dangers of "socialized medicine".

Research Question [2]

How do QALYs change the processes of health care?

The historical research of the first research question will be complemented by a different kind of analysis that is of equal significance. Although there is a growing literature from within the history and sociology of medicine that has studied the emergence and use of QALY-based measures as a part of more general trends in medical research and practice, most notably the rise of evidence-based medicine, and broader trends towards quantification, and the subsequent devaluing of other forms of knowledge (eg Asmore, Mulkay and Pinch; Coultay; Timmermanns and Berg), there is no detailed account of how the introduction of QALYs has changed the behavior of the actors within the hospital system, or the relationship between the hospital and a broader society. Using actor network theory, and an anthropological expertise in analysing group dynamics with regard to normative and regulatory systems, this will be the first detailed study at the level of actor behavior and public perception of the impact of QALYs. This element of the project will be primarily conducted by the second appointed post-doctoral fellow, under the guidance of and in collaboration with Stephen John and Ash Amin; and utilizing, along with the advisory board of the project, the unique facilities for integrating such research into professional environments provided by the University's Strategic Network in Public Health. Since QALYs are often regarded as a means to remove personal judgment from decision-making processes, this strand of the project will be a way of reinvigorating the question of how such a tendency towards quantification changes social interaction, with obvious relevance to broader claims for the merits of "Evidence-Based Policy-Making". This element of the project will be broaching, then, three topics. (1) How does the introduction of QALYs change the behaviour, social practices, strategic actions of stakeholders in public health – doctors, patients, families? (2) How has the introduction of QALYs been perceived in a broad public discourse? (This topic will overlap closely with the work of Research Question 1.) (3) What are the broad social and political conclusions to be drawn about the introduction of QALYs from these first two topics?

Research Question [3]

What arer the philosophical underpinnings, implications and limitations of QALYs as a mode of evaluation?

The third strand of the project will look most broadly at the philosophical implications of the introduction of QALYs, and will be conducted primarily by Stephen John in conjunction with Anna Alexandrova, under the guidance of Huw Price and Tony Lawson. The broadly utilitarian modeling of QALYs has been challenged in multiple ways. Interventions which have the greatest overall benefit do not necessarily benefit those who are least well-off, thereby conflicting with John Rawls' influential account of social justice (Daniels). Does utility maximization misrepresent communal concerns and fears (Wolff)? Can "equity weighting" adequately modify QALYs' distributive algorithm (Edlin, Tsuchiya and Dolan)? Can the relative judgment embodied in financial thresholds resist being reified as absolute measures (McCabe, Claxton and Culyer, Porter)? We wish first to ask a broader question which is too commonly obscured by such detailed analysis of the success or limitations of specific modeling of QALYs, namely, (1) can moral and

social concerns be quantified numerically without loss? What have been the consequences of the quantification enacted by QALYs on the normative, regulatory, and decision-making capacity of the health system? This first general question will lead inevitably to a second more detailed set of issues. (2) Complementing the work of health care scholars such as Tabuteau and Leonard, lawyers such as Laude, and social theorists such as Porter, as well as Anderson's philosophically informed critique of quantification in environmental contexts, we will interrogate what QALYs measure, whether the translation of relative concerns into numerically quantified financial thresholds influences how NICE functions, and, thus, what the place of ethics is and can be within the financial modeling of QALYs. In turn, this will provide a broader framework for assessing debates over which particular ways of quantifying health should be adopted. More generally, it is clear enough that decisions about resource distribution must be taken and that they have to be taken with regard to broad social and political issues of justice, on the one hand, and, on the other, ethical issues concerning happiness, well-being and longevity. (3): the insistent question remains: how should quantification be used in a society characterized by multiple value systems, not all of which can be translated into numerical form?

III. CONCLUSION for STRAND 2

This project is focused on Britain – only by such a delimitation is it feasible for four scholars to make significant inroads in three years. It is integral to this project, however, that it is also internationally collaborative, and that the interaction between Cambridge and the Chicago and Toronto groups will be mutually informative; that the methodological and substantive issues will be shared; and that active debate between the groups will sharpen the problems and solutions of each team. It must be underlined how rare an opportunity it is for such a collaboration to be established.

The case-study of the use of QALYs in healthcare provides a unique opportunity for a team of young researchers to investigate a set of exciting questions – about the quantification of values such as equity; about the quantification of our lived experience; and about the public life of numbers. Furthermore, as well as the intellectual interest of such a project, it occurs at a crucial transitional juncture in the development of the UK healthcare system where its results can be expected to have wide interest. It offers a chance to explore how different methodological approaches can and must be combined to produce a nuanced and sophisticated acount of such a complex social porocess, and to explore how history, anthropology and philosophy can join to have an impact on the thinking of policy.

RESEARCH PROPOSAL FOR <u>PROJECT STRAND 3</u> AT UCSB

1. THE RESEARCH BACKGROUND FOR STRAND 3

Universities around the world are subjected to unprecedented social and political pressures, and are having to adapt to sometimes contradictory demands – be better, be cheaper—at an unprecedented rate. Universities are increasingly being held responsible for the health of their nation's knowledge economy (or KBE--knowledge-based economy), for the workforce readiness of the youth population, for the technologies that are to found new industries, and for social cohesion and political capability. As university performance has become a major economic and political issue, a range of quantitative metrics has developed to assess and govern that performance.

This trend is particularly well-studied in the United Kingdom, Australia, New Zealand, and the European Union, where "new public management" and "audit culture" are established university practices and have large literatures in various languages (Bowker and Star 2000, Bruno and Didier 2013, Lampland and Star 2008, Power 1999, Strathern 2000, Vinokur 2008**).** The U.S. also has a policy tradition that goes back to "reinventing government" in the early Clinton administration, and its own practices of scientific research assessment (Gore 1993). What we could call academic knowledge performance has become the domain of quantitative procedures that are linked together in a complex assessment culture, whose influence continues to grow.

The situation cannot be framed as either novel or invasive. Numerical calculation is not new in the arts, humanities and social science disciplines (AHSS or, anagramically, SASH). Numerical calculation is not new in the management of colleges and universities. In both cases, quantification has long made central contributions to understanding cultures, societies, and institutions as systems, working in symbiosis with more established qualitative methodologies. For example, UCSB historian John Majewski, in a discussion of quantitative history, has noted that in the 1960s and 1970s, quantitative analysis "seemed more democratic than traditional history that typically focused on prominent politicians, diplomats and statesmen; ordinary men and women literally counted in a statistical table" (Majewski 2015). Descriptive statistics can have a similarly illuminating effect in university administration, and help reallocate resources away from entrenched incumbents towards emerging fields or towards higher workloads.

What *is* new about the role of the numerical in higher education? The numerical's comparative *scope* is new: it now furnishes the basic picture of a very heterogeneous international higher education sector, in part by reducing or eliminating contextual variety. The numerical's *priority to qualitative* considerations is new, and has resurged in spite of decades of criticism of the validity of traditional quantitative metrics in all levels of education (Lehman 1999, Ravitch 2014, Kamenetz 2015**)**. The numerical has gained new momentum through the worldwide interest in "big data," where advanced data processing techniques appear to offer a simplicity, convenience and *objectivity* not found in qualitative comparisons. The numerical has the ability to establish *benchmarks*, which favor processes and goals that are more measurable over those that are complicated or inseparable from their operation in specific contexts (Ordorika and Lloyd 2013). Quantitative benchmarks, rankings, and indexes have also acquired a *universality and inevitability* that give them authority over educational governance. They shape resource decisions and public policy even when their validity is disputed (Osterloh and Frey 2010).

It does not overstate the case to say that in higher education's current global phase, quantification has expanded from being used under special circumstances for specific insights about complex situations into being available under all circumstances as a privileged, decisive, and objective mode of comparison and assessment. The *qualitative* analysis of learning, research, or institutional resource allocation has come to seem partial, specific and subjective by comparison to quantitative analysis, which most decision makers now regard as more general and objective (Gates Foundation 2015). Academic research and institutional decision making require comparison of dissimilar elements across varying contexts, and higher education now generally assumes that the numerical is the best adjudicator of comparisons across varied disciplines, institutions, and cultures. It is a matter of some urgency to identify the *costs of the numerical* as universities try to adapt to conditions that are changing in different ways across the globe—and to propose better methods of evaluating and guiding these adaptations.

The UC Santa Barbara team will be led by Christopher Newfield, who is one of the few humanities scholars conversant with the full range of higher education policy analysis, and who has published widely on the relation between cultural study and quantification in university teaching and research (e.g. Newfield 2008, Newfield 2010, Newfield 2014). His co-PI is Bishnupriya Ghosh, a widely recognized scholar of postcolonial studies who analyzes the relations among global marketization, risk management mechanisms, and forms of speculation and cultural circulation that avoid standardization across international contexts (Ghosh 2004; Ghosh 2011; Ghosh 2012). They will be assisted by two newly appointed post-doctoral fellows, one in higher education data science and the other in international higher education history and policy.   This team of four cutting edge researchers will be advised by a multi-campus board of directors. As the project develops, we will build a research network of higher education scholars with an interest in this topics.

2. THE RESEARCH QUESTIONS FOR STRAND 3

Numerical systems have come to determine results in four major dimensions of higher education. (1) *stature* (of an institution); (2) *efficiency* (of resource allocation); (3) *quality* (of educational outcomes); and (4) *productivity* (of faculty). In each domain, qualitative analysis has been eclipsed by quantitative metrics that are regarded as objective, comparative, and external to the practices and relationships being evaluated.

A quantitative system of measurement has been developed for each of these dimensions, as noted in Table 1.

Table 1: University Performance and Associated Metrics

Dimension	Metric
Institutional stature	University Rankings
Financial efficiency	Return on Investment
Research Productivity	Research Bibliometrics
Educational Quality	Learning Outcomes

Each dimension of university performance has attracted a measurement system. Each measurement system is complex and heterogeneous. Each has a substantial professional literature.   Each will be approached through a case study. Each will stage

one of our four research questions, which are coordinated with the other research groups.

Research Question 1

How does the history of the introduction of university rankings explain their ongoing global influence?

<u>Metric: University rankings</u>

Rankings are very well known in the mainstream media and are used by everyone from policymakers in allocating public funds to parents helping their high school children decide where to apply to university. These rankings date from the 1980s, when they were popularized by a then-declining newsweekly called *U.S. News & World Report.* Rankings have multiplied in the past fifteen years, and are now generated in many countries (the United States, United Kingdom, Germany, Singapore, China). They have been subject to severe methodological critique for weaknesses such as an overemphasis of parameters that reinforce existing concentrations of prestige and incumbent advantage (Osterloh and Frey 2010), for mismatches between the indicator and the object measured, and for mixing heterogeneous components into invalid summary indexes (Robinson-Garcia and Calero-Medina 2014, Gingras 2014).   The rankings industry has taken such critiques seriously (Nian 2013), and they have pluralized their criteria and adjusted their weightings. Rankings are generally used "under erasure," that is, with acknowledgement of their limitations and nods to their status as mere approximations.

Rankings are also seen as "here to stay" (Marope, Wells and Hazelkorn 2013), and are used regardless of known biases as though there were an "unwritten rule that *any number beats no number*" (Gingras 2014). Rankings are used by managers, policymakers, and the general public to make comparative quality judgments in the absence of substantive knowledge of the qualities being compared. Dissimilar disciplines, goals, and practices are reduced to a common indicator that can be compared through quantification, with the crucial operation being the stripping away of contextual specifics. This is a key feature of the well-studied phenomenon of "audit culture" (Lorenz 2012, Power 1999, Shore and Davidson 2014, Wright and Ørbert 2008, Wright and Rabo 2010).   The rankings indicator has come to serve as a substitute for absent substantive and contextual knowledges.   An additional concern has been that rankings insure "institutional isomorphism" (DiMaggio and Powell 1983) by requiring mutual imitation for the sake of rankings improvement.

The literature on rankings is particularly abundant, but the history of their conceptual evolution is not well understood. Two major landmarks can be identified. The *US News & World Report*'s Best Colleges rankings began in 1983, and Shanghai Jiaotong University's Academic Rankings of World Universities, which globalized Best Colleges 20 years later. But why did these instruments become so popular? What sociocultural interests and needs did such rankings address upon their appearance, in spite of their glaring methodological flaws? We seek to identify the general features of the *decontextualization* whereby nominally similar but functionally diverse teaching and research practices are reduced to comparability.

We will compare the genealogy of three contrasting rankings systems: (i) US News & World Report (the senior ranking service in the US); (ii) Washington Monthly's "public

good" and "best bang for the buck" rankings; (iii) and the long-running project "Colleges that Change Lives," which identifies qualitative impacts on individual students enrolled in liberal arts colleges (Pope 2012).   The comparison will generate an inventory of factors omitted from the standard rankings as the developed historically.

Research Question 2

How does quantification change students' use of a college degree?

Metric: Return on investment in bachelor's degrees. A growing number of ROI measures define the value of university completion as a straightforward investment in an increase in future earnings. These rankings have emerged in the U.S. amidst a national debate about whether "college is still worth it" after several decades in which college tuition costs have risen well above the rate of consumer prices. In Anthony Carnevale's terms, "Do graduates get a job in their field, earn enough money to pay their loans?" (Stainburn 2013). In the U.K., university loan repayments are indexed to the debtor's salary, giving educational ROI a new presence in British educational life.

There are two major classes of ROI rankings. The first compares ROI by college, and the second compares differences in income by major discipline.   In the first category, Payscale.com has created a well-known website that ranks universities by 30 Year Net ROI and Annual ROI, with variations (Payscale). Many other services have sprung up, and the Obama White House has sanctioned the ROI approach by including graduate earnings in its College Scorecard project (Department of Education). In the second category, *US News & World Report*, Payscale, and other services generate widespread media attention by identifying majors that offer a higher return on the college investment (Gandel and Haynie 2013, Sheehy 2012). High ROI majors predictably skew towards tech—all types of engineering, computer science, data analytics, and the like. "The 13 Most Useless Majors," also predictably, come almost entirely from the arts and humanities (Anon 2012). The trend inspired one university consortium to issue its report on lifetime earnings by major in order to argue that humanities majors earn respectable if not top salaries and that essential vocations would be depopulated without them (Humphreys and Kelly 2014).

For ROI, our second metric, we will take Payscale's calculations as our case, compare them to Bureau of Labor Statistics data on income variation by occupation, and conduct an analysis of the media discourse of its results. We will explore our the hypothesis that ROI is a faulty indicator that measures something other than what it names: not the financial investment in college learning, but a wage price determined by demand pressures, social conventions, and other factors that have little to do with university activities. In addition to sidelining the nature of higher education as a practice of sustained learning, ROI may be systematically biased against disciplines focused on developing capabilities that cannot be codified into standard vocational skill sets because the skills remain *diverse*, and/or that work through social spillover rather than through for-profit accounting units. We will continue to monitor the evolution of ROI's use to evaluate universities internationally.

Research Question 3:

What are the philosophical underpinnings, justifications, and limitations of numerical metrics in higher education?

Metric: Evaluative bibliometrics in qualitative fields.

Bibliometrics is the philosophical and operational heart of quantitative metrics in higher education. Drawing on an elaborate history of discussions of the nature of scientific communication, Eugene Garfield gave bibliometrics its modern form when he created the Science Citation Index (SCI) in the 1950s in an effort to use computers to improve literature searches (De Bellis 2009). In 1964, Garfield founded the Institute for Scientific Information (ISI) to publish the SCI. The index now operates as Thomson Reuters, is the most comprehensive of its kind, provides a wide array of citation and other types of metrics and analytical products, is proprietary and in important ways opaque, and has spawned a large number of imitators and alternatives. In this tradition, bibliometrics rests on a set of stable assumptions that shape its effects.

A. Citation frequency is a reliable index of cognitive impact (De Bellis 2009).
B. Citation indexes measure the intellectual impact and productivity of individuals and units over time, and thus can be used for quality assurance and other management functions (Hirsch 2005).
C. Most citations reference a small number of scientists; intellectual output and impact are concentrated rather than widely distributed. The whole field is oriented around a power law function, or Pareto rather than Gaussian distribution (Lotka 1926; Bradford 1934; Zipf 1936; De Bellis 2009). Citation distributions confirm the Matthew Effect (Merton 1968): they reflect cumulative advantage in which the "rich get richer."
D. Summaries of bibliometric results that take the form of simple rankable numbers (Journal Impact Factor, h-index [Hirsch 2005]) are both popular and technically invalid.
E. Valid indicators can be constructed (Gingras 2014) and used correctly to assess research impact and productivity in the context of "informed peer review" (Wouters 59).

Bibliometrics expects highly productive scholars to be few in number, and influence and originality to be concentrated. Pareto's "80/20 rule" expresses this idea by asserting that 80 percent of effects can be traced to 20 percent of causes. In recent years, Pareto distribution has come to be associated with disruptive innovation and with entrepreneurship (Andriani and McKelvey 2009), while "normal" or Gaussian distributions reflect what Nassim Taleb called "Mediocristan" in which the typical person is mediocre and everyone suffers the "tyranny" of the not very bright "collective" (Taleb 2007). Although most scholars and researchers are concerned with protecting originality, diversity, and minority voices (Engwall, Blockmans and Weaire 2014), this philosophy of Pareto distribution functions as though originality were already identified through the "improbable" exceptions at the top of citation rankings.

In reality, a system's diversity can only be defined via the full spectrum of its members or components, not simply via those that receive the lion's share of attention. Garfield himself noted the existence of "highly useful journals that are not cited frequently," in tacit reference to the invisible college of thoughts, hints, conversations, disputes, and affiliations among scholars that citations cannot capture. The limits of bibliometric assessment in evaluating research outputs and intellectual trends is acknowledged by official bodies such as the national research councils of Canada, France, and other countries (Archambault and Gagné 2004, Institut de France 2011). And yet the international research community and policy agencies have been unable or unwilling to purge defective forms of quantification.

We will investigate the philosophical and cultural assumptions that have become embedded in research bibliometrics (as grounded in citation analysis) such that its technical flaws produce renewed efforts at quantification rather than a "qualitative turn." We will apply our philosophical analysis to two contemporary trends. The first is the attempt to extend bibliometrics to research in the arts and humanities. We will analyze Thomson Reuters Arts & Humanities Citation Index in the context of the rest of Thomson Reuters' bibliometric products. Second, we will examine the treatment of arts and humanities research in the emerging field of "altmetrics," focusing on Altmetrics.com's efforts to supplement traditional citation analysis with analysis of social media impacts (Roemer and Borchardt 2012; Priem 2014).

Research Question 4

What elements would need to be developed for a "qualitative turn" in university assessment systems, one that would support higher education diversity across the globe?

<u>Metric: Quantified learning outcomes</u>.

Learning is the area in which the numerical is likely to have the most dramatic and irreversible impact on higher education. Although quantification has enormous influence over research, and has come to assess and govern research worldwide, it does not aim to replace research. The situation for instruction is different. Numerical approaches to instruction both assess it and now, in some cases, seek to replace it with automated information delivery structured around testing.

Metrics have come to define educational quality in three major ways. The first is through various kind of degree completion and attainment measures. These have been standardized for international comparisons by the Organization for Economic Cooperation and Development (OECD 2014). The second consists of national and international standardized test comparisons such as the PISA tests (OECD 2012). The third mode became visible after 2011, during the rise to global fame of the scalable Massive Open Online Course (xMOOC). MOOCs seek to automate both instruction and testing. Their most distinctive feature was to fuse the two, so that the learning process was interrupted at intervals as short as every 3-5 minutes by a quiz whose results were recorded. The fusion of learning and measurable testing was the core of the xMOOC's most important pedagogical initiative, which was to bring "adaptive learning" to a worldwide audience (PCAST 2013).

All parties to the debate about ed-tech and contemporary instruction are agreed on the importance of learning more about learning. Some have developed alternative metrics. For example, the most widely discussed English-language book on college learning in recent years was *Academically Adrift* (Arum and Roksa 2010), which used an instrument called the Collegiate Learning Assessment (CLA) rather than standardized tests to measure university learning. In so doing, *Academically Adrift* made several important methodological moves. The authors broke through quantitative statistics about degree attainment — like what percentage of enrolled students get BA degrees — to ask what graduates actually *learned* while getting their degrees. They separated college learning from college brand. And they sought to use the CLA to assess *qualitative* intellectual development. A primary finding was that mass scale passive learning was less effective than expected. An unnoticed implication was that learning was sustained by the active approaches that still prevailed in the liberal arts and sciences core (Newfield 2014). The

current period is marked by a tension between, on the one hand, efforts to increase the sophistication of teaching and assessment programming to the point that metric both accurately assess learning while replacing the non-scalable aspects of human instruction, and, on the other, efforts to break through the limits of numerical approaches.

Our case for RQ 4 will be the MOOC provider edX, which has a branch at UC Berkeley, as it seeks to become a viable substitute for face-to-face college instruction through advanced technics in adaptive learning grounded in data analytics. We will compare the testing practices and results of this high-quality MOOC provider to (i) the CLA testing instrument; (ii) the interdisciplinary theory of learning, which is both diffuse and growing rapidly (Brown, Roediger and McDaniel 2014; Immordino-Yang 2014); and (iii) the "capabilities" approach to individual intellectual development (Nussbaum 2010). Our goal will be a full inventory of elements that should define university-level learning in the knowledge-based economies (KBEs) from which our data is drawn. We will use this inventory to develop a new qualitative understanding of the cognitive and intellectual changes that universities should provide, and that are possibly being made more difficult to achieve through the routinization of learning that MOOC-era quantification encourages. We will answer the question, how could existing instruments be changed to encourage the development of a wider range of creative capabilities in 21st century higher education?

Table 2: University Performance, Associated Metrics, Research Questions

Dimension	Metric	Research Domain
Institutional stature	University Rankings	History-Genealogy
Financial efficiency	Return on Investment	Impact-Use
Research Productivity	Research Bibliometrics	PhilosophicalUnderpinnings
Educational Quality	Learning Outcomes	Qualitative Reform

III. CONCLUSION for STRAND 3.

Our overriding concern is to identify the contexts, qualifications, and local goals—the "modalities" (Gläser and Laudel 2007)-- of the core global university activities of instruction and research that are dropped or lost through the evolution of numerical approaches to the assessment instruction and research. We note a paradox in current global university development. On the one hand, universities are being asked to adapt to a diverse range of local, national, and regional needs in a world where KBEs are following different pathways of advancement and evolving their own rules and norms. It is increasingly clear that economic development can no longer be implemented by exporting and imposing a single model devised in the West (or anywhere else) (Moore, 2015).   On the other hand, numerical indicators create uniform standards to which universities must norm themselves in order to succeed with global metrics. Universities may be making themselves less functional in their home "conditions of production" by improving functionality in areas subject to quantitative measurement. Metrics themselves are struggling with this issue: the editor of the collection, *Beyond Bibliometrics*, notes the simultaneous emergence of "total bibliometrics" and of a "culture of multiples" within bibliometrics [Cronin 2014]).

Our founding assumption is that universities cannot serve their societies by adopting one model of education and management. They must instead be sources of multiple design pathways, diverse methodologies, and even multiple rationalities. To provide these

needed varieties of thought and practice, universities will need to take multiple forms, and have the capacity for "experimental collaboration" with their regions and with parts of the population that higher education has not traditionally served. We will evaluate selected metrics for any tendency to suppress diverse knowledges and knowledge practices in favor of generalizable norms.

In other words, we will evaluate selected metrics for their impact on the effort of global universities to move in the following ways. Universities going forward will be *adaptive* to changing local, regional, and national needs. They will generate *diverse* knowledges, practices, and skills in their graduates. They will *individualize* their students by using technological and human systems to reverse the massification and anonymization that is known to impair learning. Universities will r*educe status and resource stratification*. Our group will therefore analyze numerical assessment systems for their impact on *adaptability, diversity, individualization, and destratification*.

REFERENCES

REFERENCES STRAND 1

Archer, David. 2009. *The Long Thaw: How Humans Are Changing the Next 100,000 Years of the Earth's Climate.* Princeton, N. J.: Princeton University Press, 2009.

Banerjee, Subhankar. 2012. *Arctic Voices: Resistance at the Tipping Point.* New York: Seven Stories Press.

Crichton, Michael. *The State of Fear.* New York: Harper Collins, 2004.

Hansen, James. *Storms for my Grandchildren: The Truth About the Coming Climate Catastrophe and Our Last Chance to Save Humanity.* New York: Bloomsbury, 2009.

Mann, Michael. *The Hockey Stick and the Climate Wars: Dispatches From the Frontlines.* New York: Columbia University Press, 2012.

Morton, Timothy. *Hyperobjects: Philosophy and Ecology after the End of the World.* Cambridge, Mass.: Harvard University Press, 2013.

Oreskes, Naomi and Eric M. M. Conway *Merchants of Doubt: How A Handful of Scientists Obscured the Truth from Tobacco Smoke to Global Warming.* New York: Bloomsbury, 2010.

Pearson, Charles S. *Economics and the Challenge of Global Warming.* Cambridge: Cambridge University Press, 2011.

Pierrehumbert, Raymond T. *Principles of Planetary Climate.* Cambridge: Cambridge University Press, 2010.

REFERENCES STRAND 2

Alexandrova, A (2012) "Well-being as an object of science" *Philosophy of Science* 79(5): 678-89.

Alston, W. 2005. *Beyond "Justification": Dimensions of Epistemic Evaluation* , Ithaca: Cornell University Press.

Anderson, E (1995) *Value in Ethics and in Economics*, Cambridge: Cambridge University Press

Ashmore, M Mulkay, M and Pinch, T. (1989) *Health and efficiency: a sociology of health economics* Milton Keynes: Open University Press.

Brown, M (2009) *Science in Democracy*, Cambridge, Mass: MIT press

Castells, M (2009) *The rise of the network society* (second edition), London: Wiley Blackwell

Coulter, A. (2002) "Evaluating the outcomes of health care" in Bury, M Calnan, M and

Gabe, J. *The sociology of the health service* London: Routledge.

Daniels, N (2008) *Just health* Cambridge: Cambridge University Press.

Devlin, N, Tsuchiya, A, Buckingham, K, and Tilling C (2011) "A uniform time trade off method for states better and worse than dead: feasibility study of the 'lead time' approach" *Health Economics* 20(3); 348-361.

Douglas, H (2009) *Science, policy and the value free ideal*, Pittsburgh, PA: University of Pittsburgh Press

Edlin, R, T, A and Dolan, P. (2012) "Public preferences for responsibility versus public preferences for reducing inequalities" *Health Economics*, 21(12); 1416-1426.

Fisch, M. and Benbaji, Y. (2011) *The View from Within: Normativity and the Limits of Self-Criticis*m, Notre Dame: University of Notre Dame Press.

Hacking, I (1998) *Mad Travelers*, University of Virginia Press

Harris, J. (2005) "It's Not NICE to Discriminate" *Journal of Medical Ethics* 31(7); 373–375.

Hausman, D. (2006) "Valuing health" *Philosophy and Public Affairs* 34(3); 246-274.
John, S (2010) "In defence of bad science and irrational policies: an alternative account of the precautionary principle" *Ethical Theory and Moral Practice* 13(3): 1-18

John, S (Forthcoming) "Efficiency, responsibility and disability" Forthcoming in *Politics, Philosophy and Economics*

Lambek, M (2010) *Ordinary Ethics: anthropology, language, action*, Fordham University Press

Latour, B. and Woolgar, S. (1986) *Laboratory Life: the Social Construction of Scientific Facts*, 2nd edition, Beverly Hills.

Latour, B (2007) *Reassembling the social: an introduction to Actor Network Theory* Oxford: Clarendon Press

Léonard, C (2008) *Croissance contre santé : Quelle responsabilisation du malade*, Paris: Editions Couleur livres

Lucas, R (1976) "Econometric policy evaluation: a critique" in Brunner and Meltzer (eds) *The Phillips Curve and Labor Markets* New York: American Elsevier.

Mackenzie, D (2008) *An engine not a camera* Cambridge, Mass: MIT press

McCabe, C Claxton, K and Culyer, A. 2008 "The NICE cost-effectiveness threshold: what it is and what that means." *PharmacoEconomics* 26(9); 733–44.

NICE (2013) *Guide to the methods of technology appraisal*. London: National Institute of Health and Clinical Excellence http://publications.nice.org.uk/pmg9

Nord, E. (1999) *Cost-value analysis in health care: making sense out of QALYs* Cambridge: Cambridge University Press.

Peterson, M (2013) *Dimensions of consequentialism* Cambridge: Cambridge University Press

Porter, T. (1996) *Trust in Numbers* London: Princeton University Press.

Post, H. (1971), "Correspondence, Invariance and Heuristics", *Studies in History and Philosophy of Science*, 2, 213-255.

Power, M. (1999) *The audit society: rituals of verification*, Oxford: Oxford University Press

Rawls, J. (1971) *A Theory of Justice*, Cambridge: Harvard University Press.

Pritchard, D. (2005). *Epistemic Luck* , Oxford: Oxford University Press.

Siskin, C and Warner, W (2010) *This is Enlightenment* Chicago: University of Chicago Press

Steinbrook, R. (2008) "Saying no isn't NICE - the travails of Britain's National Institute for Health and Clinical Excellence." *New England Journal of Medicine* 359(19): 1977–1981.

Steup, M. ed. (2001) *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Virtue, and Responsibility*, Oxford: Oxford University Press.

Tabuteau, D and Morelle, A *La santé publique* Paris: Presses universitaires de France

Williamson, T. (2000*) Knowledge and Its Limits*, Oxford: Oxford University Press.

Wynne, B (1996) "May the sheep safely graze?" in S. Lash; B. Szerszynski; B. Wynne, eds. *Risk, environment and modernity: towards a new ecology* London: Sage

Wolff, J. (2011) *Ethics and Public Policy: a philosophical inquiry* London: Routledge

Yearley, S (2004) *Making sense of science* London: Sage publications

Zagzebski, L. (1996) *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge* , Cambridge: Cambridge University Press.

REFERENCES STRAND 3

Andriani, Pierpaolo and Bill McKelvey. 2009. "From Gaussian to Paretian Thinking: Causes and Implications of Power Laws in Organizations." Organization Science 20 (6): November–December 2009, pp. 1053–1071

Anon. 2012. "The 13 Most Useless Majors, from Philosophy to Journalism. *Daily Beast*. April 23, 2012. http://www.thedailybeast.com/galleries/2012/04/23/the-13-most-useless-majors-from-philosophy-to-journalism.html#slide

Archamabault, Éric and Étienne Vignola Gagné. 2004. *The Use of Bibliometrics in the Social Sciences and Humanities*. Montréal: Social Sciences and Humanities Research Council of Canada.

Bhandari, Rajika, and Alessia Lefébure, eds. 2015. *Asia: the Next Higher Education Superpower?* New York: Institute of International Education.

Bowker, Geoffrey C., and Susan Leigh Star. 2000. *Sorting Things Out: Classification and Its Consequences*. Cambridge, Mass.: The MIT Press.

Bradford, Samuel C. 1934. "Sources of Information on Specific Subjects," *Engineering* 137(3550): 85-86.

Brown, Peter C., Henry L. Roediger III, and Mark A. McDaniel. 2014. *Make It Stick: The Science of Successful Learning*. Cambridge, Massachusetts: Belknap Press.

Bruno, Isabelle, and Emmanuel Didier. 2013. *Benchmarking*. Paris: Zones.

Burrows, Roger. 2012. "Living with the h-index? Metric assemblages in the contemporary academy." *The Sociological Review* 60.2: 355-372.

Cronin, Blaise. 2014. "Scholars and Scripts, Sppors ans Scores." In Cronin and Cassidry R. Sugimoto, eds. 3-22.

Cronin and Cassidry R. Sugimoto, eds. 2014. *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact.* Cambridge, Mass.: MIT Press.

De Bellis, Nicola. *Bibliometrics and Citation Analysis: From the Science Citation Index to Cybermetrics*. Lanham, Md.: The Scarecrow Press, 2009).

Department of Education. 2015. College Scorecard. Online tool at http://collegecost.ed.gov/scorecard/index.aspx.

DiMaggio, P.J. and W.W. Powell. 1983. "The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. *American Sociology Review*. 48: 147-60.

Downing, K. 2012. "Do rankings drive global aspirations at the expense of regional development?" In M. Stiasny and T. Gore, eds. *Going Global: The Landscape for Policy Makers and Practitioners in Tertiary Education*. Bingley, UK: Emerald Group: 31-39.

Engwall, Lars, Wim Blockmans, and Denis Weaire. 2014. "Bibliometrics: issues and context." In Blockmans, Engwall and Weaire, eds. *Bibliometrics: Usa and Abuse in the Review of Research Performance* London: Portland Press.

Gandel, Cathie and Devon Haynie. 2013. "11 Hot College Majors." *US News & World Report*. September 10, 2013. http://www.usnews.com/education/best-colleges/slideshows/11-hot-college-majors.

Gates Foundation. *Postsecondary Success Advocacy Priorities*. Seattle: Gates Foundation, 2015.

Ghosh, Bishnupriya. 2004. *When Borne Across: Literary Cosmopolitics in the Contemporary Indian Novel*. New Brunswick, N.J.: Rutgers University Press.

Ghosh, Bishnupriya. 2011. *Global Icons: Apertures to the Popular*. Durham, N.C.: Duke University Press.

Ghosh, Bishnupriya. 2012. "The Proximate Truth : Reenactment in the Pandemic-Era HIV/AIDS Documentaries." *BioScope: South Asian Screen Studies* (3:1): 69-86.

Gingras, Yves. 2014. "Criteria for Evaluating Indicators." In Cronin and Sugimoto, eds. 109-25.

Gläser, Jochen and Grit Laudel. 2007. "The Social Construction of Bibliometric Evaluations." In Whitley, R. and J. Gläser, eds. *The Changing Governance of the Sciences*. Springer.

Gore, Al. 1993. *The Gore Report on Reinventing Government: Creating a Government That Works Better and Costs Less*. New York: Three Rivers Press.

Hirsch, J.E. 2005. "An Index to Quantify an Individual's Scientific Research Output. *Proceedings of the National Academy of Sciences of the United States of America* 102: 16569-16572.

Humphreys, Debra and Patrick Kelly. 2014. *How Liberal Arts and Sciences Majors Fare in Employment: A Report on Long-Term Career Paths*. Washington, D.C.: Association of American Colleges and Universities.

Immordino-Yang, Mary Helen. 2013. "Studying the Effects of Culture by Integrating Neuroscientific With Ethnographic Approaches." *Psychological Inquiry* 24, no. 1 (January 2013): 42–46.

Institut de France, 2011. "On the Proper Use of Bibliometrics to Evaluate Individual Researchers." Paris. Rapport de l'Académie des sciences.

Kamenetz, Anya. *The Test: Why Our Schools are Obsessed with Standardized Testing– But You Don't Have to Be*. New York: Public Affairs, 2015.

Lampland, Martha, and Susan Leigh Star, eds. *Standards and Their Stories: How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life*. 1 edition. Ithaca: Cornell University Press, 2008.

Lehman, Nicholas, *The Big Test: The Secret History of the American Meritocracy* (New York: Farrar, Straus and Giroux, 1999).

Lim, Miguel. 2015. "Global University Rankings: Determing the Distance Between Asia and 'Superpower Status' in Higher Education." In Bhandari and Lefébure, eds. pp 23-37.

Lloyd, M. 2010. "Comparative study makes the case for Mexico's public universities," *Chronicle of Higher Education.* 11 November 2010.

If You're So Smart, Why Are You under Surveillance? Universities, Neoliberalism, and New Public Management

Lotka, Alfred J. 1926. "Statistics: The Frequency Distribution of Scientific Productivity. " *Journal of the Washington Academy of Sciences* 16 (12): 317-25.

Majewski, John. 2015. "U.S. History in a Statistical Age." Unpublished Manuscript.

Marginson, S. 2007. "Global university rankings: Implications in general and for Australia." *Journal of Higher Education Policy and Management* 29(2): 131.42.

Marope, P.T. M., P.J Wells and E. Hazelkorn, eds. 2013 *Rankings and Accountability in Higher Education: Uses and Misuses.* New York: UNESCO Publishing.

Merton, Robert K. 1968. "The Matthew Effect in Science." *Science* 159 (3810):56-63.

Newfield, Christopher. 2008. *Unmaking the Public University: The Forty-Year Assault on the Middle Class.* Cambridge, Mass: Harvard University Press.

Newfield, Christopher. 2010. "The End of the American Funding Model: What Comes Next?" *American Literature* 82:3 (September): 611-635

Newfield, Christopher. 2014. "Is College Still Worth It?" *Los Angeles Review of Books* 29 September 2014.

Nian Cai Liu, "The Academic Ranking of World Universities and its future direction," in Marope et al.: 23-40.

Nussbaum, Martha C. *Not For Profit: Why Democracy Needs the Humanities* Princeton, N.J.: Princeton University Press, 2010.

OECD. 2012. *Pisa 2012 Results in Focus: What 15-year-olds Know and What They Can Do with What They Know.* OECD Publishing.

OECD. 2014. *Education at a Glance 2014: OECD Indicators.* OECD Publishing.

Ordorika, Imanol and Marion Lloyd. "A Decade of international university rankings: a critical perspective from Latin America." In Marope et al.: 209-31.

Osterloh, Margit and Bruno S. Frey. 2010. "Academic Rankings and Research Governance," Working Paper No 482. Zurich: Institute for Empirical Research in Economics, University of Zurich: 1-29.

Payscale. 2015. "2013 College Education ROI Rankings: Does a Degree Always Pay Off? Online at http://www.payscale.com/college-education-value-2013

Pope, Loren. 2012. *Colleges That Change Lives: 40 Schools That Will Change the Way You Think About College.* New York: Penguin.

Power, Michael. 1999. *The Audit Society: Rituals of Verification*. Oxford; New York: Oxford University Press.

President's Council of Advisors on Science and Technology (PCAST). 2013. "Letter to the President on MOOCs." https://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_edit_dec-2013.pdf

Priem, Jason. 2014. Altmetrics. In Cronin and Sugimoto. 263-88.

Ravitch, Diane. *Reign of Error: The Hoax of the Privatization Movement and the Danger to America's Public Schools*

Robinson-Garcia, Nicolás and Clara Calero-Medina. 2014. "What do university rankings by fields rank? Exploring discrepancies between the organizational structure of universities and bibliometric classifications." *Scientometrics* 98.3: 1955-70.

Roemer, Robin Chin and Rachel Borchardt. 2012. "From Bibliometrics to Altmetrics: A Changing Scholarly Landscape." *C&RL News*. 596-600.

Shore, Cris and Miri Davidson. 2014. "Beyond collusion and resistance: Academic–management relations within the neoliberal university." *Teaching and Learning* 7:1.

Stainburn, Samantha. 2013. "Following the Money: Calculating the Net Worth of a College Degree. *New York Times*. August 2, 2013. http://www.nytimes.com/2013/08/04/education/edlife/calculating-the-net-worth-of-a-college-degree.html?_r=0

Strathern, Marilyn. 2000. *Audit Cultures: Anthropological Studies in Accountability, Ethics and the Academy*. 1 edition. London New York: Routledge.

Taleb, Nassim Nicholas. 2007. *The Black Swan: The Impact of the Highly Improbable*. New York, Random House.

Vinokur, Annie. 2008. "La loi relative aux libertés et responsabilités des universités : essai de mise en perspective." *la Revue de la Régulation* 2 (Janvier)

U.S. News and World Report, Best Colleges. http://colleges.usnews.rankingsandreviews.com/best-colleges

Washington Monthly. 2015. College Guide. http://www.washingtonmonthly.com/college_guide/

Wright, Susan and Jakob Williasm Ørberg. 2008. "Autonomy and control: Danish university reform in the context of modern governance." *Teaching and Learning* 1:1.

Wright , S. and A. Rabo, eds. "Anthropologies of university reform." Special issue of *Social Anthropology/Anthropologie Sociale* 18.1.

Wouters, Paul. 2014. "The Citation: From Culture to Infrastructure." In Bronin and Sugimoto, eds. 47-66.

Zipf, The Psycho-Biology of Language: An Introduction to Dynamic Philology (London: Routledge, 1936).