THE CHANGING GOVERNANCE OF THE SCIENCES

Sociology of the Sciences Yearbook

VOLUME XXVI

The titles published in this series are listed at the end of this volume.

THE CHANGING GOVERNANCE OF THE SCIENCES

The Advent of Research Evaluation Systems

Edited by

RICHARD WHITLEY
University of Manchester, UK

and

JOCHEN GLÄSER
University of Lancaster, UK

Springer

Library of Congress Control Number: 2008920619

ISBN 978-1-4020-6745-7 (HB)
ISBN 978-1-4020-6746-4 (e-book)

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springer.com

Printed on acid-free paper

PREFACE

This book arose from a conversation between Lars Engwall and Richard Whitley during the 2003 European Group for Organization Studies Colloquium in Copenhagen about important topics for future research. We agreed then, and this has been amply confirmed by subsequent events, that the proliferation of research evaluation schemes, especially in Europe, was an important feature of the changing relationships between the state, universities and scientific research more generally, which needed systematic and comparative analysis. With the support of the Bank of Sweden Tercentenary Foundation, this discussion led to workshops in Uppsala and Manchester involving colleagues from Australia, Germany and the Netherlands at which the main framework for such an analysis was developed. This framework then formed the basis for the conference at Bielefeld in 2005 at which earlier versions of most of the papers in this volume were presented and extensively discussed.

We are very grateful for the support of the PRIME network of the European Commission and the Max Planck Institute for the History of Science for this conference, as well as the invaluable efforts of Peter Weingart of the University of Bielefeld. We are also greatly indebted to the members of the Editorial Board who commented on various drafts of these papers as well as to Aant Elzinga, Stefan Kuhlmann, Philippe Laredo, Arie Rip and Nic Vonortas who acted as referees. In preparing the volume for publication we have been greatly assisted by Kathryn Morrison and would also like to acknowledge the help of Kate Barker in liaising with the PRIME network, authors, and referees.

Richard Whitley
Jochen Gläser

TABLE OF CONTENTS

Concluding Reflections

LIST OF CONTRIBUTORS

Cozzens, Susan, Professor Dr., Technology Policy and Assessment Center, School of Public Policy, Georgia Institute of Technology, Atlanta, GA 30332-0345, USA.

Cruz-Castro, Laura, Dr., Instititue of Public Goods and Policies, Consejo Superior de Investigaciones Cientificas (CSIC), Albasanz, 26-28 3D14, E-28037 Madrid, Spain.

Engwall, Lars, Professor Dr., Department of Business Studies, Uppsala University, Box 513, 75120 Uppsala, Sweden.

Gläser, Jochen, Dr., Research School of Social Sciences, The Australian National University, Canberra ACT 0200, Australia.

Kneller, Robert, Professor Dr., University of Tokyo, RCAST, 4-6-1, Komaba, Meguro-ku, Tokyo 153-8904, Japan.

Lange, Stefan, Dr., Stiftungslehrstuhl Wissenschaftsmanagement, Deutsche Hochschule für Verwaltungswissenschaften Speyer, Postfach 1409, D-67324 Speyer, Germany.

Laudel, Grit, Dr., Research School of Social Sciences, The Australian National University, Canberra ACT 0200, Australia.

Maasen, Sabine, Professor Dr., Programm für Wissenschaftsforschung, Universität Basel, Missionsstrasse 21, CH-4003 Basel, Switzerland.

Nybom, Thorsten, Professor Dr., Department of Humanities, Örebro University, SE-70182 Örebro, Sweden.

Sanz-Menendez, Luis, Professor Dr., Institute of Public Goods and Policies, Consejo Superior de Investigaciones Cientificas (CSIC), Albasanz, 26-28 3D13, E-28037 Madrid, Spain.

Schiene, Christof, Dipl.-Soz., Referat 21: Hochschulentwicklung und –controlling, Niedersächsisches Ministerium für Wissenschaft und Kultur, Leibnizufer 9, D-30169 Hannover, Germany.

Schimank, Uwe, Professor Dr., Institut für Soziologie, FernUniversität in Hagen, D-58084 Hagen, Germany.

Van der Meulen, Barend, Professor Dr., MB-STeHPS, University of Twente, PO Box 217, 7500 AE Enschede, The Netherlands.

Weingart, Peter, Professor Dr., Institut für Wissenschafts- und Technikforschung, Universität Bielefeld, Postfach 100 131, D-33501 Bielefeld, Germany.

Whitley, Richard, Professor Dr., Manchester Business School, University of Manchester, Booth Street West, GB-Manchester M15 6PB, UK.

Introduction

CHAPTER 1

RICHARD WHITLEY

CHANGING GOVERNANCE OF THE PUBLIC SCIENCES

The Consequences of Establishing Research Evaluation Systems for Knowledge Production in Different Countries and Scientific Fields

INTRODUCTION

Many recent discussions of state science and technology policies in OECD economies since the end of the Second World War have distinguished a number of distant phases or 'paradigms' (Ruivo 1994) in state-science relations, which reflected both different perceptions of the role of scientific research in industrialised societies and changes in the size and complexity of the public science systems (see e.g. Brooks 1990; Freeman and Soete 1997: 374-395; Martin 2003). From viewing academic and other research on fundamental processes as producing knowledge that would contribute to the achievement of economic and social goals in a predominantly linear fashion, many political and bureaucratic elites have come to see the connections between scientific research and public purposes as being more complex and interactive, and to believe that realising social and economic benefits involves a more active steering of research than was commonly thought in the 1950s and 1960s (Braun 2003).

Increasingly, research conducted in universities and similar organisations concerned with understanding basic processes in the material, life and social worlds is being viewed by politicians, business elites and other influential actors as a strategically important activity for national and regional development that can and should be managed for public policy purposes. Not only have the public sciences, i.e. those reputationally governed work organisations producing knowledge for publication in peer reviewed journals and books (Dasgupta and David 1994; Whitley 2000), become too large and expensive to be left to scientific elites to manage, but their outputs have become strategic assets in the international competition to dominate technological development in new industries and ensure continued economic growth. As such, their development, organisation and governance are now

R. Whitley and J. Gläser (eds.), The Changing Governance of the Sciences, 3–27.

the focus of state policies and procedures in most OECD countries, with many states establishing ministries of science and technology and specific agencies and programmes to steer research in desired directions (Drori et al. 2003).

This steering often involves the incorporation of public policy goals and evaluation standards into the selection and monitoring processes of research funding agencies, such that project proposals have to show how the research contribute to these objectives if they were to be funded (Braun 1998). Problem focused or 'strategic' funding agencies, such as the British Medical Research Council and the US National Institutes of Health, are often charged with supporting the best research that will contribute to their overall strategic objectives, such as solving specific medical problems, so that researchers are encouraged to study phenomena that will generate useful results. Much of this work falls into what Stokes (1997: 71-74) has termed 'Pasteur's quadrant', i.e. research seeking to understand the basic mechanisms producing particular effects that can be used to deal with practical problems, such as, in the case of Pasteur, spoilage in fermentation processes.

The role of these kinds of agencies, and similar attempts to steer current research goals and careers towards public policy uses, has become more significant as the provision of state funding for research stabilises or declines relative to the number of researchers needing external resources to conduct projects, and alternative sources are few in number and/or limited in their resources (Braun, 1998). Most of these attempts have focused on the conditional selection and monitoring of research projects according to their likely contribution to strategic objectives as claimed in project proposals. They have relied on the prospective evaluation of new proposals and their applicants to guide research in particular directions, as well as maintain high quality through competitive bidding for resources allocated by variously organised peer review procedures.

In most countries, state investment in this prospective evaluation of applications for funds and guiding resource allocation procedures towards the achievement of public policy goals has been considerably greater than that provided for the retrospective evaluation of the results. As Cozzens (1990: 283) puts it:

> The US funding system places overwhelming emphasis on pre-performance evaluation (in the form of the peer review system) and gives minimal attention to programme evaluation. Thus major US funding agencies devote most of their efforts to giving out grants, not to finding out what happened once they were made.

However, as Cozzens presciently suggested in the same article (ibid.: 282): "Institutions and nations that are not currently evaluating, regularly and with all the skill they can muster, will probably be doing so ten years hence", partly to ensure the high quality of research outputs, and partly to satisfy potential and actual demands for increased political accountability of the greatly enlarged public science system (see also Cozzens' and Kneller's chapters in this volume).

As developments in the 1990s and the contributions to this volume attest, retrospective evaluation systems for assessing the quality of research conducted in universities and similar research organisations have indeed proliferated across many OECD economies, such that the *ex ante* evaluation of proposals for external project

funding in many sciences and countries is increasingly being complemented by the *ex post* evaluation of project results and publications through various national research evaluation systems.

Not only have many states shifted support for basic research away from block grants to universities to research councils and specialised funding agencies more amenable to state guidance (Block 1990), thus facilitating the short-term selection of research priorities through project funding, but they are increasingly evaluating research outputs in terms of qualitative and quantitative criteria, sometimes in order to reallocate funds to 'excellent' research organisations and away from less competitive ones. More generally, the previous direct institutional control of universities as state organisations that was prevalent in Japan and many continental European countries is being widely replaced by various forms of administrative decentralisation and increased reliance on output evaluations, as indicated by many contributions to this volume.

These sorts of changes in state-science relations have, though, occurred at different rates and in different ways in different countries. Block (1990: 40), for instance, shows very clearly how research and research students in US universities have been much more dependent on external project based funding than most OECD countries in the 1970s and 1980s. Despite the broad tendency in many societies to make universities and other research organisations formally separate from the state, major divergences in the funding, organisation and control of universities remain, as the chapters by Gläser and Laudel, Kneller, and Engwall and Nybom in this volume demonstrate (see also Campbell 2003; Gaughan and Robin 2004; Liefner 2003).

Furthermore, the nature and use of state evaluation systems varies considerably between OECD countries, as do the critical characteristics of public science systems that both structure the kinds of evaluation systems that are developed and mediate their consequences. Particularly important in this respect are the nature of the funding regime for academic research, the strategic capabilities of research organisations, the structure of scientific elites, and the segmentation of universities, and research institutes and of scientists' labour markets (Whitley 2003a). Consequently, we would expect the effects of these attempts by state agencies to steer, monitor and evaluate scientific research on the organisation of knowledge production, and on how new formal knowledge differs and is used, to vary considerably between differently structured public research systems in path dependent ways.

The institutionalisation of different kinds of research evaluation systems in many national academic systems is, then, an important component of more general changes in the governance of the public sciences that have developed over the past two decades or so. As the contributors to this volume make clear, we cannot adequately understand how and why these changes occurred, or identify their consequences for the organisation and direction of knowledge production, without locating them in their specific societal contexts, particularly the changing nature of political and economic institutions and interconnections.

It is additionally important to consider how the significant differences in the intellectual and social organisation of scientific fields (Whitley 2000) are likely to affect, and be affected by, the implementation of such evaluation systems and

broader changes in state-science relationships. Accordingly, as a contribution to the comparative analysis of the role of research evaluation systems in the changing governance of the public sciences, and their consequences for knowledge production, in this paper I suggest a framework for understanding how both variations in the nature of public research systems, and differences between the sciences, can be expected to influence the development and impact of different kinds of retrospective evaluation systems.

I first discuss the major characteristics of these evaluation systems and how they differ between countries and over time, and then consider the sorts of broad effects they are likely to have on research strategies and innovation. To simplify the analysis, it is useful to combine many of the more consequential aspects into an idealised 'strong' form of evaluation system in order to highlight the sorts of impact that they seem likely to have. The next section of the paper examines the key features of public research systems that are likely to mediate these consequences in particular countries and outlines how they should do so. Finally, I discuss the key differences between scientific fields that can be expected to influence the outcome of introducing such strong evaluation systems on particular sciences.

CHARACTERISTICS OF RESEARCH EVALUATION SYSTEMS AND THEIR CONSEQUENCES

Research evaluation systems (henceforth RES) are organised sets of procedures for assessing the merits of research undertaken in publicly-funded organisations that are implemented on a regular basis, usually by state or state-delegated agencies. As the chapters in this book exemplify, their nature varies considerably between countries, and also changes over time. In particular, they differ in how they are organised and governed, and in their implications for resource allocation decisions. Considering first their governance and structure, we can distinguish between evaluation systems in terms of their frequency, formalisation, standardisation and transparency. For example, the British Research Assessment Exercise (RAE) has varied in its periodicity from three to seven years, and the Australian formula-based measurement procedure occurred every year. Similarly, the extent to which evaluations are conducted according to formally specified procedures differs considerably between the many carried out informally at the invitation of individual universities or departments on a largely ad hoc basis and those organised by central bureaucracies with systematic rules.

Standardisation here refers to the use of common evaluation procedures and practices across the sciences and over time. This varies considerably between Australia, the Netherlands and Spain as well as between other countries, as the chapters by Gläser and Laudel, van der Meulen, and Cruz-Castro and Sanz-Menendez demonstrate. It is worth emphasising that while these may be formally prescribed and intended to function in the same way between fields in some evaluation systems, actual practices may in fact vary considerably between review panels, just as journal refereeing and grant allocation practices can differ between fields and reviewers, as shown by Langfeldt (2001). Such variations seem

especially likely when peer review judgements of quality are being made in different kinds of science, as evident in the British RAE's struggles to combine formal standardisation of procedures and consequences with appropriate and meaningful judgements of research quality in very contrasting fields of enquiry.

Research evaluations also vary greatly in their public transparency. In highly transparent systems, evaluations are carried out by panels that have been appointed according to formal, published procedures and base their judgements on publicly specified criteria and ways of working. Usually, these are published and form the basis of public rankings of the relative performance of individual departments and universities. In other systems, judgements are made by small groups of colleagues appointed informally, who decide their own working procedures and report their results to universities and, sometimes, state agencies, in confidence.

These characteristics tend to occur together. Evaluations of research achievements in many Continental European countries, for instance, have been organised relatively informally on a disciplinary basis, with neither the processes involved nor results being published, although some of these have become more systematically organised around standard protocols as in the Netherlands (van der Meulen in this volume). Many also have been primarily intended to improve research capabilities, and so can be described as 'formative', as in Schiene and Schimank's account of evaluation in Lower Saxony in this volume, rather than arrive at summated judgements that rank the quality of individual departments.

In contrast, other countries have developed much more frequent, formal, standardised and transparent RES that have been managed by funding agencies or ministries who have published their results. The British RAE has probably become the most well known of these, but other countries have also begun to introduce similar schemes on a regular basis. These kinds of evaluations assess the relative quality of research outputs from particular university departments or units of assessment over the previous four or five years, and are intended to rank them on a standard scale that is publicly available. Judging panels are selected in highly formal ways, often with continuing membership, and have to follow systematic procedures that are published and standardised across the sciences and universities. The governance of research evaluations here is nationally centralised and their organisation is highly rule governed, standardised, relatively permanent and publicly transparent.

In the case of the UK, the RAE has become progressively more formalised with written procedures being widely discussed and publicised, and outcomes analysed in the press and elsewhere in terms of prestige rankings. Not least because of legal challenges and threats of such actions, panel procedures and judgemental criteria are formally prescribed and publicly justified. As a result, the RAE has become a very important process for researchers and their employers in terms of their public reputations, as well as financially.

Another aspect of the organisation of RES that can affect their impact on knowledge production concerns the ways that the units of assessment are constructed. As van der Meulen and Schiene and Schimank emphasise in their chapters in this book, this can vary considerably between funding regimes and academic systems with significant consequences for the coordination of research in

universities. In the Netherlands the primary unit of evaluation is the research programme, and universities have had to construct these organisational entities in order to gain research funding. In Lower Saxony, the individual Chair remains the key unit of research assessment, but evaluating teams have strongly encouraged collaboration between professors and the formation of research teams that could tackle 'big' problems, thus potentially generating more collective organisational research capabilities.

Elsewhere, they may either cover entire disciplines such as History or Philosophy in ways that are isomorphic with traditional university departmental boundaries; focus on more specialised fields such as development studies; or cover wide ranging areas of concern such as environmental studies. How such boundaries are drawn, and by whom they are implemented, affect the ways that judgements of intellectual significance are made, especially in fields that cross traditional academic boundaries such as economic history, innovation studies and mathematical finance.

Finally, a major difference between RES concerns their consequences for the funding of research in the public sciences, both their direct impact on resource allocation and also in the proportion of employers' incomes that is affected by them. In quite a number of states, especially where they directly control resource allocation to individual universities and departments as in many continental European research systems, evaluations have not so far been directly linked to funding decisions (Campbell 2003 and the contributions of Cruz-Castro and Sanz-Menendez, Engwall and Nybom, and van der Meulen to this volume). In contrast, in Australia and the UK they have had significant direct effects on resource allocation outcomes, albeit to different degrees across the sciences and between individual research organisations, as Gläser and Laudel show in their account of the Australian RES.

Various discussions have suggested that implementing RES in the public services can be expected to have a number of effects on researchers and their employers (See Martin and Geuna 2003; Tabil 2001). These include:

a) heightened awareness of the significance of public reputations for high quality research, especially internationally, and hence;
b) greater investment in publicising research achievements nationally and internationally and gaining scientific recognition; and
c) increased competition for publication in leading journals and for gaining project funding for research on currently significant topics.

Systematic and transparent RES are also likely to increase the visible stratification of researchers, universities and journals and encourage the standardisation of research and publication styles across the sciences and across local and national scientific schools. In some cases they may additionally discourage scientists from investing in long term, highly risky and interdisciplinary research, and from challenging current orthodoxies, as Schiene and Schimank suggest could happen in Lower Saxony if professors are forced to collaborate in research teams.

Such consequences seem especially likely when evaluations are centrally organised with rankings published in the press, conducted regularly, and have direct significant consequences for the allocation of public funds to research organisations on the basis of the quality of their research. In considering how different kinds of

RES are likely to affect the organisation and conduct of research in different public science systems and fields, it is useful to contrast two ideal types of weak and strong evaluation systems.

Weak RES can be identified in the following terms. They are typically organised informally by funding agencies and/or a consortia of universities with little standardisation of procedures or criteria. They rarely publish their conclusions, which are usually intended to encourage organisational improvements, and so be formative, rather than to arrive at summative judgements. Whilst these assessments focus on the overall effectiveness and quality of research groups and departments, they do not rank them according to an overall scale of international excellence, nor do they have direct financial consequences, although they may have some indirect ones in the medium term. While such RES may encourage scientists and universities to consider their research strategies more explicitly and to focus more on producing research that is recognised internationally as significant, their impact on current arrangements is likely to be incremental rather than radical, although they may lead to some shifts in the organisation of research, as perhaps is the case in Lower Saxony (Schiene and Schimank in this volume).

Strong RES, in contrast, institutionalise public assessments of the quality of the research conducted in individual departments and universities by scientific elites on a regular basis according to highly formalised rules and procedures. These assessments are usually ranked on a standard scale and published so that the relative standing of universities and departments can be readily ascertained. In most cases, they are organised around existing disciplines and scientific boundaries. Such peer-review-based evaluations directly affect funding decisions, often on a significant proportion of research organisations' income, and so can have a considerable impact on the management of universities and similar organisations.

The impact of developing and implementing research evaluations on knowledge production is likely to be especially noticeable when these are relatively 'strong' in this sense. Five major consequences of institutionalising such systems can be summarised in the following terms.

Firstly, by focusing attention on evaluations of the outputs of their work, researchers are likely to become more aware of the need to compete with others to gain recognition from scientific elites and coordinate their projects with those of others. This means that they will seek to contribute to the collective goals of their field as understood by current elites and so research in general should become more integrated around these goals as evaluation systems become more influential.

Secondly, as evaluators in these peer-review-based RES are forced to judge the relative merits of research outputs, they will develop and apply standard criteria of quality and intellectual significance for the field as a whole, thereby centralising judgements across individual researchers, universities and other research organisations. As they continue to do this on a regular basis, these standards and goals will become institutionalised as dominant in the field, and so the level of strategic task uncertainty, i.e. the degree of uncertainty about the intellectual importance of particular research strategies and outputs for collective goals (Whitley 2000: 123-124), should decline.

Thirdly, this centralisation and standardisation of research goals and evaluation criteria throughout scientific fields means that the diversity of intellectual goals and approaches within sciences should decline over time, especially where they challenge current orthodoxies. As evaluations become more important for both researchers and their employers, the costs of pursuing deviant strategies increase, and pressures to demonstrate how one's work contributes to dominant disciplinary goals will grow. These are especially strong for junior researchers who need to show the merits of their research as assessed by current disciplinary priorities and standards in order to gain employment and promotion.

Fourthly, such reinforcement of disciplinary standards and objectives is likely to inhibit the development of new fields and goals that transcend current intellectual and organisational boundaries by increasing the risks of investing in research projects that do not fit within them. Increasing competition for reputations and resources based on them, resulting from strong evaluation systems heighten the risks of moving into novel areas and adopting techniques and frameworks from other fields. Intellectual innovations will therefore tend to be focused on current sciences and their concerns. Radical intellectual and organisational innovation is thus less likely in societies that have strong, institutionalised research evaluation systems because these reinforce conservative tendencies in determining intellectual quality and significance.

Finally, the standardisation, formalisation and publication of quality rankings intensify the stratification of individual researchers, research teams and employer organisations. By regularly conducting and publicising such judgements, strong evaluation systems heighten awareness of one's relative position in the science system and encourage both individual and organisational strategies to enhance them. This stimulates the scientific labour market and, over time, is likely to concentrate resources and the most valued skills in elite universities, as the UK RAE seems to have done. However, such effects will differ between academic systems organised in different ways as well as between different kinds of scientific fields. I now turn to consider how the consequences of different evaluation systems seem likely to vary between differently funded and organised public science systems.

THE CONSEQUENCES OF STRONG RESEARCH EVALUATION SYSTEMS IN DIFFERENT PUBLIC SCIENCE SYSTEMS

National public science systems do, of course, continue to differ in a wide variety of ways but there are six major features that can be expected to affect the operation of research evaluation systems and their impact on the organisation of research and knowledge production.

Firstly, despite a general tendency for states to increase the conditionality of their research funding by making more resources dependent on competitive bidding for research projects, one of the most important ways in which academic systems vary remains the extent and periodicity of discretion granted to scientists and their employing organisations. In particular, the more resources are allocated as block grants to research organisations, usually on the basis of existing commitments to

departments and laboratories but sometimes according to the number of students or, later, graduates (Hansen and Borum 1999; Liefner 2003), the more discretion researchers have over the selection of research topics and approaches and the more time they have to pursue ambitious intellectual goals. In block grant funding regimes, then, the ability to continue to conduct research does not greatly depend on scientists' performance in the short to medium term.

In contrast, highly project-based funding regimes, where scientists have to gain new funds for each new project from external sources (usually state agencies) on a competitive basis, shorten the feedback loop between research performance and resource allocation and increase the costs of project failure. As many states have reduced the proportion of funds for research in the public sciences allocated as block grants, in favour of external project funding, they have increased the importance of performance-based funding and reduced the time horizons of researchers' discretion, as well as facilitating direct state guidance of research goals through making funding dependent on projects' contributions to public policy goals (Braun 2003). This feature of research funding regimes can be termed the frequency and significance of project performance review.

Secondly, another important feature of funding regimes that has become more significant as states seek to use scientific research to deal with public policy issues and improve the competitiveness of national economics, is the incorporation of state and other public policy objectives - such as the development of new technologies and industries - into funding procedures (Ruivo 1994; Braun 2003). Whether this is attempted through formal contracts, through the use of public problem solving criteria as additional selection filters for evaluating competent project proposals, or by involving user groups in the development of research programmes, it encourages the institutionalisation of multiple and often different research goals and standards for allocating funds that may legitimate and support use-oriented research careers and novel areas for study.

While such incorporation of public policy goals into project funding procedures can result more in ritualistic invocation of their desirability than substantive redirection of research activities towards problem solving goals – especially where cohesive disciplinary elites are able to control standards of research competence and significance – it can also limit the concentration of control over intellectual objectives and approaches exercised by such elites and lead to increased pluralism within cognate sciences. This is more likely when overall research funding in the public sciences is stable or declining in real terms and the proportion allocated for problem or use-oriented research is substantial, as Braun (2003) suggests. Overall, we can term this feature of funding regimes the importance of varied public policy goals in funding decisions, as distinct from those focused on intellectual purposes.

A third important feature of research funding systems concerns the variety of funding agencies and of their objectives. As Gläser and Laudel make clear in their contribution to this volume, a significant feature of the Australian public science system is the high level of dependence of scientists on a single agency, the Australian Research Council. In Britain by contrast, as well as in Sweden and other European states, there are often a number of different kinds of organisations providing funds for a variety of scientific research with diverse purposes, so that

scientists have more choice in raising money for their work. In the USA, of course, the variety of funding sources is even greater, including between different federal and local state agencies (Stokes, 1997). In general, the more varied the goals of different funding organisations with significant resources at their disposal, the less dependent researchers are likely to be on particular groups of disciplinary colleagues and agency staff, and so are the more able to pursue diverse intellectual purposes with novel research approaches. This feature can be described as the degree of external research funding diversity.

The effects of these three features of national funding regimes on the organisation of knowledge production are complicated by a fourth major difference in the organisation of public research systems: the organisational independence and capabilities of universities and other research organisations. While many states that organised universities as part of the central or local bureaucracies are now moving to a more decentralised system, their autonomy from central direction still differs considerably between, say, the UK and Spain. This means that universities' ability to exercise strategic independence and develop distinctive organisational capabilities in allocating resources, monitoring performance and pursuing distinctive purposes remains nationally variable, as is emphasised by Weingart and Maasen in their discussion of the 'enterprising' university in this volume.

In states where scientific staff and other employees are civil servants employed by national and regional governments on state-based terms and conditions, and the budgets of universities are largely determined centrally, their ability to develop specific competences through employment contracts in a comparable manner to private companies will be severely limited. In many ways, research organisations in much of continental Europe and Japan have, at least until very recently, been administrative shells in which most financial and managerial decisions are taken by ministries of education and science, while most decisions about research and teaching are made by senior academics, both individually and through faculty committees, with appointments and promotions being decided jointly (Clark 1995; Coleman 1999; Muller-Camen and Salzgeber 2005). As a result, the organisational autonomy, cohesion and distinctiveness of universities and other research organisations in such states remain relatively restricted, especially with respect to reallocating resources between fields, creating new departments and closing others down.

As Liefner (2003) suggests, the combination of general block grant funding of universities with high levels of state co-ordination of finance and administration has been associated with incremental organisational innovation and relatively slow rates of structural change. Once established in state budgets, department and institutes were largely left to pursue their own intellectual goals in many European universities, with little state or organisational steering or evaluation. This relatively high level of intellectual autonomy from bureaucratic and university pressures enjoyed by incumbent researchers – or at least by heads of research groups – was accompanied by limited flexibility in accommodating demands for new scientific fields and approaches, especially when funds became restricted. Because universities and similar organisations had limited resources that they could control

themselves, and often were constrained by incumbent professors, their ability to innovate organisationally in this situation was quite weak.

Where - in addition - authority structures within research organisations are highly stratified and administrative boundaries are coterminous with units of knowledge production – as in the German Institute System (Clark 1995; Muller-Camen and Salzgeber 2005) – researchers become highly constrained to follow the research goals and methods of their organisational superiors. Furthermore, if career mobility additionally occurs more within universities than between them, as in much of the post-war Japanese national academic system (Coleman 1999; Kneller in this volume; Sienko 1997), both intellectual and organisational innovation seems likely to be limited.

In other countries, notably the UK and USA, but also in Australia and Canada, universities have historically had much greater autonomy in allocating resources, even when the bulk of their resources have come from the state, and directly employed staff, sometimes on varied terms and conditions. The significance of university administrations has been correspondingly greater than in the previous situation, particularly in the USA, and each organisation has been more able to develop distinctive collective competences than in the more centralised academic systems. Internal resource allocation, appointments and promotions, opening and closing departments and similar decisions are here matters for individual research organisations and so they are able to cross subsidise research groups and respond to state policies in more autonomous and varied ways.

This feature of public science systems obviously depends on the overall nature of the state funding regime. Where, for example, university funds for research become increasingly dependent on external project based finance from one or two state agencies, and they have declining control over both the amount they receive for teaching and how it is allocated, their actual strategic autonomy will be limited even if they have considerable formal independence from the state. In Spain, for instance, Cruz-Castro and Sanz-Menendez, in their chapter, suggest that the inability of universities to fund research has made them incapable of acting strategically. Also, in Australia the considerable formal independence of universities has become greatly reduced in practice by both funding cuts and high dependence on project grants and evaluation-based funding, according to the chapter by Gläser and Laudel in this volume (see also Marginson and Considine 2000). It remains to be seen whether current changes in many European countries and Japan will similarly combine increased formal autonomy with considerable de facto dependence on state policies (Engwall and Nybom; Kneller, and Weingart and Maasen in this volume).

Two further features of public science systems that are likely to influence the consequences of introducing strong research evaluation systems are: a) the cohesion and prestige of scientific elites in each society – and general social standing of scientists and scientific knowledge – and, b) the segmentation of universities and other research organisations in terms of their funding, objectives and labour markets (Whitley 2003a). The first reflects the ability of scientific elites to organise themselves effectively as an interest group and to control the ways in which intellectual prestige and resources are allocated between institutions and fields. It is considerably assisted by a concentration of prestige and resources in a few elite

universities and other research organisations at the apex of a relatively stable intellectual and social hierarchy, such as that dominated by the universities of Cambridge and Oxford in the UK and those of Tokyo and Kyoto in Japan.

The organisational segmentation of research purposes, careers and labour markets in public science systems refers to the strength and stability of the organisational separation of research organisations dedicated to the pursuit of different kinds of knowledge production, such as theory driven knowledge, use-oriented research and technology transfer. In highly segmented research systems, scientists do different kinds of work, are evaluated on different criteria and have distinctively separate career paths. They are also discouraged from moving between research organisations and rarely compete in the same labour markets. As a result, relatively few are willing to leave the university sector to establish their own companies exploiting the knowledge produced by their research because of the difficulty of returning to academic posts in the event of their firm failing, especially in very high risk sub-sectors such as therapeutic biotechnology (Casper 2000; Casper and Whitley 2004).

In contrast, where organisational segmentation between universities, applied research organisations and private companies is lower, as in much of the USA, mobility tends to be greater. Additionally, US universities seem to be able to incorporate a greater variety of intellectual purposes and activities than do those in other countries, and historically they have been much more responsive to market opportunities (Clark 1995; Kenney 2000; Casper 2006; Mowery et al. 2004). On the whole, the weaker is such segmentation, the easier it should be to establish new scientific fields, especially those falling within 'Pasteur's quadrant', such as computer science, when the state or other organisations are willing to invest in them (Mowery 1999).

Variations in these six features of the organisation of public science systems are likely to affect the ways in which, and the degree to which, the implementation of strong research evaluation systems will affect knowledge production. Their reduction or reinforcement of the five main kinds of consequences outlined above are summarised in table 1 and will now be discussed in a little more detail.

Beginning with the frequency and importance of assessments of research goals and performance through competitive bidding for research project grants, the introduction of strong RES into academic systems that require researchers to compete intensively for external grants is likely to produce many of these consequences, especially where funding is concentrated in one or two agencies and project evaluations are organised around similar intellectual boundaries as those used by assessment exercises. The importance of peer review – and hence of elite conceptions of significant problems and appropriate ways of tackling them – will be greatly enhanced by such assessment systems in this situation. Competition for disciplinary and specialty reputations will be intense and scientists highly constrained to coordinate project goals and approaches with those of national and foreign colleagues. Research organisations (ROs) will also compete intensively for reputations and funds, and the stratification of researchers, departments and universities be further encouraged.

Table 1. Mediating influences of public science system characteristics on the consequences of strong RES

Characteristics of Public Science Systems	*Major Consequences of Strong RES*				
	Increasing Research Organisation Stratification	*Intensification of Reputational Competition and of Coordination of Research Goals*	*Strengthening of Central Disciplinary Standards and Priorities*	*Reduction of Intellectual Diversity and Pluralism*	*Increasing Constraints on Establishing New Fields and Approaches*
High level of project-based funding and frequency of performance review	Reinforced	Reinforced	Reinforced	Reinforced	Reinforced
High level of public policy and use-oriented programme funding	Reduced	Reduced	Reduced	Reduced	Reduced
Varied, diverse funding agencies and goals	Reduced	Reduced	Reduced	Reduced	Reduced
High level of research organisation strategic autonomy	Reduced in elite universities with control over their own resources	Reduced in elite universities with control over their own resources	Reduced in elite universities with control over their own resources	Depends on significance of RES funding	Reduced in elite universities with control over their own resources
Strong and cohesive scientific elite implementing RES	Reinforced	Reinforced	Reinforced	Reinforced	Reinforced
Segmentation of research organisation goals and careers	Reinforced in academic ROs	Reinforced in academic RO's	Reinforced in academic RO's	Reinforced in academic RO's	Reinforced in academic RO's

Such competition for favourable evaluations from scientific elites for both project funds and the outputs from research activities seem likely to limit the ability of universities and similar organisations to implement distinctive research strategies

that challenge current orthodoxies. The more dependent they become on both *ex ante* and *ex post* peer reviewed judgements, the less intellectual autonomy they are able to exercise and the more focused they will become on achieving highly favourable assessments, especially when rankings of individual departments and fields are published. While some cross subsidy of less central scientific fields and unorthodox approaches may be feasible for elite organisations in the short term, this seems unlikely to be tenable as a long-term commitment for most research organisations in this situation.

The mutual reinforcement of project based finding and strong retrospective evaluation systems can be expected, then, to limit university independence in pursuing deviant goals with unorthodox methods, and scientists will be discouraged from pursuing research careers that involve working on problems deemed uninteresting or developing theories that are regarded as wrong by the current scientific consensus, such as the theory of continental drift in the 1920s and 1930s (Hallam 1973). Even in public science systems that formally grant considerable autonomy to universities, then, this complementary impact of pre- and post-research evaluations would make it difficult for them to avoid implementing their results more or less directly and discourage the establishment of wholly new scientific fields that challenge dominant scientific ideals. While perhaps not completely realised in any funding regime as yet, this combination could become established in the UK if, as the Royal Society has advocated, all public research funds are allocated through the Research Councils and there is no block grant funding of academic research in UK universities.

Where researcher dependence on project based research funding is lower, the effects of introducing strong evaluation systems may be less marked, especially where universities have greater levels of organisational autonomy and distinctiveness and there are a variety of funding agencies. Where they can control employment policies and financial allocations, research organisations are, in principle, able to invest in different kinds of scientific fields and switch resources between topics in response to changing opportunities. They should therefore be more organisationally innovative than in centralised state co-ordinated academic systems, and more diverse in terms of the fields they research and the kinds of approaches used. Some intensification of reputational competition and of national and international coordination of research goals can be expected to result from the implementation of strong evaluation systems in these kinds of public science systems, thus reducing the potential diversity of research styles and length of projects, but this should be incremental rather than radically transforming.

Considering next the role of major state and/or other funding bodies' provision of project funds for strategic research programmes that contribute to public policy goals and problems, this seems likely to mitigate the centralising and standardising effects of strong evaluation systems, especially if resource allocation decisions involve non-scientific groups and criteria. Conflict between assessments based on the goals and standards of disciplinary elites and those tied more to the objectives of project funding agencies focused on public policy problems could result in increased differentiation within and across universities between groups and departments focused on primarily disciplinary research goals and those more concerned with

use-oriented ones. It could also reduce the attractiveness of research careers based on contributions to use-oriented research programmes.

While strong evaluation systems may well intensify reputational competition for contributions to disciplinary goals and encourage intellectual coordination of research goals and approaches in such contexts, this seems less likely where assessments are implemented by problem or use-oriented scientific groups as well as by disciplinary ones, and substantial project resources are available from agencies following public policy goals that cross disciplinary boundaries. Here, their impact should be correspondingly weakened, and research organisations more able to exercise strategic autonomy in developing distinctive research goals and allocating resources between groups and approaches.

More generally, the more diverse and varied are funding agencies' goals and associated criteria for assessing project proposals, the weaker we would expect the effects of implementing strong RES to be, except perhaps for their impact on universities' research management. This is because scientists' dependence on peer review judgements of the quality of their research will be limited by their access to different funding sources that judge the worth of intellectual contributions in contrasting ways. Thus, scientific elites will not be so able to dominate the allocation of resources by monopolising assessment criteria and procedures.

In particular, greater pluralism of research agency goals and evaluation standards should enable scientists to develop new fields and approaches, as well as careers based on their contributions to these, which may not be highly valued by current discipline-based elites controlling quality judgements. Depending on the balance of resource allocation between research assessment-based funding and that provided by agencies with diverse objectives and standards, researchers may be able to establish careers in more problem-oriented fields despite the development of strong, discipline-elite dominated evaluation systems.

As the previous discussion has emphasised, the actual strategic autonomy of universities and other research organisations depends greatly on how funding is organised, but it is worth considering how high levels of organisational autonomy might influence the impact of strong evaluation systems on knowledge production as a separate mediating factor. Where such independence reflects elite status and control of significant resources, through for example endowments and control over other kinds of resources that do not depend on state support or peer review, we might expect universities to pursue distinctive strategies and support more diverse intellectual goals and approaches than would be viable for lower status organisations.

While they may not always do so, such prestige and independent control of significant resources could enable them to invest in long term and risky projects, as well as recruiting and supporting scientists who do not wish to follow established approaches. At the very least, elite research organisations should be able to continue to employ staff researching unfashionable topics that are outside the intellectual mainstream of the field. They should additionally be in a position to support research that transcends established disciplinary boundaries and contributes to the development of new scientific fields. Thus, elite organisations that have some autonomy and independently controlled resources could be expected to limit many of the general consequences of introducing strong RES, although many may choose not to do so.

Considering next the cohesion and prestige of scientific elites, this affects the likelihood that they will be able to influence the organisation and direction of evaluation systems. Powerful elites are likely to design them to favour the best research in terms of their conceptions of quality and contribution to their central intellectual objectives. They can be expected to prefer evaluation systems based on peer review judgements since these enable disciplinary elites to set assessment standards. As a result, where elites are cohesive and able to influence the development and implementation of evaluation systems, we would expect such assessments to intensify further reputational competition and coordination of research goals around elite intellectual goals, as well as reinforcing current disciplinary and specialty boundaries.

This combination of powerful scientific elites and strong evaluation systems would inhibit the establishment of new scientific fields based on radically novel goals and approaches, as well as limiting the attractiveness of use-oriented research careers. Whatever the public genuflections to technological and public policy goals that sometimes accompany the implementation of research evaluation systems, strong scientific elites should, in practice, be able to establish their performance criteria as the dominant ones, as seems to have happened in the UK RAE and in the evaluation system developed in Lower Saxony.

When, in addition, universities and other research organisations are organised into a strong and stable hierarchy of intellectual and social prestige, such elite influence over the design and operation of research evaluation should reinforce their stratification, as leading scientists would tend to be based in the top organisations and quality based evaluation should encourage concentration of resources in them, because the elite established the standards being used. This is even more likely when research funds are concentrated in a small number of agencies that rely on peer review for allocating resources. In strongly hierarchical academic systems, this will reinforce existing distinctions and criteria for evaluating research outputs, as seems probable in Japan (Kneller in this volume). In contrast, where there is a plurality of different kinds of research support agencies pursuing diverse intellectual and policy objectives, research groups and universities will be more able to gain resources for different kinds of purposes and the stratification of research organisations may not be so stable, as is arguably the case in the postwar USA (Stokes 1997).

Turning next to consider the likely mediating effects of organisational segmentation of research purposes, careers and labour markets on the effects of strong evaluation systems, high levels of such segmentation can be expected to reinforce many of these in the major research universities. This is because the organisational and labour market separation of ‘pure’ research from more use-oriented work limits the feasibility of pursuing varied intellectual objectives and obtaining funding for diverse kinds of research within academic research organisations. Since strong research evaluation systems focusing on peer review assessments of quality are likely to be implemented by leading academic scientists in highly segmented public science systems, their standards and goals will dominate the evaluation system and so intensify reputational competition amongst university researchers.

Insofar as new scientific fields and approaches are encouraged by use-oriented research programmes, and multiple funding agencies pursuing varied objectives do exist in a public science system, then, high levels of organisational segmentation will tend to ensure that these pressures for greater diversity will be channelled into specialist research organisations rather than impinging greatly on the leading research universities, as perhaps has happened in postwar Germany. This suggests that establishing new research areas that cross disciplinary boundaries and involve the use of new kinds of intellectual approaches will be difficult in highly segmented public research systems, at least in the major universities.

In general, then, many of the expected consequences of implementing strong RES should be considerably reduced in public science systems that combine considerable diversity of research funding agencies and foundations with high levels of university autonomy and strategic capacity based on control of their own resources. Especially for elite research organisations in societies that accord them considerable social prestige and independent access to finance, the largely intellectually conservative implications of adopting strong evaluation systems may be restricted. Conversely, where research funding is available from only one or two state agencies and private funds are in short supply, and elite universities have limited social and political support, these consequences can be expected to be quite marked, as Gläser and Laudel suggest has been the case in Australia in their chapter in this book.

THE EFFECTS OF STRONG RESEARCH EVALUATION SYSTEMS ON DIFFERENT KINDS OF SCIENTIFIC FIELDS

Just as strong research evaluation systems are likely to reinforce many conservative tendencies in national public science systems, so too they can be expected to encourage greater elite coordination of research goals and standards within established scientific fields, especially where intellectual elites have to rank the quality of research outputs publicly. By institutionalising and standardising such judgements across researchers in universities and other research organisations, national research evaluation systems force leading scientists to set intellectual priorities for disciplines and research specialties that are likely to lead to increased integration of research strategies and projects within these fields. However, the intellectual and social organisation of different sciences varies considerably, and changes over time, in ways that can be expected to affect these and other consequences of implementing research evaluation systems (Whitley, 2000).

In particular, scientific fields differ in: a) the diversity of agencies and organisations funding research and of legitimate audiences for research results; b) the extent to which researchers are able to obtain externally provided project-based funding; c) the extent of local organisational variability of research goals and frameworks; d) the tendency of researchers to concentrate on a limited number of central problems and approaches as opposed to studying a wide range of different topics and problems; e) their centrality and prestige in the broad hierarchy of the sciences; and f) the extent of elite cohesion and consensus on intellectual goals,

research methods and the value of results. How these characteristics are likely to mediate the effects of strong RES, and their possible change as a result of implementing strong RES, are summarised in tables 2 and 3 and will now be further discussed.

Table 2. Likely Effects of Different Characteristics of Scientific Fields on the Consequences of Implementing Strong RES

Characteristics of Scientific Fields		*Mediation of Impact of Strong RES*
Diversity of Funding Sources and Legitimate Audiences for Research Reputations	High	Reducing
	Low	Increasing
Variability of Research Goals and Approaches	High	Reducing
	Low	Increasing
People to Problem Ratio	High (urban sciences)	Increasing
	Low (rural sciences)	Reducing
Elite Structure	Cohesive and agreed on core goals and approaches	Increasing
	Heterogeneous	Reducing
Centrality and Prestige of fields	High	Increasing
	Low (peripheral fields)	

Table 3. Likelihood of strong RES affecting characteristics of scientific fields

Characteristics of Scientific Fields		*Changes in characteristics due to strong RES*
Diversity of Funding Sources and Legitimate Audiences for Research Reputations	High	None
	Low	
Variability of Research Goals and Approaches	High	Diminishes because of increased coordination of research goals
	Low	Reinforced
People to Problem Ratio	High (urban sciences)	Reinforced
	Low (rural sciences)	Growing because of reduced reputations for work in marginal areas

Characteristics of Scientific Fields		*Changes in characteristics due to strong RES*
Elite Structure	Cohesive and agreed on core goals and approaches	Reinforced
	Heterogeneous	Increasing fragmentation of fields and control of resources
Centrality and Prestige of fields	High	Further increased because of strengthening stratification of fields; reinforce elite control and ideals of science
	Low (peripheral fields)	Further diminished because of strengthening stratification of fields; imitation of dominant research styles and norms

Considering first the research funding characteristics that also apply to public science systems as a whole, we would expect similar consequences for individual sciences as those for national research systems. Fields where researchers are able to obtain research funds from diverse organisations and have opportunities for seeking legitimate intellectual reputations for contributions to varied goals and audiences are less likely to be greatly affected by strong evaluation systems than are those where scientists are highly dependent on funding based on the judgements of such assessment systems. Thus, sciences where there is a diversity of goals and frameworks that are supported by different funding agencies and organisations will be less tightly coordinated around disciplinary objectives as a result of strong evaluation systems being implemented than will those with more limited sources of support. In particular, fields that are able to gain funding for use-oriented research, such as engineering, medicine and many of the policy and practice-oriented social sciences, should be less affected by the establishment of formal and consequential research evaluation systems than fields more dependent on a small number of state funding agencies.

Additionally, fields where scientists are able to gain most of the resources required for their work from external agencies such as national academies and research councils are less likely to be greatly affected by the implementation of strong RES than are those more dependent on RES-based assessments. While many of the referees and members of decision making committees in the former may well also be involved in the operation of RES, and so the criteria and standards of the latter will often reinforce those used in project selection, the impact of strong evaluation systems is likely to be incremental and limited since the proportion of total research support governed by them in this situation is relatively low.

Conversely, the impact of strong evaluation systems on scientists' research activities should be more marked in sciences where major resources – especially those funding the time of academics – depend on their judgements and where researchers have not hitherto been subject to regular assessment through peer review. Here, the effects of introducing systematic evaluations of research quality that affect universities' budgets directly are likely to be much more visible, resulting in qualitative increases in reputational competition and in the level of intellectual coordination of research goals and approaches across research sites. At least in terms of public debate and the selection and organisation of research projects, they can be

expected to have stronger consequences in these kinds of scientific fields than in those where scientists are more able to gain resources from external funding bodies.

Similarly, such assessment systems should significantly increase the degree of intellectual coordination and integration of research around common topics and concerns in sciences where goals and approaches have tended to differ between groups in different universities, such that separate schools of thought have developed and been reproduced in different research institutes. Especially where local traditions in, say, philosophy, have been important sources of intellectual variation and loyalty, the establishment of systematic, public and consequential evaluations of research quality are likely to encourage standardisation of research styles and strategies, as well as increased agreement about important research problems and contributions. This seems to have happened in Dutch philosophy when funding became linked to evaluations of international excellence (Van der Meulen and Leydesdorf 1991). In fields that already manifest considerable national and international coordination of research objectives and styles, on the other hand, the introduction of systematic research evaluation systems may simply reinforce this incrementally rather than leading to major shifts in problems and approaches.

This more marked influence of strong RES on sciences with relatively low levels of intellectual integration can also be expected to apply to more 'rural' fields that have low researcher to problem ratios. As summarised by Becher and Trowler (2001: 106-107), these kinds of science tend to be more diverse in the range of problems tackled by researchers and less concentrated in terms of the proportion of scientists working on the same or closely related issues. In contrast, 'urban' fields are those in which researchers focus intensively on a limited number of central problems and compete for reputations on the basis of their contributions to dealing with these.

Implementing strong RES in which disciplinary elites are forced to make judgements about the merits of different kinds of problems and contributions to them seems quite likely to reduce the diversity of topics and approaches deemed to be within the discipline in highly rural sciences, and so encourage more intellectual coordination and competition as researchers respond to this hierarchical arrangement of problems. Over a series of such evaluations, research on certain issues will come to be seen as more important and rewarded than that on others and so competition between scientists for the significance of their contributions to these core problems will grow. In fields that are already quite urban in this sense, on the other hand, strong RES should intensify competition and coordination.

The extent of this effect will, though, depend on the cohesion of scientific elites and their consensus on the central problems of the discipline. Where this is considerable, then we would expect strong RES to reinforce their authority and their control over the intellectual agenda of the field, especially where RES-based funding constitutes a major proportion of the resources needed for carrying out research. Since they will typically implement the RES, they will be able to determine the quality standards and intellectual priorities for judging research outputs and so intensify competition for reputations based on contributions to the central problems of the discipline. Research that deviates from current orthodoxies is unlikely to be regarded as worthwhile, or even competent, in these circumstances, and so intellectual pluralism will decline further.

In contrast, where elites are more differentiated, if not fragmented, in their intellectual goals and approaches, and there is limited consensus on the key problems of the field, the development of a strong RES may help to reproduce such separation of specialisms and topics rather than intensifying competition and coordination across the discipline. As long as no school or group is able to dominate the process and centralise control of key resources, the leaders of the major subfields could tacitly or explicitly agree to recognise each others goals and standards as equivalently high quality and so rank the 'best' research outputs dealing with each of these as excellent for the purposes of the evaluation exercise. Essentially, the RES would here increase coordination within each subfield and intellectual school but not strengthen integration between them, thus confirming the existing intellectual power structure in the discipline. This outcome is especially likely if the concentration of control over key research resources is low so that each group can gain access to them relatively easily.

Finally, just as we would expect the response of universities and other research organisations to strong research evaluation systems to vary according to their prestige and control over resources, so too the relative prestige and centrality of scientific fields should make a difference to their impact. Since such evaluation systems are likely to reflect the conceptions and standards of leading researchers in dominant scientific fields, and to be organised and implemented in ways that are complementary to current elites' perceptions of the best kinds of work, strong RES are more likely to reinforce their views, and the degree of elite control over research agenda and priorities, than to follow the research styles and goals prevalent in peripheral sciences. As they impinge upon funding decisions and university research strategies, they will therefore encourage the less prestigious fields to imitate the procedures and approaches – at least superficially – of the more central ones.

The impact of strong RES on marginal fields in terms of increasing competition and coordination should, then, be more visible than that on more central ones, especially when scientists depend greatly on RES-based funding. Hierarchies of journals and devaluation of alternative publication media for research results, formalisation of ranking systems, and increased awareness of 'international' standards will become more institutionalised in these kinds of fields, largely in imitation of the more successful sciences and current perceptions of their norms. Again, these might be expect to be especially marked in the less prestigious and autonomous research organisations, whose managers feel compelled to follow the 'rules of the game' established by the leading organisations and agencies, and are unable to insist on the significance of differences between the sciences. Reliance on bibliometric indicators of intellectual impact seems likely to reinforce this narrowing of research styles as well as having other consequences as discussed by Gläser and Laudel in this volume.

This may lead to changes in prevalent research styles and skills in the more peripheral sciences as traditional elites are challenged by more 'scientific' colleagues who seek to institutionalise common goals and standards as means of integrating research and demonstrating the scientific status of the field. In some cases, this search for scientific respectability has led to calls for the establishment of Kuhnian-style paradigms to deliver intellectual progress, apparently regardless of

their merits – any candidate will do (Pfeffer 1993). Depending on the cohesion and strength of current elites in a field, as well as the importance of evaluation system-based funding and reputations, this means that the establishment of strong research evaluation systems could result in qualitative shifts in elite membership and purposes in the less prestigious sciences.

CONCLUSIONS

This discussion has emphasised a number of points about the development and effects of establishing RES in the public sciences that can be summarised as three main conclusions. First, such retrospective evaluation systems are highly variable and path dependent in the sense that both how these institutional innovations in state-science relations became institutionalised and changed in different academic systems, and how they are affecting the direction and organisation of scientific research, strongly reflect the differing organisation of the public science system in each country, as well as the overall structure of professional labour markets and certain characteristics of the state. Differently organised groups of politicians, elite civil servants, business elites and other influential policy advisers have developed contrasting RES in varied ways in different kinds of state systems, and their effects have been mediated by varying characteristics of national academic systems, as the chapters on Australia, Germany, the Netherlands, Spain and Sweden in this book demonstrate.

Thus, although imitation and borrowing of particular aspects of RES have occurred between many of the OECD countries, sometimes to a considerable extent, as many contributions to this book demonstrate, their implementation in practice has been greatly affected by the continued substantial variations in how state agencies allocate funds for scientific research and higher education, how scientific elites are organised and manage relations with politicians and the state bureaucracy, and how the employment and careers of researchers are organised.

Second, a crucial factor in examining the development and consequences of RES in the public sciences is the changing nature and behaviour of universities and other research organisations in many of the societies where they have been introduced. How these have reacted to, and influenced, the implementation of different kinds of evaluations have had major effects on the organisation of scientists' careers, and professional labour markets more generally, as well as on the organisation and direction of research groups and fields. As many RES have been established as part of a general process of delegating some financial and administrative authority and resources to universities and similar organisations, they have become more significant collective actors in the development and organisation of scientific knowledge, and mediated the consequences of state policies, such as the introduction of RES, to a much greater extent in many countries than hitherto.

Although the extent of this process of financial and administrative decentralisation has varied considerably between countries in Europe and elsewhere, and in some is more apparent than real, to the extent that it has produced qualitative shifts in resource control and strategic autonomy from the state to research organisations, then how this autonomy has been used by different groups of

managers and researchers becomes more important than when they largely implemented the decision of ministries and faculties. This is especially so if some develop a more independent research policy that does not simply reproduce the results of peer review judgements in the RES and project proposal assessments.

How universities and other research organisations develop different kinds of research management capabilities as separate employment organisations becomes much more significant for the organisation and development of the public sciences as they gain more independence and control their own resources. While unlikely perhaps to generate such distinctive, organisation-specific competences through authority relations and managerial hierarchies as private companies in market economies are seen to do by adherents to the resource-based theory of the firm (see e.g. Metcalfe and James 2000; Penrose 1959; Whitley 2003b), the combination of RES and managerial delegation to research organisations makes the management and behaviour of universities and research institutes more consequential for scientific research than before in many countries. However, the actual autonomy and strategic capabilities developed by many universities and other research organisations in Continental Europe and Japan seems rather limited so far.

Third, while there are some general implications of adopting strong forms of RES for the sciences in general, especially where they are organised around established disciplinary boundaries and elites, their impact can be expected to vary considerably across different kinds of scientific fields, often in ways that are interdependent with particular features of national public science systems. Whatever kind of RES is implemented and changed in a country, it will complement and be more appropriate for some kinds of sciences and research styles than other ones, and so will potentially change some fields more than others.

Since the sciences vary in many ways, as well as changing over time as some fields become more iconic than others and come to dominate implicit and explicit hierarchies of scientific prestige, any evaluation system that tries to standardise assessments of the productivity and quality of scientific research is bound to have differential effects across the sciences. Whether this will in time lead to greater standardisation of research styles, publication practices and the organisation of scientific work depends, as I have suggested, on the nature and importance of the RES, how it is implemented, and on the nature of the public science system and behaviour of leading universities and other research organisations in each country.

REFERENCES

Becher, Tony and Paul Trowler (2001), *Academic Tribes and Territories: Intellectual enquiry and the culture of disciplines*, Buckingham: Open University Press and the Society for Research into Higher Education. Second edition.

Block**,** Hans-Juergen (1990), 'The University Systems in Transition: Possibilities and limitations of universities in the 'steady state', in Susan E. Cozzens, Peter Healey, Arie Rip and John Ziman (eds.), *The Research System in Transition*, Dordrecht: Kluwer, pp. 35-50.

Braun, Dietmar (1998), 'The Role of Funding Agencies in the Cognitive Development of Science', *Research Policy,* 27, 807-821.

Braun, Dietmar (2003), 'Lasting Tensions in research Policy-Making – A delegation problem', *Science and Public Policy,* 30, 309-321.

Brooks, Harvey (1990), 'Lessons of History: Successive challenges to science policy', in Susan E. Cozzens, Peter Healey, Arie Rip and John Ziman (eds.), *The Research System in Transition*, Dordrecht: Kluwer, pp. 11-22.

Campbell, David F. J. (2003), 'The Evaluation of University Research in the United Kingdom and the Netherlands, Germany and Austria', in Philip Shapira and Stefan Kuhlmann (eds.), *Learning from Science and Technology Policy Evaluation. Experiences from the United States and Europe,* Cheltenham: Edward Elgar, pp. 98-131.

Casper, Steven (2000), 'Institutional Adaptiveness, Technology Policy and the Diffusion of New Business Models: The Case of German Biotechnology', *Organization Studies,* 2, 887-914.

Casper, Steven (2006), 'How Do Technology Clusters Emerge and Become Sustainable? Social network formation and inter-firm mobility within the San Diego biotechnology cluster', Working Paper: Keck Graduate Institute of Applied Life Sciences.

Casper, Steven and Richard Whitley (2004), 'Managing Competences in Entrepreneurial Technology Firms: A comparative institutional analysis of Germany, Sweden and the UK', *Research Policy,* 33, 89-106.

Clark, Burton R. (1995), *Places of Inquiry: Research and advanced education in modern universities*, Berkeley: University of California Press.

Coleman, Samuel (1999), *Japanese Science: View from the inside*, London: Routledge.

Cozzens, Susan E. (1990), 'Options for the Future of Research Evaluation', in Susan E. Cozzens, Peter Healey, Arie Rip and John Ziman (eds.), *The Research System in Transition*, Dordrecht: Kluwer, pp. 281-294.

Dasgupta, Partha and Paul A. David (1994), 'Toward a New Economics of Science', *Research Policy*, 23, 487-521.

Drori, Gili S., John W. Meyer, Francisco O. Ramirez and Evan Schofer (2003), *Science in the Modern World Polity: Institutionalization and globalization,* Stanford: Stanford University Press.

Freeman, Christopher and Luc Soete (1997), *The Economics of Industrial Innovation*, London: Pinter.

Gaughan, Monica and Stephane Robin (2004), 'National Science Training Policy and Early Scientific Careers in France and the United States', *Research Policy*, 33, 569-581.

Hallam, A. (1973), *A Revolution in the Earth Sciences: From continental drift to plate tectonics*, Oxford: Oxford University Press.

Hansen, Hanne Foss and Finn Borum (1999), 'The Construction and Standardization of Evaluation: The case of the Danish university sector'. *Evaluation*, 5, 303-329.

Kenney, Martin (ed.) (2000), *Understanding Silicon Valley: The anatomy of an entrepreneurial region*, Stanford: Stanford University Press.

Langfeldt, Liv (2001), 'The Decision-Making Constraints and Processes of Grant Peer Review, and Their Effects on the Review Outcome'. *Social Studies of Science*, 31, 820-841.

Liefner, Ingo (2003), 'Funding, Resources Allocation, and Performance in Higher Education Systems', *Higher Education*, 46, 469-489.

Marginson, Simon and Mark Considine (2000), *The Enterprise University: Power, governance and reinvention in Australia*, Cambridge: Cambridge University Press.

Martin, Ben (2003), 'The Changing Social Contract for Science and the Evolution of the University', in Aldo Geuna, Ammon J. Salter and W. Edward Steinmuller (eds.), *Science and Innovation: Rethinking the rationales for funding and governance*, Cheltenham: Edward Elgar, pp. 7-29.

Martin, Ben and Aldo Geuna (2003), 'University Research Evaluation and Funding: An international comparison'. *Minerva,* 41, 277-304.

Metcalfe, J. Stanley and Andrew James (2000), 'Knowledge and Capabilities: A new view of the firm', in Nicolai Foss and Paul Robertson (eds.), *Resources, Technology and Strategy: Explorations in the resource based perspective*, London: Routledge, pp. 31-52.

Mowery, David C. (1999), 'The Computer Software Industry', in David C. Mowery and Richard R. Nelson (eds.), *Sources of Industrial Leadership*, Cambridge: Cambridge University Press, pp. 133-168.

Mowery, David C., Richard R. Nelson, Bhaven N. Sampat and Arvids A. Ziedonis (2004), *Ivory Tower and Industrial Innovation: University-Industry technology transfer before and after the Bayh-Dole Act,* Stanford, California: Stanford University Press.

Muller-Camen, Michael and Stefan Salzgeber (2005), 'Changes in Academic Work and the Chair Regime: The case of German business administration academics', *Organization Studies*, 26, 271-290.

Penrose, Edith (1959), *The Theory of the Growth of the Firm*, Oxford: Blackwell.

Pfeffer, Jeffrey (1993), 'Barriers to the Advance of Organization Science'. *Academy of Management Review*, 18, 599-620.

Ruivo, Beatriz (1994), ''Phases' or 'Paradigms' of Science Policy?', *Science and Public Policy*, 21, 157-164.

Sienko, Tanya (1997), *A Comparison of Japanese and U.S. Graduate Programs in Science and Engineering*, Tokyo: National Institute of Science and Technology Policy, Discussion Paper no. 3.

Stokes, Donald E. (1997), *Pasteur's Quadrant: Basic science and technological innovation*, Washington, D.C: Brookings Institution Press.

Tabil, Ameen Ali (2001), 'The Continued Behavioural Modification of Academics since the 1992 Research Assessment Exercise', *Higher Education Review*, 33, 30-46.

Van der Meulen, Barend and Loet Leydesdorff (1991), 'Has the Study of Philosophy and Dutch Universities Changed under Economic and Political Pressure?', *Science, Technology, and Human Values*, 16, 288-321.

Whitley, Richard (2000), *The Intellectual and Social Organization of the Sciences*, Oxford University Press. Second edition.

Whitley, Richard (2003a), 'Competition and Pluralism in the Public Sciences: The impact of institutional frameworks on the organisation of academic science', *Research Policy,* 32, 1015-1029.

Whitley, Richard (2003b), 'The Institutional Structuring of Organisational Capabilities: The role of authority sharing and organisational careers', *Organization Studies*, 24, 667-695.

Universities as Strategic Actors: State Delegation and Changing Patterns of Evaluation

CHAPTER 2

LARS ENGWALL AND THORSTEN NYBOM

THE VISIBLE HAND VERSUS THE INVISIBLE HAND

The Allocation of Research Resources in Swedish Universities

INTRODUCTION

The allocation of resources is a fundamental research problem within the social sciences, particularly within the economic sciences. The basic task is to allocate resources in such a way that the result is optimal for society. The allocation of research resources is no exception to this principle. However, it is somewhat more complicated, since research is associated with fundamental uncertainty and has demonstrated that it is characterised by long-term rather than short-term returns. Nevertheless, some form of central strategic control over resources and investment is necessary. In corporations, this control is exercised through research investment budgets and management control systems. For government-funded research, the control function is implemented through (a) institutional control, (b) input control, and (c) output control.

Among these, the first: *institutional control*, has to do with the acknowledgement of institutions and the conditions they have to operate under. It is exercised through chartering and institutional rules. Chartering refers to the power of government to recognise certain institutions and to deny others the status and privileges of a university or university college. This is essential for the trust of the institution and its access to research resources. Institutional rules, on the other hand, have a bearing on the rules that selected institutions have to follow as they are acknowledged as a university or university college.

The second category, *input control*, is achieved through a mechanism similar to the chartering – that is, *the screening of teachers and research personnel*. This control includes the procedures for selecting students, giving access to research careers, and selecting holders of top research positions. The relevance of these procedures is, of course, important for the efficiency of the research system. Researchers are not enough, however. They also need the financial resources to carry out their research. This leads to a demand for *a system of resource allocation* – procedures for channelling resources from governments and legislative bodies to the

R. Whitley and J. Gläser (eds.), The Changing Governance of the Sciences, 31–49.

researchers selected through the screening procedure. A significant issue in this context is whether resources should be allocated on the basis of block grants to institutions or on the basis of various types of individual and institutional competitive funding. A related issue concerns the time perspective of the funding.[1]

The third category, *output control*, concerns the way in which the outcome of research is followed up. Traditionally, this control has been exercised primarily through input control, that is, the output of individuals is evaluated in relation to hiring and promotion. Recently, this form of control of research has also been used as university colleges have applied for upgrading to university.[2] However, in addition to these evaluations there is an increasing tendency today to develop continuous systems of output control and to a varying degree link the outcome of these evaluations exercises to resource allocation.[3]

The presented framework will now be used to demonstrate changes in the system of allocation of research resources in Sweden. The basic question concerns how politicians over time have chosen strategies to govern the Swedish research system. Since Sweden largely subscribes to the Humboldtian model – that research and higher education should be combined – the changing patterns of steering and organisation of the higher education system – has to play a significant role in our analysis.[4]

INSTITUTIONAL CONTROL

Entry Control

Unlike banks that require a charter to use the label 'bank', the label 'university' is not protected in the same way. It is thus open to anybody who wants to start a university, and this mechanism is used to a considerable extent. In modern times, large corporations have thus tended to label their educational activities 'corporate universities' (e.g., McDonald's Hamburger University, see also Crainer and Dearlove 1999). However, although the use of the label 'university' is not restricted, governments play another significant role in guaranteeing quality through their control over the rights of institutions to award degrees, that is, significant quality control through entry control.

The entry control is particularly important for research in countries, like Sweden, where practically all non-commercial research is concentrated in universities.[5] To

---

[1] For a general discussion and analysis of this issue, see Richard Whitley´s contribution to this volume.

[2] For an example of such an evaluation, see Högskoleverket (1998).

[3] In addition, recent heated public controversies over certain research procedures and results with the ensuing tendency of intensified media intervention have led to a growing amount of oversight by ethical and other 'non-scientific' committees and bodies which have been introduced into the research funding system. For a general discussion of the consequences for different funding regimes and for different disciplines, see Richard Whitley´s contribution to this volume.

[4] For a similar historical case, see the contribution to this volume on the German case by Peter Weingart and Sabine Maasen.

[5] When a formal science policy was introduced in the mid-1940s, the Swedish Government and Parliament declared that Sweden would abstain from establishing extramural research institutes – like,

understand the development of quality control in the Swedish research system it is therefore very important to know about the institutional developments and structural features of the system for higher education and research. It is also important to note that the business sector plays an enormously important role for R&D in Sweden. Public funding thus accounts for less than 30% of the total, which together with Japan, represents the lowest public share among all OECD countries (OECD 2005).[6]

For many centuries, the Swedish university system consisted of only two universities, one founded in Uppsala in 1477 and a second in Lund in 1668.[7] Through private initiatives, university colleges were created in the latter part of the nineteenth century in the two largest cities of Sweden, Stockholm (1878) and Gothenburg (1891). Both institutions became state universities in the mid-twentieth century (1960 and 1954, respectively).

In 1965, a fifth university was founded in Umeå in northern Sweden. This event can also be said to have marked the starting point of a regionalisation of higher education. In the late 1960s, the existing universities were given the task to establish branches in Karlstad (Gothenburg), Linköping (Stockholm), Örebro (Uppsala), Sundsvall (Umeå) and Växjö (Lund). Among the latter, Linköping was upgraded in 1975 to university status and the other branches became independent university colleges. In 1997, the remaining institutions also applied for university status. In a peer review evaluation only Karlstad received approval.[8] Nevertheless, the Government decided in 1999 to upgrade all except Sundsvall to formal university status. The latter was granted university status in 2005, thus bringing the total number of universities up to ten.

In addition, Sweden has six professional schools in the fields of agriculture, management, medicine, and technology. They were originally not admitted to university status but acquired it in the course of time. They are the Karolinska Institute (founded in 1810, entitled to give examinations in 1861), the Royal Institute of Technology (founded in 1827, upgraded to academic status in 1877), Chalmer's Institute of Technology (founded in 1829 and given full academic status in 1937), the Stockholm School of Economics (1909), the Luleå Institute of Technology (founded in 1971 and upgraded in 1996) and the Agricultural University (1977, created by a merger of several institutions).[9] These institutions are not

---

for instance, the West German Max Planck Institutes. Instead the principle of university-based research was elevated to *the* particular 'Swedish model' of research organization. See also Nybom (1997).

[6] The implications and consequences of private (industry) funding are discussed in Elzinga (1993) and recently, at length, in Schilling (2005).

[7] As a matter of historical fact, the number of Swedish universities was larger when Sweden had a more geographical extension in the seventeenth century. Swedish universities were thus founded in Dorpat (Tartu) in Estonia and Åbo (Turku) in Finland in 1632 and 1640, respectively. Through the Peace of Westphalia in 1648, the University of Greifswald in Germany became Swedish and remained so until 1815.

[8] See Högskoleverket (1998).

[9] The information about founding years is taken from *Nationalencyklopedien*. The Agricultural University is a merger of several older colleges of agriculture, forestry, horticulture, and veterinary medicine. For historical reasons the Agricultural University is under the auspices of the Ministry of Agriculture. The university has made efforts for a transfer to the Ministry of Education and Research, so far without success. It should also be pointed out that the Stockholm School of Economics is a

multidisciplinary in the same sense as the first ten but have the same rights to appoint professors and to run doctoral programmes. Thus, Sweden has in total sixteen institutions with university status.

The drive towards regionalisation has in the latter part of the twentieth century resulted in fourteen additional regional university colleges throughout Sweden, located in Borås, Eskilstuna/Västerås (Mälardalen), Falun/Borlänge (Dalarna), Gävle, Halmstad, Jönköping, Kalmar, Karlskrona/Ronneby (Blekinge), Kristianstad, Malmö, Skövde, Södertörn, Trollhättan/Uddevalla, and Visby.[10] In addition, there are ten colleges specialising in artistic performance, nurse training and teacher training, all located in Stockholm.[11]

Thus, altogether, the Swedish system of higher education today contains 40 institutions: 10 universities, 6 specialised schools with university status, 14 university colleges, and 10 specialised colleges.[12] They vary in their profile and in their traditions. However, they basically share the same governing principles in terms of resource allocation and government involvement. The major difference is that the 16 institutions with university status receive government block grants specifically intended for research and research education.

Entry control thus provides a basic screening mechanism of quality control. It determines the institutional actors in the field of research. This in turn is vital for resource allocation, i.e., the input control of research funding (see also the section on allocation procedures below).

Institutional Rules

With a few exceptions, Swedish universities and university colleges are formal government agencies under the jurisdiction of the Government and the Parliament (*Riksdag*) and are subject to the same body of regulations as any other government authority or agency. The basic documents regulating the universities and university colleges are the Higher Education Act (*Högskolelagen*) and the Higher Education Ordinance (*Högskoleförordningen*). These reforms were adopted in 1992 (SFS No. 1992:1434) and 1993 (SFS No. 1993:100), respectively and have since been revised a number of times. In their initial form, they replaced much more detailed regulations. The Higher Education Act consists of five short chapters: 1. Introductory Regulation, 2. The Organisation of Government Universities and

---

private institution with a special agreement on State funding, and that Chalmer's Institute of Technology since 1993 has been a foundation.

[10] Like Chalmer's Institute of Technology, the University College in Jönköping is a foundation. It has a parent foundation with four wholly owned limited companies for research and education.

[11] Danshögskolan, Dramatiska institutet, Idrottshögskolan i Stockholm, Konstfackshögskolan, Kungl. Konsthögskolan, Kungl. Musikhögskolan i Stockholm, Lärarhögskolan i Stockholm, Operahögskolan i Stockholm, Röda Korsets Högskola and Teaterhögskolan i Stockholm.

[12] These figures could be compared with the corresponding figures from the state of California, which has about four times as many inhabitants as Sweden. The University of California has ten campuses with more than 200,000 students; California State University has 23 campuses with more than 400,000 students and local colleges at more than a hundred locations. See also Trow (1998) and the website of the University of California at http://universityofcalifornia.edu/campuses/welcome.html.

University Colleges, 3. Professors and Other Teachers, 4. The Students, and 5. Special Regulations; altogether about ten pages.

The Higher Education Ordinance is much more detailed and longer, but not more than forty-two pages. The Act and the Ordinance also differ in status in that Government can change the Ordinance, while changes in the Act must go through Parliament.

Of the five chapters in the Higher Education Act, Chapter 2, on the organisation of universities and university colleges, has particular relevance for institutional governance. It states that such institutions shall have a board, and its actual composition has gone through a series of changes. Presently, the governing board of Swedish higher education institutions consists of 15 members. Government appoints the majority (8). The students and the teachers appoint three board members each, and the Vice-Chancellor (*rektor*) also sits on the board but no longer as chairperson. Since 1997, the chairperson has been appointed by government and is required to be 'a well-qualified and experienced person who is not employed at the institution in question'. Thus, both through the majority principle and through the appointment of the chairperson, government communicates a wish to control the strategic decisions of universities and university colleges. Political ambitions are also manifest in the choice of chairpersons. A study of the appointment of chairpersons at the sixteen universities and the fourteen university colleges as of January 1, 2004 revealed that 18 chairpersons (60%) had links to political parties. Of these three-quarters were associated with the government party (Engwall 2007). Thus, we note clear political ambitions in the governance of academic institutions.

It is quite clear that the traditional, and highly formalised and intermediary, central bureaucratic control and governance that characterised the Swedish higher education system until the late 1980s, and which had been, at least partly, abolished in the early 1990s, from the mid 1990s, was transformed into a system of more or less direct political intervention by the Social Democratic government. It is fair to say that during the last decade higher education institutions are less and less regarded as national 'cultural institutions', but rather primarily as means to achieve other important political and economic ends. At the same time, universities and public research founders have also explicitly been ordered to include certain politically defined values ('quality' 'sustainability' and 'ethnical inclusiveness') as 'integrated and measurable' dimensions of their traditional activities and duties.

The increasing external influence on the governance of Swedish universities and university colleges has implied a shift in the recruitment of vice-chancellors. Previously this used to be a process among peers, i.e., vice-chancellors were selected through an internal procedure within the universities and the university colleges. With the introduction of external chairpersons, it has become common to search for external candidates, often using search consultants. Vice-chancellors are thus more and more seen as corporate executives running a corporation of learning. However, they are still appointed by Government, and hence are more or less ordinary Government-appointed civil servants. Their term is at the most six years, with the option of extending the appointment for two additional three-year terms at most.[13]

[13] For an evaluation of the new statutes, see Högskoleverket (2000). See also SOU (2001: 101).

So far, this has not led to the introduction of a full-fledged 'spoil-system', where Vice-chancellors and Chairpersons automatically resign when the appointing government loses power. But in the public debate this possibility has, nevertheless, been put forward as a more or less logical end-result if this development continues much longer.

Simultaneously, as government has advanced its positions in the formal governance of universities and university colleges; the regulation of higher education has become less strict. While the early 1970s were characterised by a high degree of centralisation, the last 25 years have been characterised by decentralisation.[14] The central political and bureaucratic authorities are still in command of decisions of a more general nature while widening the scope for local political decision making in higher education. Decentralisation was probably most marked in the area of financial management, where the earlier focus on types of costs was gradually replaced by a system closer to management by objectives. A government bill in 1988 presented general guidelines for a new central government and control system, giving more responsibility to individual authorities and agencies and introducing three-year budget cycles, a system that had been introduced already in 1982 for research policy planning (Premfors 1980). This bill marked the start of extensive development work on the central government budget and control system, which eventually would have substantial implications on the higher education and research funding system.[15]

In the 1990s, the higher education system as well as the research funding system was subject to a series of deep-reaching changes (Benner 2001):

1. The Higher Education Act (1992) and the Higher Education Ordinance (1993) introduced a new system of management by objectives and results for higher education, with Government and Parliament setting objectives and the higher education institutions being assigned the task of meeting the objectives within given parameters.
2. The existing central coordinating and supervising body for the allocation of block grants, the National Board of Universities and Colleges *(UHÄ)* was abolished. Instead, Government now formulates annual appropriation directions for each institution as 'educational assignments'.
3. The institutions gained extended control over their premises. They had to pay rents, however, to a newly founded government corporation owning most of the premises for academic institutions (*Akademiska Hus*). This corporation has specified demands for rates on return on capital from the Ministry of Finance. Needless to say, this principle of quasi-market rents has been debated.
4. Like all other authorities accountable to central government, higher education institutions are required to prepare an annual report, interim report and budget documentation for submission to the Government. In

[14] On the centralized period, see e.g. Premfors (1980), Lane and Fredriksson (1983), and Lindensjö (1981).

[15] As an immediate answer to the rapid expansion of the higher education system in the 1960s and 1970s, a dual career-pattern with a 'teaching only' *(lectors)* path and a research-oriented path was introduced which further increased the internal strains of the university departments (Nybom, 1997).

addition, higher education institutions are also obliged to deliver four periodic forecasts of the estimated outcome for the appropriation for higher education (SOU 1996: 21).

5. The appropriations system in force until 1993 in which resource allocations were detailed and decided by actual costs, was, by and large, replaced by a new system of resource allocation inspired to some extent by the 'buying and selling system' that had already begun to permeate other parts of the public sector. The state grants for institutions depend on the results they achieve – by 'results' were meant the number of full-time equivalent (FTE) students and the number of FTE course completions per year. In research funding, a simultaneous and gradual shift from 'peer review' towards what John Ziman (1994: 94) has labelled "merit review" can also be clearly observed.

Conclusions

In terms of the institutional control of the system for research in Sweden, we can thus conclude that government plays a vital role in controlling entry. However, political pressures to distribute institutions of higher education and research across different regions of the country have entailed a considerable expansion of the number of universities and university colleges. This in turn has led to an increasing decentralisation of decision-making. Today, government is exercising power more through rules and economic incentives than by formal decisions. At the same time, government is also striving for more influence in the governance of universities and university colleges through the appointment of chairpersons and the majority of board members in these institutions.[16]

INPUT CONTROL

Screening of Candidates

With regard to the screening of candidates up to the 1960s, evaluations in the Swedish research sector, as in practically all European university and research systems, were restricted to the initial control of incoming students and tenured research personnel. First, the quality of the incoming students was guaranteed through a rigorous national secondary school examination (*studentexamen*). Second, the quality of the upcoming generation of scientists and scholars was guaranteed through the *Habilitation* procedure followed by a public defence and examination procedure (*disputation*) often eight to ten hours long. Third, all academic positions – and especially the few chair holders (*lärostolsprofessorer*) – were appointed by the King in Council after a rigorous and lengthy process of reviewing and appeal, which

[16] For a discussion see Strömholm (2006).

in itself was supposed to be a sufficient guarantee for the preservation of academic excellence.[17]

However, over time, these three screenings of candidates have undergone changes. As the number of students passing secondary school examinations increased, academic institutions had to introduce *numerus clausus*, i.e., students had to be selected through quality standards. These have been generally based on school leaving certificates but have been supplemented with national tests (*Högskoleprovet*) and other criteria (*Högskoleförordningen* 2000, Ch. 7).

Similarly, in 1969 the traditional German system of dissertation *cum habilitation*, for different reasons, was changed to a system similar to the North American Ph.D., with course requirements and a total study time of four years. External examiners still scrutinised the dissertations, but the grading was abolished and initially there were no examination committees. However, after a few incidents involving dissertations that were considered to have been of disputable quality, examination committees were reintroduced.[18]

In addition, the recruitment of professors changed in the last decade of the twentieth century. The Higher Education Act of 1992 stated that institutions could create new professorial positions. The salaries also became individually negotiable. All this had hitherto been the prerogative of Government to decide. Nevertheless, the Higher Education Ordinance continued to state what categories of teachers the institutions could employ and also to decide the provision concerning the proper recruitment procedures for certain categories of teachers, not least professors.

Furthermore, in 1999, the special security of employment (Royal Charter) for the professoriate under the Higher Education Act was abolished. At the same time, it also became possible for persons who had an academic competence corresponding to the professorial level to be promoted to professor (*Högskoleförordningen* 2000, Ch. 4, 11). In the evaluation of these candidates, the established professors demanded considerably high – some would say excessive – standards of quality.

Finally, it should be mentioned that publications in reviewed journals and citations have become an increasingly common feature in the screening of candidates in Sweden. This has created particular tensions between representatives of disciplines with different publication traditions, such as whether they published mainly in journals or in monographs. These differences have particular importance, since candidates for academic appointments are screened, after expert evaluations, by representatives of the relevant faculty of the university and not by their particular department. These differences also have implications for arguments related to the allocation of resources (Regeringens Proposition, 2004/05:80).

Allocation Procedures

The Swedish system for higher education and research was governed through centralised resource allocation for a long time. Universities were given block grants

[17] Not infrequently, these processes constituted the basis for lifetime animosity among candidates, as well as between candidates and external evaluators.

[18] For a discussion of the system, see e.g. Engwall (1987) and Nybom (1997).

with detailed specifications. However, in the 1940s the government decided to support research not only through block grants but also through grants for specific projects. For this purpose, research councils were created for different areas of research. Their task was to allocate research money after applications from project leaders. These grants were given as a supplement to the block grants.[19] Earlier, private foundations had also provided such grants; the Knut and Alice Wallenberg Foundation being the most significant.[20]

An important change occurred in the Swedish research funding system in 1977. In line with the proposals presented by a government committee (*Forskningsråds-utredningen) (*SOU 1975:26) the research council organisation, originally established directly after the Second World War, was restructured. This process did not just include a merger of some of the existing rather small and discipline-oriented research councils; it also meant the establishment of a new research-supporting agency primarily devoted to the support of interdisciplinary research – the Council for Planning and Co-ordination of Research *(Forskningsrådsnämnden*, *FRN)*. This implied that academic and disciplinary excellence were no longer the only criteria and prerequisites for the allocation of grants from the research council organisation: an additional criterion of *societal relevance* was introduced (Premfors 1980; Landberg et al 1995).

More important for the research funding system was the build-up of a sectorial - i.e., applications-oriented - research apparatus on a fairly massive scale. This development in the 1970s implied that most ministries and government agencies set up some kind of research granting body of their own (Stevrin 1978). Given the Swedish research system, these new and sometimes quite substantial resources were channelled into the traditional university system, often as comparatively short-term assignments. Often they were not distributed after a process of traditional peer review but rather were an outcome of an internal politico-bureaucratic process where academic and disciplinary competence played a secondary role to political urgency (Elzinga 1985; Gustavsson 1989).

This second stream of public research funding was not confined to the technical and medical fields of research but became of great importance for other fields and sub-fields in the social sciences. This in turn led to at least a partial disintegration of the existing research organisations in the universities. Gradually there was a growing criticism, not least in academic circles, against this type of research funding, due to its alleged lack of quality, and the Social Democratic government made an attempt in the late 1980s to restructure the sectorial bodies by turning them into more research council-like entities where scientific expertise was supposed to have a final say.

In view of the expansion of the system and international pressures for market solutions, a government in the early 1990s, breaking a long-standing run of Social Democratic leadership in Sweden, made several changes in the Swedish system for higher education and research. It followed, with a slight delay, the international trend towards what has sometimes been labelled ‘academic capitalism’ (Bok 2003;

[19] See also Nybom (1997: 65-104).
[20] For an account of the Knut and Alice Wallenberg Foundation, see Hoppe et al. (1993).

Slaughter and Leslie 1997; Slaughter and Rhoades 2004). As previously mentioned, the changes implied that resources for higher education now are granted on the basis of a combination of input and output indicators.[21] Similarly, block grants for research, including research training and doctoral studies, are allocated under special appropriations to each university and those university colleges that have been approved for examination of doctoral students in a particular area of research. Furthermore, government specifies objectives for doctoral degrees in each of the four 'areas of research', a newly invented bureaucratic or cameral term replacing the traditional faculty designation.

In addition, specific guidelines and directions for the use of some of the allocated resources for research training were occasionally issued. One of the most recent examples was the establishment of 16 national 'graduate schools' in 2000, where research orientation, forms of organisation, and collaboration were clearly stated.[22] Two years earlier, the legal and financial prerequisites under which the universities were allowed to recruit doctoral students were specified and made mandatory, with four years guaranteed funding being stipulated, amongst other things (SFS 1998: 80).

Until the late 1980s, competitive funding constituted an important but still relatively minor part (10-20%) of the total research budget, with the exception, of course, of fields like technical and (parts of) medical research. As pointed out earlier, external funds were considered to be top up money in an organisation where normal research was supposed to be financed through regular block grants paid directly to the universities. In the 1990s, this funding system was more or less reversed quite suddenly.

The process started in 1994, when the Liberal-Conservative government decided to transfer a substantial part of the so-called Wage Earners Funds to establish a handful of autonomous research foundations for funding research in a number of what were considered to be strategic areas of knowledge and innovation (Regeringens Proposition 1991/92:92). Some of these 'strategically' defined foundations were either given explicit policy goals (The Foundation for Strategic Environmental Research – MISTRA) or statutes where co-operation and joint ventures with industry and business sectors became mandatory (The Knowledge Foundation – KK-Stiftelsen). The non-academic presence on the boards and in review groups was quite substantial. But even if the government-appointed boards consisted of a mixture of lay (industry and society) and academic members, the money would, nevertheless, be allocated after a due process of traditional peer review. There are, however, notable differences between the academically controlled

[21] Higher education in Sweden, except doctoral studies, is nearly exclusively financed by state block grants. The only exception is education commissioned by employers. Individual tuition fees are not allowed in Sweden. In 2004, the Government appointed a commissioner to investigate the possibilities to charge fees from non-European students (see SOU 2006:6). In relation to this it is worth noting that, since the mid-1990s, the revenues per FTE student and FTE course completions were reduced. According to the annual report for 2002 by the National Agency resources for undergraduate education had fallen by 17% in real terms between 1994 and 2001. This trend has continued ever since.

[22] See Regeringens proposition (2000/01: 3) and Högskoleverket (2004). For a presentation of one of them see http://www.forskarskolan-mit.nu.

research councils and the foundations in the way they interact with their respective grant-recipients. Whereas the research councils still adhere to the principle of implicit trust in their founding, some foundations have introduced a more hands-on policy of continuous evaluations and deliverables (Sörlin 2005).

Initially, these foundations were supposed to function as an additional boost for the regular research organisation, but they soon turned into an integral and decisive part of the existing research system. And even more seminally, they became an excuse for substantial cutbacks in public funding in the latter half of the 1990s, particularly in basic research in the university sector. In 2002, the new research foundations were responsible for MSEK 1,145 or 5% of all research funding.[23]

The end result was a relatively dramatic undermining of the research infrastructure in Swedish universities. In the beginning of the 1980s, external competitive financing accounted, *on average*, for about 33% of the total research budget of the universities, while the corresponding figure in 2002 was 55%. Thus, in less than two decades there was a rather dramatic shift in the relation between block grants and competitive funding. At the same time, the level of overhead paid by the external funding bodies rose from 3% at the beginning of the 1980s to over 50% on average in 2004. The latter also led to an almost perpetual and counter-productive controversy over the just and proper level of overhead for universities, on the one hand, and funding bodies and researchers, on the other (Nybom and Stenlund 2004; Engwall 2005).

The changes we have described were reflections of a principal policy statement by the Social Democratic government, that an increasing part of the state research funds available to the universities should in future be transferred only after a due process of competitive funding, controlled by the state research council organisation. Government considered the model as a chief means to guarantee quality and to 'sharpen the profile of Swedish research'. The sharp reduction in research funding provided directly to the universities by the State has in reality lead to a marked deterioration in their capacity to set priorities internally and to give their research activities a clearly defined unique profile. Instead, the profile-building ambitions seem to lead to a growing tendency towards mainstreaming among universities, research founders, and research units.

The expansion of the higher education sector, with a deliberate political commitment to university colleges, and the introduction of an initiative to give new institutions the right to award doctoral degrees, have led to a further shift in the distribution of research funds, to the disadvantage of the traditional research universities.[24] Of the estimated growth of MSEK 2,900 over the period 1998-2000 for research and doctoral studies, more than 30%, or MSEK 1,000, went to the three newly established universities and to university colleges. The problems of underfunding became even more pronounced by a simultaneous rapid expansion of undergraduate and vocational education.

---

[23] For an examination of the first ten years of these foundations, see Sörlin (2005). For The Bank of Sweden Tercentenary Foundation, see Nybom and Stenlund (2004).

[24] A few of the university colleges are also entitled to receive limited block research grants (and award doctorates) in designated 'research areas'.

To promote its new policy of competitive funding and to strengthen the efficiency of public research financing, in 2001 the Government decided to amalgamate a number of state basic research councils that had hitherto been organised according to disciplinary principles, and established a single body, the Science Research Council (*Vetenskapsrådet, VR*). Simultaneously, the more applied research funding agencies were amalgamated into the national bodies *VINNOVA* (technology), *FORMAS* (environment) and *FAS* (social policy).

To further stress the element of strategic steering and earmarking, the government bill of 2000 (*Research and Renewal*) also identified eight different areas of strategic importance which government considered the Science Research Council and other public providers of research funding should pay particular attention to. In addition, government also stressed the need for interdisciplinary and multidisciplinary research.

In the last few years, there has also been a tendency for the funding over and above block grants to be given to large programmes, to *strong research environments* or *centres of excellence* rather than to small projects. Increasingly, the opinions of international peers are sought for evaluations. This generally accepted strategy of research funding agencies quite consciously indicates that the agencies are moving from the traditional form of individual research funding towards for a more institutionally based funding strategy.

Conclusions

With respect to input control, we note a certain decentralisation. The screening of students, research education, and the selection of professors have undergone changes implying that, to a certain extent, institutions have a higher degree of freedom. At the same time, there have been increased demands on institutions regarding their treatment of students (particularly doctoral students), while the conditions of professors have deteriorated, particularly through the abolition of the special rules of employment security and guaranteed sabbaticals. This change has been one effect of the increasing emphasis on the allocation of resources through competition between short-term project proposals rather than through long-term block grants.[25] However, in the last few years both Government and the research funding bodies have tended to favour larger programmes over small individual projects. As a result, we can see a trend back towards block grants, although this time not coming directly from Government but instead channelled through research funding bodies. A significant difference between this process and the earlier one is that allocations are based on peer reviews rather than political decisions.

[25] As pointed out by Richard Whitley in his contribution to this volume, similar tendencies can be observed in many other countries. Role models appear to come from the United States, as is often the case (see Susan Cozzen's contribution to this volume).

OUTPUT CONTROL

As research councils developed and became more and more central to the allocation of research resources, these organisations found it appropriate to systematically evaluate the research they had funded. This was the start of a process of evaluation throughout the whole Swedish research system. The development of publication databases and techniques for citation analysis provided significant tools to facilitate such analyses. Already in the late 1960s and early 1970s, the Natural Science Research Council had taken the initiative to carry through regular evaluations of different broad disciplinary fields within the natural sciences, e.g., physics (*Regeringens proposition,* 1981/82: 37; NFR 1981). The evaluations, which normally were carried out by teams of non-Swedish peers, had two main objectives: (a) to do a general stocktaking of that particular disciplinary area, and (b) to review the council's own policy of research funding. The evaluations usually included a number of very general research policy recommendations. In the 1980s and 1990s, the other state research councils (Humanities and Social Sciences) followed suit and initiated similar discipline-based evaluations (history, economics, sociology, and education).[26]

These internally initiated evaluations had no links whatsoever to funding, and they had very little, if any, effect on general policy or the actual allocation of resources.[27] Their recommendations were by no means mandatory and were very rarely implemented. From the late 1990s and onwards one can, however, detect an element of open ranking of the existing research institutions. This shift could very well be seen as a direct effect of Government's directives to the Research Council to increase the competitive component and quality considerations in their funding policy.

The impact and formalisation of evaluation schemes in the 1990s entailed a fairly fundamental change in policy, both in higher education in general and in the research community. The deregulation process of the early 1990s included, almost automatically, the implementation of some kind of output control system, i.e., formal evaluation. The Liberal-Conservative government took the initial initiative with the establishment of the Secretariat of Evaluation. Apart from being the chief accreditation office of academic programmes, the main objective of the Secretariat, as defined by then-University Chancellor, Stig Hagström, was not the 'evaluation of results' but rather the 'evaluation of evaluation procedures' implemented by the higher education institutions. The Secretariat explicitly stated that these evaluations should not be in any way connected with funding.[28]

The Social Democratic government transformed the Secretariat of Evaluation in 1995 into a new National Agency for Higher Education (*Högskoleverket*). The Agency was given the task to conduct quality audits of higher education, to exercise supervision of the higher education institutions, to

[26] For an example of a research council evaluation, see Engwall (1992). Marton (2005: 167-169) provides some general conclusions from the evaluations.

[27] At least one effect of the evaluations can be reported. After the evaluation of economics reported in Engwall (1992), economics professors got together and decided on more realistic requirements for dissertations within their discipline.

[28] For an account of the system, see Engwall (1997).

review and analyse the higher education system, to evaluate qualifications awarded abroad, to support educational innovation and development, to provide information on higher education, and to encourage student recruitment. In 1999, the government instructed the National Agency to carry out an examination of the quality of all disciplines and programmes in higher education over a six-year period beginning in 2001. These examinations involve both a self-evaluation by the departments concerned and assessment by external reviewers. The team of reviewers submit a report to the National Agency including a statement of proposed measures.

The subject and programme evaluations have three main purposes: (a) to contribute to quality development, (b) to examine whether the education offered matches the objectives and provisions set out in the Higher Education Act and Higher Education Ordinance, and (c) to provide information to interested parties. Nine evaluation teams carried out a total of 222 examinations in 2001, of which only eight led to the questioning of the right to award degrees. However, after the report is presented, the department or programme concerned is given the opportunity to correct the alleged deficiencies. Only after a second examination may a withdrawal of the right to award degree be considered. The evaluations performed by the National Agency, which also considers research education and research, have no direct links to funding, and the recommendations are not mandatory.[29]

In addition to the evaluations performed by the Research Council and the Agency of Higher Education, two of the Royal Academies also conduct similar exercises. They were given this task when the new autonomous research foundations with resources from the Wage Earners' Funds were established (see the section on allocation procedures, above). The Royal Swedish Academy of Sciences *(KVA)* and The Royal Swedish Academy of Engineering Sciences *(IVA)* were given statutory rights to review and evaluate the policies and procedures of all the newly created foundations, with the exception of The Bank of Sweden Tercentenary Foundation *(RJ)*. These evaluations have been carried out on a regular basis and have included not only formal procedures for evaluation of the foundations but also an investigation of the quality of the research projects granted.[30]

There is also an increasing tendency for individual institutions in Sweden to perform their own evaluations. One example is the BASTU project at Uppsala University (SAUNA IV 2005). It required the different sections within the university to summarize their research accomplishments and to set priorities for the future. These reports then formed the basis for recommendations from an evaluation team with distinguished members from Berkeley, Edinburgh, and Helsinki. They led to some reallocation of resources. A long-term effect was the creation of a university database of publications (OPUS), which is used for analyses within the university. So far, no direct links between publication patterns and resource allocation have been established.

Finally, it should be mentioned that certain areas of research, like business and engineering, are subject to international evaluations through accreditation and

[29] For a Nordic comparison of evaluation systems, see Kim (2002).
[30] For an example, see KVA (2001).

ranking. Although the focus is primarily on education, research is also taken into consideration.[31]

The foregoing account implies that we can conclude output control is still relatively soft in Sweden.[32] Evaluations of disciplines are undertaken by the Research Council and by the National Agency of Higher Education. In addition, the Royal Academies have evaluated research funded by foundations created out of the Wage Earners' Funds. None of these evaluations are directly linked to funding but are, of course, significant input for decisions on resource allocation. Throughout the system, bibliometric methods are increasingly, if not systematically, used to discriminate between scholars (on bibliometrics, see Gläser and Laudel in this volume). The effects and consequences of the gradual importance of bibliometric criteria – not least 'international visibility' – are not only observable in external (research councils etc.) funding. It has become even more prominent and significant for the allocation of research resources within universities and not least in tenure and employment procedures.

CONCLUSIONS

Our point of departure for this paper was that the allocation of research resources belongs to a larger set of allocation decisions, with the common purpose of channelling available resources to the most efficient use. However, we have also pointed out that these decisions are particularly difficult for research, since uncertainty is high and the time horizon is often very long. Under these circumstances, we have argued that governments tend to control the field through *institutional control*, *input control,* and *output control*. Applying this framework to the Swedish research system we have found the following:

1. Government plays a significant role in controlling the entry of new actors. However, gradually, the number of universities and university colleges has grown from a limited number in the post-war decades to 40 institutions of higher education today. Pressures to allocate research funds to all these units are strong but have mainly been resisted.
2. Government, both an earlier Liberal-Conservative and the present Social-Democratic, has introduced New Public Management into the system of universities and university colleges. Institutions of higher education and research have been given specified tasks and are rewarded when achieving these tasks. The Social-Democratic government has also pushed universities and university colleges towards more corporate-style governance, with a majority of external members on the board and one

[31] For analyses of accreditation and ranking, see Hedmo (2004) and Wedlin (2006), respectively. As pointed out by Peter Weingart and Sabine Maasen in their contribution to this volume, various kinds of such rankings have also been developed for whole universities, the most spectacular being the Shanghai Ranking of World Universities (http://ed.sjtu.edu.cn/ranking.htm). One such ranking of universities was made in Sweden in 1999 by the magazine *Moderna tider*.

[32] Evidence of tougher systems in other countries is provided in the contributions to the present volume by Jochen Gläser and Grit Laudel for Australia and Barend van der Meulen for The Netherlands. See also again the general discussion in the contribution of Richard Whitley.

external chairperson. This person has increasingly tended to be selected from among politicians. In terms of institutional control, we have thus observed a combination of market and political forces.[33]

3. We have also been able to show that government has left it to the academic community to appoint professors. Government thus no longer formally controls the input into the system of the most qualified researchers. In this way, academic institutions have obtained a certain degree of freedom in selecting professors. At the same time, professors have become ordinary state employees without their earlier privileges of employment security.
4. Another feature of the changing Swedish system has been that resources for research to an increasing extent are allocated through short-term project grants rather than long-term block grants. The logic behind this change has been that project grants would be more flexible and would imply that researchers were under continuous pressure to perform high-quality research. However, we have also noted a tendency among research councils and research foundations to support fewer and larger programmes rather than many small projects. Thus, block grants can in a way be said to have returned through the back door. But it could also be said that the politicians have turned over a genuine political responsibility and decision-making role to the research community.
5. There has been an increasing tendency for output control, i.e., to evaluate research. In Sweden, this trend started with the research councils in the 1960s and has since been continued both by these organisations and by other bodies like the National Agency for Higher Education and the Royal Academies. The link to funding has so far been weak.

The overall conclusion of our paper is that decentralisation has characterised the system for research and higher education in the last decades. The earlier 'visible hand' of closed governance, through the State budget and through the appointment of professors, has thus, at least rhetorically, ceded power to the 'invisible hand' of market governance, both in terms of higher education and research. Whether this has led to a more efficient system is difficult to say. Studies of citations, however, appear to indicate rather the opposite, i.e. that the role of Swedish researchers on the international scene has deteriorated.[34] Needless to say, there are many explanations of these findings offered by various actors. The suggestions for solutions to change this state of affairs for the better are also numerous. They can all be seen as elements in the everlasting struggle over research resources between disciplines, between institutions and between generations. Fortunately, it is not the task of the present paper to solve these conflicts, only to point to features of the system as it has developed into its current state. The only forecast we dare to make about the future is that the conflicts will remain.

---

[33] A non-Socialist government taking office in September 2006 gave universities and colleges the right to propose their board members to the government (Government Bill 2006/07:43).

[34] Heyman (2003) reports a distinct decline in citations of Swedish research, particularly within medicine. Similar results are also reported in a Norwegian study (Aksnes and Sivertsen 2003, see also http://www2.db.dk/jws/Files/Workshop%20Presentations/Askness_Sivertsen.pdf).

Finally, it is of course appropriate to briefly discuss the extent to which the Swedish findings have a broader relevance, i.e. whether they can be observed or can be expected to be found in the future in other countries. Without going into extensive comparisons between countries there seem to be reasons to believe so. First, research and higher education have become increasingly significant in the political debate in most countries. Scientific innovations and academic competence have thus since long been mentioned as means for future prosperity and competitive advantage (e.g. Porter 1990). Second, there has been a rapid diffusion of managerial models from the corporate world to public sector. A particular significant feature in this process has been the movement towards an audit society (Power 1999), i.e. a shift towards an increasing output control.[35] Needless to say this is part of an extensive diffusion of management models in the modern society (Sahlin-Andersson and Engwall 2002).

REFERENCES

Aksnes, Dag W. and Gunnar Sivertsen (2003), *How Should One Measure the Overall Citation Rates of Countries?*, Discussion Paper. Oslo: Norwegian Institute for Studies in Research and Higher Education.

Benner, Mats (2001), *Kontrovers och konsensus. Vetenskap och politik i svenskt 1990-tal*, Nora: Nya Doxa.

Bok, Derek (2003), *Universities in the Marketplace: The Commercialisation of Higher Education*, Princeton: Princeton University Press.

Crainer, Stuart and Des Dearlove (1999), *Gravy Training: Inside the Business of Business Schools*, San Francisco: Jossey-Bass.

Elzinga, Aant (1985), 'Research, Bureaucracy and the Drift of Epistemic Criteria', in Björn Wittrock and Aant Elzinga (eds.), *The University Research System: The Public Policies and the Home of Scientists*, Stockholm: Almqvist &Wiksell, pp. 191-217.

Elzinga, Aant (1993), 'Universities, Research and the Tranformation of the State', in Sheldon Rothblatt and Björn Wittrock (eds.), *The European and American University since 1800. Historical and Sociological Essays,* Cambridge: Cambridge University Press, pp. 191-233.

Engwall, Lars (1987), 'An American Dream. Postgraduate Research Training in the Social Sciences in Sweden', in *Postgraduate Research Training in the Social Sciences*, Copenhagen: International Federation of Social Science Organizations, pp. 122-128.

Engwall, Lars (ed.) (1992), *Economics in Sweden*, London: Routledge.

Engwall, Lars (1997), 'A Swedish Approach to Quality in Education', in John Brennan, Peter de Vries and Ruth Williams (eds.), *Standards and Quality in Higher Education*, London: Jessica Kingsley, pp. 220-244.

Engwall, Lars (2005), 'Hur har vi det med beredskapen idag? Utbildning och forskning som modern försvarspolitik', in Anders Björnsson, Martin Kylhammar and Åsa Linderborg (eds.), *Ord i rättan tid*, Stockholm: Carlssons, pp. 281-300.

Engwall, Lars (2007) "The Universities, the State and the Market: Changing Patterns of University Governance," *Higher Education and Policy*, 19, 87-104.

Government Bill 2006/07:43. Frihet att välja ett ökat inflytande för universitet och högskolor när styrelseledamötor utses, Stockholm: Fritzes.

Gustavsson, Sverker (1989), 'Den sista forskningspropositionen', in Thorsten Nybom (ed.), *Universitet och samhälle. Om forskningspolitik och vetenskapens samhälleliga roll*, Stockholm: Tiden, pp. 165-178.

Hedmo, Tina (2004), *Rule-making in the Transnational Space: The Development of European Accreditation of Management Education*, Uppsala: PhD Dissertation, Department of Business Studies, Uppsala University.

[35] Similar conclusions are drawn by Peter Weingart and Sabine Maasen in their contribution to this volume.

Heyman, Ulf (2003), *Sveriges vetenskapliga publicering – en analys av NI*, Stockholm: Vetenskapsrådet.
Högskoleförordningen (2000), in *Utbildningsväsendets författningsböcker*, Del 3, Universitet och högskolor, Stockholm: Norstedts, pp. 27-108.
Högskolelagen (2000), in *Utbildningsväsendets författningsböcker*, Del 3, Universitet och högskolor, Stockholm: Norstedts, pp. 13-22.
Högskoleverket (1998), *Högskola i dynamisk utveckling: fyra högskolors förutsättningar att bli universitet*, rapportserie, 11 R, Stockholm: Högskoleverket, http://www.hsv.se/download/18.539a949110f3d5914ec800087718/9811R.pdf (accessed 29 May 2007).
Högskoleverket (2000), *Att leda universitet och högskolor: en uppföljning och analys av styrelsereformen 1998*, rapportserie, 15 R, Stockholm: Högskoleverket, http://www.hsv.se/download/18.539a949110f3d5914ec800083506/0015R.pdf (accessed 29 May 2007).
Högskoleverket (2004), *Uppföljning av 16 nationella forskarskolor – samverkan, rekrytering, handledning och kurser*, rapportserie, 18R, Stockholm: Högskoleverket, http://www.hsv.se/download/18.539a949110f3d5914ec800092647/0418R.pdf, (accessed 29 May 2007).
Hoppe, Gunnar, Gert Nylander and Ulf Olsson (1993), *Till landets gagn: Knut och Alice Wallenbergs stiftelse 1917-1992*, Stockholm: Norstedt.
Kim, Lillemor (2002), *Lika olika. En jämförande studie av högre utbildning och forskning i de nordiska länderna,* Stockholm: Högskoleverket.
KVA [Kungl. Vetenskapsakademien] (2001), *Stiftelsen för forskning inom områden med anknytning till Östersjöregionen och Östeuropa (Östersjöstiftelsen). En granskning av verksamheten 1994-juni 2001*, KVA rapport 2001: 3, Stockholm: KVA, http://www.kva.se/KVA_Root/publications/reports/reportbaltic.pdf (accessed 29 May 2007).
Landberg, Hans, Olle Edquist and Uno Svedin (1995), *Riksdagen, regeringen och forskningen. Några drag i svensk forskningspolitik under två decennier*, Stockholm: Fritzes, SOU 1995:121.
Lane, Jan-Erik and Bert Fredriksson (1983), *Higher Education and Public Administration*, Stockholm: Almqvist & Wiksell.
Lindensjö, Bo (1981), *Högskolereformen*. *En studie i offentlig reformstrategi,* Stockholm Studies in Politics 20, Stockholm: Gotab.
Marton, Susan (2005), 'Humaniora och samhällsvetenskap – 'business as usual'?', in Sverker Sörlin (ed.), *'I den absoluta frontlinjen'. En bok om forskningsstiftelserna, konkurrenskraften och politiken möjligheter*, Nora: Nya Doxa, pp. 159-199.
NFR [Naturvetenskapliga forskningsrådet] (1981), International Evaluations of Research Projects Supported by the Swedish Natural Science Research Council. Summary of Reports, Achievements and Criticisms, 1977-1980, Stockholm: NFR.
Nybom, Thorsten (1997), *Kunskap, politik, samhälle. Essäer om kunskapssyn, universitet och forskningspolitik 1900-2000*, Stockholm: Arete.
Nybom, Thorsten and Bengt Stenlund (eds.) (2004), *'Hinc robur et securitas'. En forskningsstiftelses handel och vandel – Stiftelsen Riksbankens Jubileumsfond 1989-2003,* Hedemora: Gidlunds.
OECD [Organisation for Economic Cooperation and Development] (2005), *OECD Science, Technology and Industry Scoreboard*, Paris. OECD.
Porter, Michael E. (1990), *The Competitive Advantage of Nations*, New York: Free Press.
Power, Michael (1999), *The Audit Society*, Oxford: Oxford University Press.
Premfors, Rune (1980), *The Politics of Higher Education in Comparative Perspective: France, Sweden, United Kingdom*, Stockholm Studies in Politics 15, Stockholm: Gotab.
Regeringens proposition 1981/82:106, *Proposition om forskning*, Stockholm: Fritzes, 1981.
Regeringens proposition 1991/92:92*, Om utskiftning av löntagarfondernas tillgångar m.m.*, Stockholm: Fritzes, 1992.
Regeringens proposition 2000/01:3*, Forskning och förnyelse*, Stockholm: Fritzes, 2001.
Regeringens proposition 2004/05:80, *Forskning för ett bättre liv*, Stockholm: Fritzes, 2005.
SAUNA IV (2001), *Rektors förslag till förnyelse och fördelning av BASTU-medele*, Uppsala: Uppsala universitet.
Sahlin-Andersson, Kerstin and Lars Engwall (eds.) (2002), *The Expansion of Management Knowledge. Carriers, Flows and Sources*, Stanford, CA: Stanford Business Books.
Schilling, Peter (2005), *Research as a Source of Strategic Opportunity? Re-thinking Research Policy Developments in the Late 20th Century*, Umeå Studies in Economic History 32, Umeå: Umeå universitet.

Slaughter, Sheila and Larry L. Leslie (1997), *Academic Capitalism. Politics, Policies, and the Entrepreneurial University*, Baltimore: The Johns Hopkins University Press.

Slaughter, Sheila and Gary Rhoades (2004), *Academic Capitalism and the New Economy*, Baltimore: Johns Hopkins University Press.

Sörlin, S. (ed.) (2005), *'I den absoluta frontlinjen'. En bok om forskningsstiftelserna, konkurrenskraften och politikes möjligheter*, Nora: Nya Doxa.

SOU 1975:26 [Statens Offentliga Utredningar] (1975), Forskningsråd. Betänkande avgivet av forskningsrådsutredningen. Stockholm: SOU. Liber.

SOU 1996:21 [Statens Offentliga Utredningar] (1996), Reform och förändring : organisation och verksamhet vid universitet och högskolor efter 1993 års universitets- och högskolereform, Stockholm: Fritzes.

SOU 2001:101 [Statens Offentliga Utredningar] (2001), Högskolans ledning. Slutbetänkande av Utredningen om högskolans ledning, Stockholm: Fritzes.

SOU 2006:6 [Statens Offentliga Utredningar] (2006), Studieavgifter i högskolan. Betänkande av studieavgiftsutredningen, Stockholm: Fritzes, 2006.

Stevrin, Peter (1978), *Den samhällsstyrda forskningen: en samhällsorganisatorisk studie av den sektoriella forskningspolitikens framväxt och tillämpning i Sverige*, Stockholm: Liber.

Strömholm, Stig (2006), 'Tankar om universitetsstyrning', in Kjell Blückert and Eva Österberg (eds.), *Gränslöst – forskning i Sverige och världen*, Stockholm: Natur och Kultur, pp. 422-441.

Trow, Martin (1998), 'Governance in the University of California: The Transformation of Politics into Administration', *Higher Education Policy*, 11, 201-215.

Wedlin, Linda (2006), *Ranking Business Schools: Forming Fields, Identities and Boundaries in International Management Education*, Cheltenham: Edward Elgar.

Ziman, John (1994), *Prometheus Bound: Science in a Dynamic Steady State,* Cambridge: Cambridge University Press.

CHAPTER 3

ROBERT KNELLER

PROSPECTIVE AND RETROSPECTIVE EVALUATION SYSTEMS IN CONTEXT: INSIGHTS FROM JAPAN

INTRODUCTION

Research evaluation in Japanese universities has relied upon prospective evaluation of competing research proposals until very recently. Over the past decade, the process for evaluating and selecting among competing research proposals, which I henceforth refer to as *prospective peer review*, has evolved so that in 2006 it is more transparent and based upon expert input than it was in the mid 1990s. However, because each program usually has its own peer review process and the number of government programs that might provide competitive funding is quite large, potential applicants face a variety of peer review systems.

Despite the improvements in prospective peer review, two troubling questions remain: First, does the Japanese system of research funding and prospective peer review still encourage young researchers to follow the leads of senior professors, and thus discourage innovative, pioneering research? Second, the distribution of competitive research and development (R&D) funding (which accounts for roughly half of total R&D support in universities) remains highly skewed in favour of a few elite universities, as shown in the Appendix. Does this reflect an equally skewed distribution of talent or a tendency for the peer review system to allocate funding on the basis of institutional status and reputation?

The inequality in funding is a central issue in the recent effort to implement a *retrospective* evaluation system. This system is aimed at providing objective measures of individual researchers' performance that might be used in promotion decisions and that would encourage individual researchers to be more productive. But another purpose is to differentiate between a group of approximately thirty *research oriented* universities that will continue to receive substantial government funding in the hope that they will become world class research centres, and a larger group of *education oriented* universities where research will be viewed as a subsidiary activity and will be funded accordingly. Such differentiation is a common

R. Whitley and J. Gläser (eds.), The Changing Governance of the Sciences, 51–73.

aim of national research evaluation systems, most explicitly perhaps in Australia (see Gläser and Laudel, this volume) and the UK.

Japan is implementing its retrospective system slowly, and it is not expected to result in differentiated budget reductions until 2010. However, already second and third tier universities are complaining that they should not be judged by the same criteria as elite universities. A key question about this new retrospective system, then, is: Are its results pre-ordained by the skewed results of the prospective system and by equivalent discrepancies in the distribution of general purpose funding? Known as *unei koufu kin* (literal translation: *operational and administrative subsidies*), this Japanese equivalent of block grant funding is also skewed in favour of the same elite universities, particularly, as shown in the Appendix table 3, the seven state universities designated as *imperial universities* in the pre-war era. Only a portion of the general purpose funding supports research, but the table nevertheless shows the overall disparity in distribution. If the results of retrospective evaluation are indeed pre-ordained, what additional purpose does it serve?

In attempting to deal with these questions, it is crucial to understand how evaluation systems and processes function in their broader institutional context, especially the funding and control of academic systems and reward systems, as the contributions to this volume make clear (see also Whitley 2003). Accordingly, in this paper, I summarise the key features of the post-war Japanese research system and the role of prospective evaluation in it before considering the likely impact of retrospective evaluation on the direction and organisation of research in the future. Thus, the following section describes the Japanese university research system and the role of prospective peer review, including a detailed description of the prospective evaluation process and the principal funding sources. Next, I outline the retrospective review system that is in its first years of trial implementation, while the final section offers some tentative remarks about coordinated reforms that might improve the climate for innovative science in Japan and other countries.

UNIVERSITY RESEARCH AND PROSPECTIVE PEER REVIEW

The University Research System

In 2004, the Japanese system of higher education consisted of 87 national universities and 4 national academic research institutes under the Ministry of Education, Sports, Culture, Science and Technology (MEXT); 80 prefectural, municipal and local government universities; and 542 private universities. In total, the universities accounted for about 14% of Japanese R&D in 2002, with government research institutes (GRIs) contributing a further 9.5% (NSB 2006). National universities conducted about 75% of this research and graduated 78% of all science and engineering doctorates in 2004 (MEXT 2004). A significant proportion (55%) of the research activities carried out in the national universities is supported by 'outside sources.' This percentage is around 50 in the private universities that fund a larger proportion of research from tuition fees. The most important of these

external sources is government competitive research funding but industry funding and overhead support are also important components.

Japanese national universities receive almost no funding from prefectural or local governments, except for some sponsored research, despite many being major contributors to regional economies. As the premier educational institution in many of the outlying prefectures, they have the potential to contribute substantially to local economic development. In the case of the wealthier prefectures, bureaucratic rivalry between the prefectural governments and the prefectural universities they support, on the one hand, and MEXT, on the other hand, have prevented cooperation on a regional level. In the case of poorer prefectures, local governments lack the means to support resident national universities.

The Japanese academic system is highly stratified with the older state universities, particularly the universities of Kyoto and Tokyo, being the most prestigious and controlling the most resources (see Appendix). It is trite but nevertheless true that most academically inclined high school students (or at least their parents) dream of entering the University of Tokyo or Kyoto University, and most academics dream of ending their careers there. Because of these strong regional and institutional preferences, the system of recruitment and promotion in a few highly regarded universities influences academic career strategies throughout the nation. The elite universities do try hard to recruit and promote able and productive researchers, but nevertheless the selection of lead candidates usually depends upon small internal committees in which a single professor often has a dominant voice. The committees' selection of a lead candidate is rarely questioned by the larger faculty and university. Open debate is also unusual and solicitation of outside opinions even more so.

The leading universities manage to recruit creative and capable persons using this system, because they attract interest from bright younger researchers throughout the country, but the need for patronage probably discourages young researchers from pursuing unorthodox themes or using unorthodox research approaches (Coleman 1999). The price of failure is not simply losing an opportunity to work in a prestigious university. Because funding is consistently skewed in favour of the same small number of universities, it may mean spending one's career in a university with scant research resources.[1]

---

[1] It might be argued that, because the Tokyo, Nagoya, and Osaka-Kyoto-Kobe metropolitan regions (hereinafter, group A regions) account for a large proportion of Japan's population, it is appropriate for universities in these regions, as well as the four next most populace regions: Sapporo, Sendai, Hiroshima and Fukuoka (hereinafter, group B regions); to receive a large proportion of research support. However, assigning the universities that account for 95 percent of 2006 MEXT Grants-in-aid to standard-classification metropolitan regions shows that universities in group A regions receive 63 percent of grants-in-aid, although these regions account for only 43 percent of the 2006 population. Including group A and B regions together, universities in these regions receive 83 percent of grants-in-aid, although they account for only 49 percent of population.

Even classifying universities and population according to *prefecture* rather than *metropolitan area* shows skewed funding in relation to population. Thus, if the Tokyo area is considered to include all of Chiba, Saitama and Kanagawa prefectures, and the Osaka-Kyoto-Kobe area to include all of Shiga, Nara and Hyogo prefectures (most of these prefectures include substantial rural areas distant from the

In 2004, national universities became independent administrative organizations (officially university corporations) under MEXT. Previously they were simply branches of MEXT, and personnel and financial matters were all subject to MEXT control. As a result of incorporation, they now have legal authority over most key areas. They are free, in theory, to determine the number of permanent faculty positions and their allocation between existing or new departments, as well as being able, in principle, to raise funds from alumni, local governments and other sources, and to vary tuition charges, although not by more than 10% a year. In fact, however, they all remain dependent upon MEXT for infrastructure costs and salaries of permanent staff. These expenses are covered mainly by the *operating and administrative subsidies* mentioned above.[2]

In order to reduce government expenditures while encouraging more self-reliance, beginning in fiscal year (FY) 2005, subsidies for all national universities are being reduced by about 1 percent per year compared to the 2004 base. Beginning in FY 2010 these reductions are to increase to 2 percent annually for universities that are judged to be *education oriented*, while the rate of reduction will remain at 1 percent for approximately 30 universities judged to be *research and education oriented*. The annual data being collected on research performance are intended to play a key role in determining which universities fall into each category.

So far, the subsidy reductions have not lead to decreases in numbers of permanent faculty. Instead, cuts are being absorbed by shedding temporarily unfilled faculty positions that are still receiving subsidies. But as the cuts continue, reductions in faculty and other permanent staff are expected, especially in the education oriented universities. Also some of the weaker education oriented universities are expected to merge or be absorbed by stronger universities. In view of the large number of universities, many with meagre research resources, some consolidation is appropriate. Nevertheless, to suggest that world class research ought to be consolidated within approximately 30 institutions seems to assume that elitism, not broadly based competition, leads to scientific excellence. It also presumes that it is possible to teach higher level subjects effectively without doing research or having contact with research colleagues, and that the solution to Japan's low share of Nobel Prize level scientific research is even greater concentration of research resources.

The Organisation of University Research and the Academic Reward Systems

The basic organizational unit in Japanese universities is the *kouza*, modelled on the *professor chair* system in early 20th century German universities. A *kouza* typically consists of one full professor, the laboratory head, an assistant/associate professor

main cities), the 63 percent of grants-in-aids received by group A region universities would still be attributed to only 49 percent of the national population. Also the 83 percent of grants-in-aid received by the combined group A and B regions would be attributed to only 58 percent of population.

[2] Tuition and patient hospital charges are combined with the subsidies to form the main pool of funds out of which salaries and general infrastructure costs are paid.

(*jo kyouju*), who is usually the lead candidate to inherit laboratory leadership when the professor retires, and one assistant (*joshu*). There is usually one laboratory per *kouza*. Thus laboratory facilities are under the *kouza* head. Applications for research funding from junior *kouza* members usually include the *kouza* head as a co-applicant, and of course must be coordinated with him/her.

Some *kouzas* contained an instructor (*koushi*) intermediate in rank between a *joshu* and the assistant/associate professor. A *koushi* was expected to emphasize mainly teaching and sometimes was not considered to be in line to fill a vacancy at the assistant/associate professor level. In 2007, these titles changed. Assistant professors became associate professors (*jun kyouju*). Assistants became assistant/associate professors (*jokyou*).

Academic careers still depend upon patronage rather than a record of individual achievement (Coleman 1999). Well into the 1990s, it was still common for vacancies to be filled from within the *kouza*. The *kouza* represented a narrow career ladder where vacancies were usually filled by the person next below in the hierarchy, and the professor essentially picked his second generation successor when he selected a new joshu. Now internal promotions to the assistant professor level are discouraged, and joshus/jokyous usually find their first assistant/associate professorship in a different *kouza*, sometimes in a different university. Nevertheless, academic recruitment and promotions still depend mainly upon the recommendations of key senior professors. Open recruitment, in the sense of widely soliciting applications to fill vacancies and a commitment to select among applicants on the basis of merit, is still rare.[3] Rarer still is soliciting in-depth, objective evaluations of candidates' achievements from *outside* experts and giving considerable weight to these outside evaluations.[4]

The *kouza* system is, however, becoming more flexible. In a few departments, formal *kouza* affiliations have been abandoned and professors make real collective recruitment decisions based upon individual merit and the needs of the department, not upon applicants' past affiliations with members of the department or the closeness of their research interests to those of particular senior professors. Even in such departments, however, there is usually no objective outside input into the process.

Research Funding and Prospective Peer Review

Various Japanese government ministries have implemented a variety of programmes for funding research. Procedures for judging applications and awarding funds vary,

[3] I am familiar with recruitment and promotion practices in only a few Japanese universities, but these include two of the leading national universities and one leading private university. Within each of these three universities I know of one department that practices this form of open recruitment. But persons within these departments themselves say that they are pioneers within their universities. In other words, they are exceptions that prove the rule.

[4] Such steps are under consideration in a few departments, but I know of no department that has implemented such procedures.

reflecting the different missions of these ministries, their various *modus operandi*, and in some cases their rivalry. It is worthwhile summarizing some of the more important ones in order to contrast them with the procedures for retrospective research evaluation. The following summaries cover MEXT Basic Research Grants-in-aid, the Strategically Promoted Creative Research Programme of the Japan Science and Technology Corporation (JST), MEXT Special Coordination Funds for Promoting Science and Technology, and the Centres of Excellence Programme.

The *MEXT Basic Research Grants-in-Aid* are the mainstay of support for R&D projects in Japanese universities. Although the applicant can name collaborators, one person should have main responsibility for carrying out the project. The applicant chooses among 276 subject categories covering most fields of humanities, social science, law, natural science, engineering, agriculture and medicine. Each category has a review committee consisting of six to fifteen persons, most of whom are university professors, who maintain their normal responsibilities. For example, there are 13 categories under 'chemistry,' one of which covers 'physical chemistry.'[5] The applicant does not know the names of the panel members who will review the current year's applications, but the names and affiliations of panel members two years before are listed on the Japan Society for the Promotion of Science's (JSPS) website.[6] For example, an applicant under physical chemistry in 2004 would know that there were twelve members on the 2002 physical chemistry review committee.[7]

The chair of the committee distributes the application by mail to six members on the basis of a quick overview of the research theme. Nevertheless, each reviewer often must review over 150 applications covering a wide range of subjects within five weeks spanning the busy New Year's period. Reviewers rate each application on a 0 to 5 scale for research theme (originality, importance, etc.) and also for research plan (clarity, feasibility, ability of researchers, etc.). The six sets of scores are then sent to a higher level committee that ranks applicants according to their scores, makes adjustments if necessary and also makes preliminary funding decisions. There is one such higher level committee for all fields of chemistry, and it

[5] This includes topics such as molecular structure, crystal structure, electron states, radiation chemistry, chemical reactions, fluid chemistry, molecular spectroscopy, high molecular energy state processes, electrical chemistry, radiation chemistry, electron energy processes and surface and boundary chemistry.

[6] JSPS is a corporation under MEXT that has handled many of MEXT's extra mural funding programmes as well as scholarships for collaborative activities. JSPS administers the Grant-in-aid programs that tend to have lower per project funding, while MEXT directly administers the programs with large budget projects. In addition to Basic Research grants, MEXT Grants-in-aid include other subprograms with different prospective review mechanisms (see Table 2).

[7] Panel members are nominated by various professional societies to the Japan Science and Technology Council which then suggests names to JSPS. Panel members serve two year staggered terms. The 2002 physical chemistry review panel consisted of one associate professor and 11 professors, ten men and two women, drawn from Tohoku University (the chair), the University of Tokyo, the National Institutes of Natural Science (a major MEXT research center), and Okayama, Kyushu, Hiroshima, Ochanomizu, Nagano, Hokkaido, Hokuriku and Keio Universities.

consists of 12 professors from science, medical and engineering faculties in various universities.[8]

Finally a super committee of 20 persons overseeing all JSPS administered programs certifies the funding decisions of the 18 higher level committees and addresses any global or contentious problems. In fact, however, funding decisions are determined by the scores assigned by the six reviewers from the initial specialty area review committees. *At no point in this process is there discussion among reviewers on the merits of particular proposals.* Reviewers are not required to explain their scores, and reasons for success or failure are not conveyed to applicants. However, applicants whose applications are rejected can request the average of the six reviewers' scores and their approximate percentile ranking.

Approximately 96 billion yen or 850 million USD was distributed in this manner in 2002, over one-third of MEXT's direct competitive support for university R&D. Moreover, in terms of numbers, this accounts for the vast majority of university R&D projects. The average yearly size of Basic Research awards ranged from 1.4 million yen (~13,000 USD) for category C awards to 20 million yen (~180,000 USD) for category S awards.

After MEXT Basic Research Grants-in-Aid, JST's Strategically Promoted Creative Research Program (also known as the *JST Basic Research Program*) is the largest source of university R&D funding. The three main types of projects under this program are CREST (Core Research for Evaluational Science and Technology)[9] PRESTO (Precursory Research for Embryonic Science and Technology)[10] and ERATO (Exploratory Research for Advanced Technology)[11].

CREST applications must be for collaborative research involving several laboratories, and they should be targeted on one of the approximately 12 new research themes that JST announces each year.[12] JST selects a respected scientist to supervise review of applications under each theme. Most of the scientists-supervisors are professors in elite universities. These supervisors in turn select 6 to 8 advisors for their review team. Most of these advisors are also professors in elite universities, although many panels include one person from a large company. The supervisor assigns each application to two advisors who read and score them. Then the committee meets to decide on a short list of candidates who will be interviewed. After the interviews the committee agrees on final awardees. Unsuccessful applicants are informed why their applications failed. In 2003, 117 new awards were made, most for five years. The average annual amount of support was slightly under 1 million USD.

---

[8] The names and affiliations of these committee members are available online from MEXT. In the field of chemistry, three of the twelve members reviewing applications for funding beginning in 2004 happened to be from the University of Tokyo.

[9] In Japanese: senryaku teki souzou kenkyuu suishin jigyou.

[10] In Japanese: sakigake kenkyuu.

[11] In Japanese: souzou kagaku gijutsu suishin jiggyou.

[12] Examples of new themes in 2003 are 'nano-scale processes and manufacturing for high performance communication' and 'molecular bio-elements and biosystems for medical therapies'.

PRESTO projects usually involve a single laboratory. However the process of designating new priority research fields each year and selecting among applicants is similar to CREST. In 2002, 147 new projects were funded. Funding averages about 150,000 USD annually, with most projects lasting three years. PRESTO has a special subprogram to support the training of postdocs and PhD students pursuing dissertation research. Funding under this subprogram supports the junior researcher as well as the mentor and thus is generally higher than projects that support only an individual researcher.

Each *ERATO* project is centered around an innovative scientist with charisma and organizational skills whom JST recruits both to initiate and oversee the project. Only four ERATO projects are initiated each year, but each receives about 3 million USD annually for five years – a generous level of funding. JST staff consult with scientists of all ages for suggestions about persons who are doing pioneering research in areas where Japan needs to boost its S&T capabilities and who are also good team leaders and mentors. After a multistage vetting and interview process, JST selects four project supervisors who are given funding to set up a new laboratory and nearly free reign to recruit a research team. Usually the research teams have consisted of about 30 persons divided among two or three research centres (space is often rented from universities or GRIs).

In the past, few *kouza* members participated in these projects (except for the project supervisor, who was often an academic and was expected to devote about 20 percent of his/her time to the project). The rest were usually researchers from companies seconded to the project for two or three years, and postdocs or PhD candidates finishing their thesis research. Now the proportion of *kouza* participants has increased and the proportion of industry participants has decreased. In the first ten years of the ERATO program (1981 to 1990) about one quarter of the projects were headed by industry scientist, but since then only about five percent have had industry supervisors, indicating a shift to more basic research themes. Supervisors are predominantly from elite universities, even more so over the most recent ten year period.

Of all of Japan's government science programs, ERATO has received the most praise from Japanese and foreigners. It has been one of the most successful programs in terms of generating patent applications and academic papers co-authored by scientists in different institutions.[13] Foreigners and Japanese based in foreign universities have participated, even as project supervisors.[14] Examples of successful ERATO research show that a top down, non-peer review process of project selection by a small scientific staff can work in some cases.[15] Nevertheless,

---

[13] Source information for this paragraph is provided by a report from the Japan Technology Evaluation Center (JTEC 1996) and by Hayashi (2003).

[14] For example, Dr. Shuji Nakamura, the inventor of the blue diode laser, is supervisor of an ERATO project that began in 2001, one year after he left Nichia Chemicals and became a Professor in the University of California, Santa Barbara.

[15] One of the first projects in 1981 studied ultra-fine particles and developed methods for depositing thin films of such particles now in commercial use. One of the participants in this project, Dr. Sumio Iijima later discovered carbon nanotubes while working at NEC. Most of the supercomputing electronics

now that the program has shifted to more basic research, a follow-up evaluation may be in order to determine whether significant achievements are arising as frequently as in the past.

Although JST's budget for its Strategically Promoted Creative Research Program is substantial, the number of scientifically trained staff who manage the process of selecting research themes and scientist-supervisors is probably only around twenty, much less than the in house scientific staff of the US NSF or NIH. A small number of senior scientist-advisors to JST make key decisions on research priorities and who will constitute the review committees. These advisors have eminent reputations, but sometimes based upon work many years in the past. Many are still active in universities, GRIs or corporate research, but they are also very busy. This pyramidal top-down decision making system that characterizes the JST programs probably can make good decisions for a country that is still catching up to forefronts of science established in other countries. However, persons who are following the lead of eminent scientists probably have an advantage in applying for funding.[16] For a scientific community that is already at the forefront of human knowledge, such a system may not be the most effective to support research that will extend those frontiers further.

MEXT Special Coordination Funds for Promoting Science and Technology have been in use for several years. Beginning 2001, however, on the recommendation of the Prime Minister's Science and Technology Council, about 150 million USD began to be made available annually for new programs to increase Japan's capabilities in key areas of science, technology and medicine. Emphasis was placed on boosting Japan's international S&T competitiveness, improving university-industry coordination, and ensuring that promising publicly funded discoveries are developed. A unique aspect of the Special Coordination Fund is that they can be used to pay salaries for non-permanent *kouza*-equivalent positions, in other words, for non-permanent (usually five year duration) professorships, associate professorships, etc. In 2003 about 90 awards were made under the various Special Coordination Fund subprograms. Most of these subprograms involve funding for specific projects. The smallest were awards to young researchers, about 20 in number. The largest were three Strategic Human Research Resources awards, each of approximately 10 million USD each annually, made to an entire department or centre to employ non-permanent research staff.

Regardless of subprogram, award decisions were made by 15 Working Groups. One working group, for example, reviewed all applications related to life sciences. It was chaired by the Director of the National Neurological Centre. Its other 13 members included the head of basic research at Ajinomoto, the head of the Intellectual Property (IP) Department of Kissei (a medium size pharmaceutical

---

research worldwide is now based on the manipulation of single flux quanta that builds upon research carried out in an ERATO project begun in 1986. Ultra-fine resolution dual laser interferometers for use in X-ray lithography to manufacture computer chips were developed in another ERATO project. Dr. Ryoji Noyori, who won a Nobel Prize for chemistry in 2001, headed an ERATO project on molecular catalysis beginning in 1991.

[16] Based upon conversations with researchers who have applied to these programs.

company), the head of research at GeneCare (a biotechnology company) the Director of the National Centre for Cancer Research (Japan's largest cancer research centre), a research group leader at Riken (MEXT/JST's flagship GRI), and eight professors, five of whom were from either the University of Tokyo or Kyoto University. In other words, members include many busy, prominent persons who can probably provide helpful insights on broad policy issues, but may have difficulty evaluating the details (or lack thereof) in the various proposals.[17] After initial reviews, finalists are called for interviews. Persons who have taken part in these interviews report that questions tend to be general. Following the interviews, the committees make their final recommendations, which are reviewed by an advisory committee consisting of the chairs of the 15 working groups plus six other persons.

The distribution of new awards in 2003 mirrored the distribution of MEXT Grants-in-aid and government commissioned research shown in the Appendix[18] - predominantly to a few elite universities, except that a substantial number of awards went to GRIs.[19] Although there has been no comprehensive evaluation of program results, concerns that some projects were poorly conceived and wasteful have been frequently voiced off the record by senior academics and government officials.

The *Centres of Excellence (COE) Program* was one of the main outcomes of the 2001 Toyama Plan[20] to reform Japanese universities in order to help revive the Japanese economy. It was originally conceived as a means to raise education and research standards in approximately 30 universities so that they could meet world standards of excellence (for the discussion of a similar attempt in Germany see Weingart and Maasen, this volume). It was also intended to introduce the concept of competitive resource allocation based upon external evaluations. Universities could submit applications in 2002 and 2003 for funds to employ post doctoral level researchers and research assistants, pay stipends for graduate students, purchase equipment, build or rent research space, invite leading researchers from overseas and to support international collaborative research. 113 projects were initiated in 2002, 133 in 2003 and 28 in 2004.[21] None were initiated in 2005 and 2006. New projects are expected to be funded beginning in 2007 under a new Global COE Program, with funding concentrated on a smaller number of universities in order to implement the original intent of this program - to develop up to thirty world class academic centres. The retrospective evaluation system mentioned in the text will be used to select recipient universities or departments.

---

[17] The committees can request that outside experts review particular applications. It is not clear how often they did so.

[18] Sources for the information in this paragraph are in documents accessible under various URLs beginning http://www.mext.go.jp/a_menu/kagaku/chousei.

[19] Also, only one of the Young Researcher awards went to an applicant from one of the seven former Imperial Universities, i.e., those listed in table 3 of the Appendix.

[20] The Plan was named after Ms. Atsuko Toyama, Minister of MEXT, who issued the plan in June 2001 after consulting with Prime Minister Koizumi.

[21] Applications for 2002 had to relate to life science; chemistry or materials science; IT or electronics; human literature; or new interdisciplinary fields. Applications for 2003 had to relate to medical science; mathematics, physics or earth science; mechanical or civil engineering; social science; or new interdisciplinary fields. The 2004 applications simply had to be innovative (NSF 2004).

The award process was similar to that for Special Coordination Fund applications. Each of the ten areas mentioned in the previous note had a review committee of 22 to 27 members. The COE review committees tended to be made up of persons from even more diverse backgrounds than the Special Coordination Fund working groups. For example, the 2003 medical science review committee had 23 members including heads of cancer research centres, professors of nursing, the head of a rehabilitation technical college and even the head of a technical high school. These review committees had the option of asking outsiders to comment on particular proposals. Final decisions were made by a committee of six MEXT administrators and 17 other persons, mainly university professors or GRI laboratory heads. Each project receives on average 1.5 million USD per year. The following table shows the disbursement of COE funds in 2006 for projects initiated 2002-2004. It is heavily weighted in favour of the same elite universities that are leading recipients of Grants-in-aid and government commissioned research (Appendix).

Table 1. Centres of Excellence disbursements in 2006 for projects initiated in 2002-2004[22]

Rank	*University*	*10^8 yen*[2]	*Percent*
1	U Tokyo	44.24	12.7
2	Kyoto U	33.35	9.6
3	Osaka U	24.14	6.9
4	Tohoku U	20.06	5.8
5	Keio U	17.69	5.1
6	Hokkaido U	17.39	5.0
7	Tokyo Inst. of Technology	17.21	4.9
8	Nagoya U	17.07	4.9
9	Kyushu U	12.15	3.5
10	Waseda U	10.19	2.9
11	Kobe U	8.51	2.4
12	Tokyo Medical & Dental U	5.02	1.4
13-91	All other recipients	121.81	34.9
Total		348.83	100

The Japanese government has recently placed priority on increasing research funding opportunities for young researchers. However, having to rely on the professor for laboratory space, key equipment, supplemental funding and support staff, means that even recipients of such awards still must coordinate their research with the *kouza* head (Normile, 2004). Some major funding programs involve the

[22] Source: http:// www.mext.go.jp/b_menu/houdou/18/04/06041308/003.htm. Most projects last five years. Throughout this chapter, funding figures are given in units of 10^8 yen, which is approximately equal to 0.9 million US$, or roughly 1 million US$. The exact equivalence depends upon fluctuating exchange rates and purchasing power parities.

distribution of large funds to a senior principal investigator who then distributes the funds to other collaborating *kouza* heads in other departments or universities.[23]

More generally, over one-third of competitive funds available for universities come from programs that tend to fund large projects involving multiple laboratories. These include the following programs listed in table 2, which account for 36% of MEXT's total competitive research budget: Grants-in-aid for Specially Promoted Research; CREST; ERATO; Research for the Future (JSPS);[24] COE; [25] and Special Coordination Funds (SCFs) for Strategic Human Research Resources, Pioneering Research in new Fields and Training for Emerging Fields.[26] In addition, most funding from the New Energy Development Organization (NEDO)[27] of the Ministry of Economy, Trade and Industry (METI), and most Ministry of Public Management funding for universities also involve either multiple laboratories or (more frequently) a combination of university laboratories and companies (table 3). Table 4 provides the funding totals for competitive research funding in Japan in FY 2002.

---

[23] MEXT's Priority Area Research projects (recently folded into the new Development of Innovative Seeds and the Promotion of Key Technologies Programs) and JST's CREST and ERATO projects tend to be of this type.

[24] The latest projects were started in 1999. The program is being phased out. Some functions will be carried on by new programs.

[25] The Cabinet Office's 2002 list of competitive S&T funding programs that is the main data source for table 2 (note 28) does not classify the COE Program as 'competitive research support.' I have nevertheless included it in my analysis because of its importance as a major new source of S&T research support, and because project proposals are solicited and evaluated competitively. The COE data in table 2 are for projects initiated in FY 2002, the first year of the program (source: NSF Tokyo Regional Office Report Memorandum 02-08 available at www.nsftokyo.org/rm02-08.html). See table 1 for 2006 funding levels.

[26] Information on the various SCF Programs to Promote Reform of the Science and Technology System is at http://www.mext.go.jp/a_menu/kagaku/chousei/gaiyo4.html. Not included in table 2 is another category of SCF--continuation funds for multi-year 'general research' programs begun before 2001. No new projects were initiated in either 2001 or 2002 under these continuing SCF programs, and I do know how these projects were initially selected. These continuation funds amounted to 187 x 10^8 yen in 2002, and are included in the Cabinet Office's 2002 list of competitive S&T funding programs that is the main data source for table 2 (note 28). Thus the total SCF budget in the Cabinet Office list is 187 x 10^8 yen greater than the 178 x 10^8 yen budget shown in table 2 specifically for SCF programs to promote S&T reforms. The total budget for competitive S&T programs shown in the Cabinet Office list is also correspondingly higher than the totals in table 4 (after adjusting for COE funding, which is not included in the Cabinet Office list).

[27] NEDO is a corporation under METI that carries out most of METI's competitive extramural research funding. Its scope covers many fields in addition to energy.

Table 2. Major MEXT competitive research funding programs[28]

Program name	*2002 budget (10^8 yen)*	*No. new projects in 2001*	*Funding range per project (10^8 yen)*	*Duration of projects (yrs)*
University, industry and gov't. cooperation for innovative enterprise creation[A,I]	*71*	*28*	*0.1-0.5*	*3-5*
Grants-in-Aid for Research	*1,703*	*21,000*		
JSPS: Basic research	812	9466	< 1.0	1-5
JSPS: Exploratory research	40	1074	< 0.05	1-3
Support for researchers younger than 37	134	4170	< 0.3	2-3
Specially promoted research	127	13	< 5.0	3-5
Priority area research	386	3394	0.2-6.0	3-6
Disseminating res. results	34	780	Varies	1-5
New Special Coordination Funds to Promote S&T System Reform[26]	*178*	*~150*		
Industry-university-government results-oriented joint research.[A]	28	35	0.27	3
Strategic human research resources development	40	2	< 10.0	5
Research support for researchers younger than 35	15	66	0.05-0.15	$\leq$ 5
Pioneering research in new fields[I]	66	24	0.5-2.0	5
Training for emerging fields	19	7	< 2.0	5
JST Basic Research Program	*427*	*~370*		
CREST[I]	289	173	0.83	$\leq$ 5
PRESTO[I]	64	184	0.17	1-5
ERATO[I]	62	4	3.2	5
International Coop. Research	16	2	?	5
Centres of Excellence[25]	*182*	*113*	*0.10-5.0*	*5*
JSPS Research for the Future	*90*	*0*[7]	*?*	*5*

[A] Program generally has applied research focus or aims to develop competence in particular technical areas.

[I] Program open to applicants or co-applicants from private industry.

[28] Main (umbrella) programs are in *italics*. Subprograms are in normal type. Figures include funding to private companies, GRIs, and universities. The main source is the Prime Minister's Cabinet Office's official list of competitive S&T research funding programs at http://www8.cao.go.jp/cstp/project/compe/haihu02/siryo1.pdf. Data for the subprograms of JST's Basic Research Program are available at www.jst.go.jp. See notes 25 and 26 for additional sources and explanations related to the SCF and COE programs. Since subprograms with annual budgets below 1 B yen (~10 M US$) are not listed in this table, the sums of the budgets for the listed subprograms are less than the total funding levels for the main programs.

Table 3: Major Competitive Government Research Funding Programs of Funding Agencies other than MEXT[29]

Program name	*2002 budget (10^8 yen)*	*No. new projects in 2001*	*Per project funding range (10^8 yen)*	*Duration of projects (yrs)*
Ministry of Public Management (includes former Ministry of Posts and Telecommunications)				
Strategic Communications R&D[A,I]	14	?[30]	0.1-0.5	3-5
Japan Key Technology Centre: Promotion of corporate research in basic technologies[A,I]	107	11	no set limit	usually ≤ 5
Ministry of Health, Labour and Welfare				
Grants in Aid for Health Research[I]	393	1251	0.01–10.0	1-3
Basic Research in Health and Medicine	98	10	0.5-1.0	≤ 5
Ministry of Economy, Trade and Industry (METI)				
NEDO: Industrial Technology Research[A]	53	93	0.3-0.4	2-3
Ministry of Agriculture, Forestry and Fisheries				
Research to Apply Advanced Agricultural Technologies[A,I]	18	?	0.1-1.0	≤ 3
Basic Research to Create New Technologies[I]	42	13	≤ 1.0	< 5
New Enterprise Creation R&D[A,I]	17	6	≤ 0.6	< 5
Environment Agency				
General Environmental Research[I]	29	13 (+ 7 smaller for young researchers)	0.02-1.0	≤ 3
Grants in aid for Research into Fields such as Environmental Disruptors of Biological Pathways[I]	10	30	0.01-1.0	≤ 3

[A] Program generally has applied research focus or aims to develop competence in particular technical areas.

[I] Program open to applicants or co-applicants from private industry.

[29] Includes funding to private companies, GRIs, and universities. Source: document issued by the Prime Minister's Cabinet Office: http://www8.cao.go.jp/cstp/project/compe/haihu02/siryo1.pdf. Programs with annual budgets below 1 B yen (~10 M USD) are not listed in this table. Thus the sum of the listed programs is less than the total for the non-MEXT programs in table 4.

[30] Probably no new projects were approved in 2002. Most of participants in projects initiated in previous years are large electronics companies.

Table 4. Funding totals

Programs	*Funding Total (10^8 yen)*
Non-MEXT programs (including those not listed in table 3)	817
Total for MEXT programs (including those not listed in table 2)	2,657
Total for all competitive funding programs	3,474

In addition, private companies have great influence over Japanese university research. Upwards of a third of inventions in major universities are attributed to company-sponsored research, the percentages being even higher for inventions on which patent applications are actually filed. Private company funding accounts for about 20% of activity-specific university research (i.e., funding net of salaries for full time faculty and administrators, construction, and operational and administrative subsidies). But only about 10% of this is contract research that gives the companies rights to jointly own IP. The other 90% is funding in the form of donations that do not give the donors rights to IP. However, many inventions that probably were made with government or donation support are attributed to corporate-sponsored contract research allowing the corporate sponsors to control resulting IP. Industry researchers engaged in joint research doubled from 1398 in 1992 to 2821 in 2002 (MEXT 2003). This has facilitated the pre-emption by companies of university discoveries by large companies (Kneller 2006).[31] It may also reflect an increasingly applied emphasis in university research.

The *kouza* (laboratory) head is usually responsible for coordination with companies and with laboratories in other universities. Thus young researchers who want to participate in these multi-laboratory projects must do so as part of the larger *kouza*. Such research may have the advantage of bringing many minds to bear on a problem, but it tends to foreclose opportunities for young researchers to pursue new lines of inquiry. It probably also favours the larger universities that already have the equipment, networks and prestige useful for large projects.

Many of the government funding programs have an applied emphasis and some of these encourage collaborations with industry in consortium-like arrangements.[32] Even programs labelled as Basic Research Programs such as JST's CREST, PRESTO and ERATO stress the need for research results to have practical applications and social contributions. In the case of programs funded by METI's NEDO and the Ministry of Public Management, the aim is more explicitly to

[31] In major US universities, probably less than 10% of inventions are attributed to company-sponsored research, and the proportion of these that have company co-inventors is low.

[32] In most of the JST's CREST review panels, one of the (six to eight) members is a representative from industry. Kneller (2007) describes the predominance of consortium research in many government programmes funding cutting edge university research, with the notable exception of MEXT Grants-in-aid.

achieve advances with direct applications for industry. The question arises whether it is appropriate to allocate a large proportion of (young) talent and public funding to topics that are someone else's brainchild, and that may constitute translational or applied research that private companies could fund on their own. Under such circumstances, the need for rigorous, expert based peer review is particularly great.[33]

THE ADVENT OF RETROSPECTIVE RESEARCH EVALUATION

A retrospective RES system has been in place since 2000. Moreover, improving the research evaluation system was one of the main goals of the Second Science and Technology Basic Plan adopted in 2001 (Blanpied 2003). Organizationally, research evaluation has grown out of the accreditation process under the purview of the National Institute for Academic Degrees and University Evaluation (NIAD-UE).[34] Under the NIAD-UE system, universities have considerable leeway how to evaluate themselves, including their research.[35] NIAD-UE has itself reviewed specific departments or faculties within a growing number of universities to check the quality of self-evaluations, and offer feedback. Sometimes the focus is on research within a particular department and sometimes on themes such as international cooperation and exchange that span all the activities of a university.[36]

Many universities are using quantitative indices of research progress such as numbers of publications, research grants, prizes, patent applications, and invited presentations by faculty (MEXT 2005). As a particular example, the University of Tokyo collects information annually from individual faculty on numbers of publications in international journals, contributions to society,[37]contributions to new fields of science and education, numbers of international presentations, number and titles of international collaborations, input into scientific data bases,[38] and major awards.

---

[33] In addition to the careers of young researchers and public resources being at stake, results of applied research projects sometimes are not subject to retrospective evaluation in academic publications. Also such funding may have anti-competitive effects (Kneller, 2007). While NEDO staff often do consult with industry and academic advisors before announcing research themes and they occasional solicit outside evaluations of particular proposals, to a large extent selection decisions are made internally by METI/NEDO staff without broadly solicited, critical debate.

[34] NIAD-UE is an independent administrative agency under MEXT established in 1991 to award university degrees. In 2000 its mandate was broadened (and its name changed from NIAD to NIAD-UE) to include university evaluation.

[35] Self-evaluation should address at least eleven themes (goals). Research is actually a twelfth and optional theme. However, universities that want research accreditation should be evaluated by a working group of NIAD-UE experts every seven years (MEXT 2005).

[36] Thus in 2000, NIAD-UE reviewed research in the faculties of general science in six universities and faculties of medicine in six others universities. In 2001, it reviewed research in law in six universities, education in six other universities and engineering in six other universities. In 2002, it reviewed research in the humanities in nine universities, economics in eight universities, agriculture in seven universities and general science in six universities.

[37] Especially practical benefits of one's research such as products on the market, or actual applications in industry, health, etc.

[38] E.g., gene and protein data bases.

Through to 2009, the results of the evaluations will be for MEXT's and the individual universities' internal use – in the latter case, mainly for self appraisal, improvement of data collection, and promotions. But beginning in fiscal year 2010, they are supposed to be the basis of determining which universities should receive accelerated reductions in operational and administrative subsidies.

The combination of the COE programme, retrospective research evaluation and budget cuts may be seen as a well conceived long term strategy to shift support for Japanese university research entirely to competitive funding. The plan incorporates a generous transition period that will be particularly long in a few universities - not necessarily the elite universities appearing in the Appendix but universities where scientific output in proportion to students or courses is high. In other words, universities where many faculty are engaged primarily in research will be spared the swiftest cuts, precisely to enable productive scientists who spend most of their time on research to continue to do so. This may be in keeping with the advantage often attributed to block grant funding as it enables researchers to undertake riskier, longer term projects than if they had to apply for funding for a new project every several years.

But as suggested above, an alternative interpretation is that this combination reflects a misguided conviction that the path to good science involves conferring elite status on a small number of institutions and providing them preferential access to resources beyond even the highly skewed allocations that exist today – rather than facilitating mobility of researchers and competition among them and their institutions.

In contrast, the US has approximately 200 research universities, of which 96 are classified as research intensive.[39] The distribution of government funding among universities is more even. The vast majority of federal funding is distributed through competitive grants or contracts – the largest portions through the peer review mechanisms of the National Institutes of Health (NIH) and the National Science Foundation (NSF) (NSB 2006). The NIH and NSF peer review committees usually have more expert reviewers and more focussed critical debate on individual proposals than any of the Japanese peer review mechanisms. Thus, they are probably better at detecting novel research proposals from researchers who are not well known (i.e., young applicants) or not from elite universities (Hayashi 1996; Coleman 1999; Suga 2004; Normile 2004; Kneller 2007).[40] Regardless whether it would make sense for Japan to try to implement NSF or NIH-style systems of peer review,[41] the US experience suggests that relying on competitive funding, rather than block grants, is compatible with a high level of university autonomy and a great variety of intellectual approaches in research, as noted by Whitley in this volume – provided funding is plentiful and fairly allocated.

---

[39] According to the classifications of the Carnegie Foundation for the Advancement of Teaching (http://www.carnegiefoundation.org/).

[40] For a critical perspective on NIH peer review, see Kaplan (2005).

[41] MEXT is debating implementing a NSF style system.

However, a system that relies upon competitive, time-limited funding for most research expenses (i.e., a soft money system), must rely on such funding for a significant proportion of the salaries for full time university staff. The prospective peer review system and the administrative competence of Japanese universities must be improved to make a soft money system of salary and research funding work – and to avoid the pitfalls of such a system noted by Whitley, Engwall and Nybom, and others in this volume. So far, however, no system to supplement salaries from competitive research funds has been approved. This is under active discussion, as the subsidy reductions begin to cut into salaries. One sticking point is how the funding agencies will monitor faculty assertions that they are devoting a specified proportion of their time to particular projects. Another issue is that the amounts of competitive awards will have to be increased in order to cover salaries. In other words, much of the money saved from cutting back on block grants ought to be transferred to competitive programs to cover salary costs.

Doubt remains as to whether these issues can be resolved. Even leading Japanese universities have yet to establish strong, competent contract offices than can handle research funds from government and industry. For such a system to work effectively, these offices would have to be able to:

- Ensure the universities' intellectual property interests are protected;[42]
- Collect appropriate overhead (indirect costs) and distribute these to appropriate parts of their universities;[43] and
- Monitor compliance with award terms, including whether faculty are allocating their time appropriately between projects.[44]

But if they did develop such administrative competence, not just with respect to contracts, but also with respect to financial management in general, they could begin to break out of their dependence on government subsidies. They could begin to act entrepreneurially, which (despite risks related to conflicts of interest) might generate more research opportunities, especially for young scientists seeking to pursue new ideas. They could begin to hire permanent or tenure-track faculty using competitively awarded funds, a step no national university has yet taken.[45] Because

---

[42] Currently, the most valuable inventions pass directly to companies with minimal royalty and development obligations (Kneller 2006).

[43] Currently overhead rates are low (10% to 30%, compared with 50% or more in US universities). However, they are set arbitrarily without any attempt to estimate actual indirect costs for university research. In contrast, although US rates are high, they are supported by calculations of various indirect costs that are reviewed by the funding agencies and the General Accounting Office. Moreover, overhead payments in Japan have essentially been hijacked by the individual laboratories and departments performing the research so that they can be ploughed back into additional expenditures by these laboratories. While this boosts resources in these laboratories, it deprives the universities of funds that they might otherwise use for broader benefits, such as providing young researchers with startup funds and smoothing out gaps in competitive funding for individual researchers, so that a system of soft money competitive funding can work smoothly.

[44] Leaving this up to the funding agencies would risk undue bureaucratic delays and intrusion of government into academic affairs.

[45] Competitive funds have been used to pay stipends for graduate students and, less frequently, time-limited faculty positions (primarily research oriented positions, sometimes bestowed upon retirement age researchers from companies sponsoring joint research). The closest approximation to soft-money

these would be university-funded positions, the persons who fill them need not be attached to particular *kouza*.

So far, the impact of retrospective research evaluation is greatest in non-university intramural research, i.e., research conducted within GRIs. For example, in 2003 METI was allocating 10% to 15% of the budget of its flagship GRI, the National Institute for Advanced Industrial Science and Technology (AIST), according to the results of annual evaluations of research results. Researchers whose work was rated as excellent were receiving bonuses. Plans were under discussion to link the budgets of AIST's various component institutes to the results of the annual research results evaluations - to integrate retrospective evaluation into the setting of mid and long term goals (Blanpied 2003). Most of AIST's research funds are allocated from METI, and each component institute has discretion regarding the allocation of funds among various projects. In other words, budget allocation is largely internal, not by external peer review. In such cases, retrospective evaluation is the only means to evaluate research output and to determine whether this output matches agency goals. Thus, retrospective evaluation may be particularly appropriate for GRIs whose budgets are allocated internally among various projects, provided their research mandates are clear.[46]

FINAL REMARKS

As suggested by Whitley in this volume, many features of national public science systems influence the effect of retrospective evaluations and block grant funding. Given the nature of the postwar Japanese research system as summarised in this chapter, the most likely result of the strong system of retrospective evaluation scheduled to take effect in 2010 is justification of budget cuts that will reinforce the elite status of a few universities - although less pessimistic outcomes are possible. In any case, the current soft evaluation system is providing feedback for individual researchers and their universities.

But the larger issue is what combination of funding and review systems is most likely to provide judicious support for a large number of researchers to pursue original creative science. As an initial proposition, it probably makes little sense to impose a strong evaluation system on top of the kind of prospective peer review that has been established in Japan. Here, the likely result is for retrospective evaluation to reinforce and perpetuate the effects of the prospective evaluation system, that is, for the retrospective results to be pre-ordained by the basic funding system. Prospective peer review usually takes into careful consideration past accomplishments, either by individuals (in the case of individual applicants) or by institutions (in the case of applications for broader scope institutional programs). Thus, to do it retrospectively in a separate process would be duplicative.

---

funding of permanent faculty positions have been a few large company funded 'endowed chairs' that provide enough funds to cover the salary of a single professor until retirement.

[46] This is consistent with Cozzens' description in this volume of research evaluation conducted by US GRIs.

On the other hand, in the absence of effective prospective peer review, retrospective evaluation that will guide future funding is probably essential. This is the case in those government laboratories whose funding is distributed by legislative or ministry appropriations. Cozzens' chapter in this volume shows how a semi-weak[47] evaluation system can be helpful to science agencies in determining the success of current programs and for planning new programs. NIH intramural site reviews are clearly a strong type of retrospective system, but largely successful due to the high level of expertise and objectivity of the reviewers.

A tentative conclusion from these examples is that the stronger the review system (either prospective or retrospective), the greater the need for objective, specialist, yet also diversified, expertise among the reviewers. Another lesson that emerges from the Japanese experience is that block grants can reinforce elitism, rather than equalizing opportunities among universities and promoting diverse competing centres. On the surface, MEXT's system of allocating operational and administrative subsidies is objective and formulaic, based upon numbers of faculty, graduate programs, special research facilities, etc. But ultimately these independent variables are decided by MEXT, and the result is the skewed distribution of subsidies shown in the Appendix. Universities that receive relatively low allocations should ask whether, rather than seeking a continuation of subsidies, they should campaign for replacing subsidies with competitive funding based upon objective, expert-based, and transparent peer review. Indeed, this may be the ultimate direction of current MEXT policies.

Good prospective peer review could go a long way to provide young researchers with funds to pursue their own scientific directions, help universities build administrative competence, and establish an alternative to the current semi-feudal system where key administrative decisions impacting education and research are made at the level of individual *kouzas*. Even assuming that the current uneven distribution of resources reflects a corresponding uneven distribution of talent, competitive funding and competent peer review will probably lead to a more even distribution. Capable researchers will be more willing to work in lesser known universities, if they know they can receive funding there. Finally, objective, expertise-based peer review would enable even agencies whose main missions are *not* to support of basic science to make better funding decisions. It should encourage them to articulate their goals more clearly and not leave decisions up to bureaucratic fiat or to the opinions of a few senior scientists. But if a shift to competitive funding is to be successful, other reforms must occur concurrently.

REFERENCES

Blanpied, William (2003), *The second science and technology basic plan: a blueprint for Japan's science and technology policy*, US National Science Foundation Tokyo Regional Office Special Scientific Report no. 03-02 (19 May 2003).

[47] Here I am using 'weak' and 'strong' in Whitley's sense to distinguish between evaluation systems that provide feedback primarily for self-evaluation (weak), and those whose results are used to determine funding allocations (strong, see Whitley, this volume).

Coleman, Samuel (1999), *Japanese Science, From the Inside*, London: Routledge.

Hayashi, Masao (pen name: Hakuraku Rokubiru) (1996), *Amerika no kenkyuuhi to NIH [NIH and Amercian Research Funding]*, Tokyo: Kouritsu.

Hayashi, Takayuki (2003), 'Effect of R&D programmes on the formation of university-industry-government networks: comparative analysis of Japanese R&D programmes', *Research Policy*, 32, 1421-1442.

JTEC [Japan Technology Evaluation Center] 1996, *Japan's ERATO and PRESTO Basic Research Programs*, http://www.wtec.org/loyola/erato/toc.htm (accessed 24 May 2007).

Kaplan, David (2005), 'How to improve peer review at NIH', *The Scientist*, 19, no. 17, 10.

Kneller, Robert W. (2003), 'University-industry cooperation and technology transfer in Japan compared with the US: another reason for Japan's economic malaise?', *University of Pennsylvania Journal of International Economic Law*, 24, 329-449.

Kneller, Robert (2006), 'Japan's new technology transfer system and the pre-emption of university research by sponsored research and co-inventorship', *Journal of the Association of University Technology Managers*, XVIII (no. 1, summer), 15-35.

Kneller, Robert W. (2007), *Bridging Islands: Venture Companies and the Future of Japanese and American Industry*, Oxford: Oxford University Press.

Matsuo, Yoshiyuki (1997), 'Chiteki sangyou e no toushi de aru kakenhi wa jinrui no atarashii bunka o umidasu [Grants-in-aid as an investment for [producing] intellectual property gives rise to a new human culture]', in Tetsuo Kobayashi (chief ed.), *1997 Daigaku Rankingu [1997 University Rankings]*, Tokyo: Asahi shimbunsha shuppan kikaku shitsu [Asahi Newspaper Corporation Publication Project Division]

MEXT [Ministry of Education, Sports, Culture, Science and Technology] (2003), *Kokuritsu daigaku nado no (Kigyou nado tono kyoudou kenkyuu) no heisei 14 nendo no jisshi joukyou ni tsuite [Actual Situation in 2002 Related to Joint Research with Industry in National Universities]*, issued 31 July 2003.

MEXT [Ministry of Education, Sports, Culture, Science and Technology] (2004), *Heisei 16 nendo Gakkou kihon chousa jouhou sho [2004 Basic School Survey]*, Tokyo: Kokuritsu insatsu kyoku [National printing office].

MEXT [Ministry of Education, Sports, Culture, Science and Technology] (2005), *Kenkyuu kaihatsu hyouka ni kansuru genjou oyobi mondaiten (Monkashou ga gisshi suru hyouka) [Current status of research evaluation by MEXT]*, 15 Feb. 2005.

MEXT [Ministry of Education, Sports, Culture, Science and Technology] (2005b), *Heisei 16 nendou daigaku nado ni okeru sangaku renkei nado jisshi joukyou houkokusho [Report on Industry-Academic Cooperation in Higher Education Institutions in 2004]*, June.

Normile, Dennis (2004), 'Older scientists win majority of funding', *Science*, 303, 19 March, 1746. (This summarizes the original May 2003 Japanese report available at www.mhlw.go.jp/shingi/2003/05/s0509-6i4.html.)

NSB [National Science Board] (2006), *Science and Engineering Indicators 2006*, Arlington, VA: National Science Foundation.

NSF [National Science Foundation] (2004), *Tokyo Office Report Memorandum* No. 04-06, http://www.nsftokyo.org/rm04-06.html (accessed 24 May 2007).

Suga, Hiroaki (2004), *Sessatakuma suru Amerika no kagakusha tachi [Diligent American Scientists]*, Tokyo: Kyouritsu.

Uekusa, Shigeki and Kano Takaoka (2005), *Kokuritsu daigaku houjin no zaimu shihyou to sono kanousei [Financial indicators of national universities and their availability]*, Research on Academic Degrees and University Evaluation 1, National Institute for Academic Degrees and University Evaluation.

Whitley, Richard (2003), 'Competition and Pluralism in the Public Sciences: The impact of institutional frameworks on the organisation of academic science', *Research Policy*, 32, 1015-1029.

APPENDIX: DISTRIBUTION OF PRINCIPAL FORMS OF GOVERNMENT SUPPORT FOR UNIVERSITY RESEARCH

Table 1. Monbusho/MEXT Grants-in-aid (all types, new and continuing projects)

Rank	1995			2005		
	University	*Amount (10^8 yen)*	*% of total*	*University*	*Amount (10^8 yen)*	*% of total*
1	U of Tokyo	125.5	13.6	U of Tokyo	201.2	11.7
2	Kyoto U	72.7	7.9	Kyoto U	131.1	7.6
3	Osaka U	61.3	6.6	Tohoku U	94.8	5.5
4	Tohoku U	41.6	4.5	Osaka U	89.8	5.2
5	Nagoya U	34.9	3.8	Nagoya U	64.6	3.8
6	Kyushu U	30.0	3.3	Kyushu U	56.8	3.3
7	Tokyo Inst Tech	30.0	3.2	Hokkaido U	56.1	3.3
8	Hokkaido U	28.5	3.1	Tokyo Inst Tech	45.4	2.7
9	U of Tsukuba	22.2	2.4	U of Tsukuba	30.2	1.8
10	Hiroshima U	13.2	1.4	Riken	26.3	1.5
11	Okayama U	9.5	1.0	Keio U	24.9	1.5
12	Keio U	9.1	0.9	Kobe U	24.7	1.4
13-~500	Other universities	445.5	48.3	Other universities	868.5	50.7
Total		924.0	100		1714.4	100

Sources: For individual universities 1995 see Matsuo (1997). For total 1995 and all 2005 data, see www.jsps.go.jp.

Appendix Table 2: Commissioned Research in 2004 (Source: MEXT 2005b)[48]

Rank	*Institution*	*Amount (10^8 yen)*	*% of total*
1	U Tokyo	177.6	17.5
2	Kyoto U	81.4	8.0
3	Osaka U	77.4	7.7
4	Waseda U	44.1	4.4
5	Tohoku U	42.2	4.2
6	Kyushu U	38.9	3.8
7	Keio U	38.2	3.8
8	Hokkaido U	34.9	3.4

[48] Commissioned research includes most project-specific funding from government agencies. Referring to table 2 in the main text, it includes all funding except MEXT Grants-in-aid, Special Coordination Funds, and COE funding. It also includes contract research from private companies that does not involve company researchers working collaboratively in university laboratories. However, such funding probably accounts for less than 5% of Commissioned Research funds, at least in major universities. Most industry funding is either under Joint Research contracts or donations (Kneller 2003).

Rank	*Institution*	*Amount (10^8 yen)*	*% of total*
9	Tokyo Inst Tech	29.9	3.0
10	Nagoya U	21.1	2.1
11	National Institutes of Natural Science (a MEXT GRI)	19.1	1.9
12	Tsukuba U	13.0	1.3
13-~500	Other universities and academic research centres	394.5	38.9
Total		1012.3	100.0

Table 3. Budget for National Universities' Operational and Administrative Subsidies, April 2004 to March 2010 (projected)[49]

Rank	*Institution*	*Amount (10^8 yen)*	*Approx. % of total*
1	U of Tokyo	5,364	7.3
2	Kyoto U	3,676	5.0
3	Tohoku U	3,122	4.2
4	Osaka U	3,008	4.1
5	Kyushu U	2,819	3.8
6	Hokkaido U	2,541	3.4
7	Nagoya U	2,066	2.8
8-87	Other national universities	51,304	69.4
Approx. total		73,900	100.0

Note: the individually listed top seven recipients happen to be the seven universities that had *Imperial University* status in the prewar era.

[49] These amounts represent total operational and administrative subsidies, not only funds to support research. Sources: For the seven universities, see Uekusa and Takaoka (2005). The six year overall total is an estimate based on overall totals for FYs 2004 and 2005 (drawn from www.mext.go.jp), calculated by applying the same rate of decrease over the entire period, as between FY2004 to 2005, i.e., 98 x 10^8 yen.

CHAPTER 4

PETER WEINGART AND SABINE MAASEN

ELITE THROUGH RANKINGS – THE EMERGENCE OF THE ENTERPRISING UNIVERSITY

ENTERPRISING UNIVERSITY AHEAD?

A little more than a decade ago, 1994, the Center for University Development (CHE) began its mission to organise the 'revolution' (*Umbruch*) of the German higher education system. In other European countries such as the UK and the Netherlands this process was already well under way. The German science system, being more hierarchical and less responsive to the public than their Anglo-Saxon counterparts and additionally burdened with the structural conservatism of a federalist regime, was reluctant and late in reacting to the introduction of numerical indicators, evaluations and rankings to universities and research institutes. Until then, the university system operated under the principle of equality mandated by the constitution for all *Länder* to assure equal living conditions between them, thus also to assure free mobility of students, and the Max Planck Society operated on the assumption that it produced world-class research. Hence, no need to move, so it seemed – despite first initiatives advanced by the German science council suggesting, as early as in 1985, that accomplishments of universities should be judged publicly and in a comparative fashion (Wissenschaftsrat 1985).

The shift only occurred when in 1989 the East German science system had to be downsized and de-politicised before integration into the Western system was possible. For the first time in Germany, formal evaluation processes were applied to institutions of science, albeit to those of the former GDR, but henceforth no credible arguments could be mobilised that would prevent their generalised use in the country as a whole. Thus, 1989 and re-unification marked for Germany, as did the fall of the Soviet empire for the Western industrialised countries, the end of the post-war 'social contract' between science and society. One of the latter's central elements had been the institutionalised trust in the self-regulating mechanisms of science assuring the prudent use of public funds and the ultimate utility for the common good of their expenditure. The erosion of this leading principle gave way to a 'new deal' between science and society, basically resting on the idea of

R. Whitley and J. Gläser (eds.), The Changing Governance of the Sciences, 75–99.

universities becoming both efficient and responsible organisational actors, largely governed by a managerial regime.

Since then a plethora of practices pertaining to 'new public management' have been institutionalised in German universities. They make use of a rhetoric that owes its concepts and rationale mostly to the world of business administration, management schools and consultancies and focus on concepts such as *accountability*, *transparency*, and *efficiency* – increasingly used in more and more societal domains. Be it politics, administration, art, education or individual conduct of life, there is virtually no domain left that has not been influenced by the ideas of efficient management. Regardless of whether the object of management is an individual or an institution, the main techniques are continuous (self-)observation and (self-)intervention: As inner or outer conditions change, a new (self-) intervention is called for. Today, management systems abound that offer help for systemic inspection and the search for adaptive responses. From New Public Management in administrative domains via evaluation systems in science to individual self-management aids, managerial procedures currently pervade contemporary societies (Power 1997).

Note that concepts of accounting and its corollaries are not just rhetoric, they are also based upon techniques that ultimately produce the accountable entities they are targeted at. This is true for individual selves who by way of self-management techniques become capable of steering themselves and others in most flexible ways (Foucault 2000; Bröckling et al. 2000), thereby turning themselves into 'enterprising selves' (Miller and Rose 1995). This is also true for institutional selves such as universities that by way of accounting and other managerial procedures become capable of steering themselves and others in most flexible ways, thereby turning themselves into what we call *enterprising universities.*

It is to the emergence of the new identity of universities that this paper would like to make a contribution. In contrast to the use of related concepts such as "Entrepreneurial university" (Clark 1998) or "the enterprise university" (Marginson and Considine 2000), we consider managerial processes to be a prime indicator of the new university and of the rearrangement of science and society at large for two major reasons. First, the 'entrepreneurial' is not restricted to selected domains of academic activity, notably to technology transfer, but rather about to become characteristic of all academic processes (e.g., teaching, research, governance, knowledge transfer, public outreach). Granted this perspective, the enterprising university, secondly, becomes a prime mover in rearranging the relation of science and society by way of counting and accounting. The 'new deal' rests on enterprising the university so as to render it an efficient and responsible actor, always directed towards the Common Good. While busy subjecting themselves to ratings and rankings, evaluations and excellence initiatives, they seem to lose sight of the contents: what exactly do they consider high quality *Bildung*, where should research be headed? Are the answers really to be found in the multitude of mission statements presented on the Internet by programs, faculties and universities? Indeed, in our knowledge society, knowledge seems to be defined by (managerial) processes rather than by (knowledge) politics.

Before considering the broader claims implicit in this topic at the end of this paper, we will elaborate our argument in four steps: First, we will show how the university is currently turning into an organisational actor. Secondly, we will refer to one example that, if largely implicit, popularises the idea of universities as organisational actors which are badly in need of entrepreneurial spirit: it is the idea of establishing elite universities announced by the German government, hotly debated in the media early in 2004. Thirdly, we will delve into a particular instance of public accounting, that is, into rankings. Both the discussion on establishing elite universities and the recent hype over rankings published in the media testify to the readiness with which a managerial regime has gained almost unquestioned evidence. The debate is more about what exactly we mean by elite university and how exactly to perform and communicate rankings than about the restructuring effects on science (and society) at large. The fourth section will therefore embed this example into some thoughts about the enterprising university in neo-liberal society, a society that is based on innumerable forms of auditing. Our final thesis is that the enterprising university as the core institution of science will also shape the contemporary 'audit society' science (Power 1997).

UNIVERSITIES: ELEMENTS OF ORGANISATIONAL ACTORHOOD

The emerging enterprising university is characterised by four elements of organisational actorhood:[1] organisational accountability (evaluation procedures); the tendency towards defining own organisational goals (mission statements); the implementation of formal procedures and practices serving these goals; and the professionalisation of university management.

Accountability: Quality assurance practices like evaluation and accreditation are important indicators of the overall trend towards accountability. Among many others, the statement by the European University Association is prototypical: "Progress requires that European universities be empowered to act in line with the guiding principle of autonomy with accountability" (EUA 2001: 7). The growing importance of evaluations and accreditations is accompanied by the implementation of specialised organisations and associations. In submitting academic work to standardised techniques of counting and accounting, there is more involved than just organised scepticism and collective criticism. Rather, these practices reflect a broader societal trend towards what Michael Power has called the 'audit society' (Power 1997). The problems of formal measurements notwithstanding (as regards bibliometry, see, e.g., Weingart 2005: 197 and Gläser and Laudel in this volume), these techniques rapidly diffuse into academia. In so doing, the attribution of responsibility, which traditionally has been much more individualised, is now transformed into an organisational account. The university as an organisation has "to explain, to justify, to answer questions" (Trow 1996: 310).

The definition of goals: Today, the universities' homepages abound with mission statements and visions. Their vocabulary regularly highlight 'centres of excellence',

[1] In the following chapter, we draw on a paper by Krücken and Meier (2006) who outlined those four elements that we find most convincing.

'national and international orientation', 'life long learning', 'interdisciplinarity' – in the effort to render the organisation accountable. Yet, what is the effect of only half-believed mission statements? On the one hand, they are designed to instigate organisational change, on the other; they are regularly only loosely coupled to day-to-day decision-making. This ambivalence notwithstanding, these statements should not be underestimated as arbitrary or non-functional. Rather, by mission statements, universities give (albeit ceremonial) testimony to what a higher education organisation today is expected to be. Moreover, this expectation is tightly coupled to evaluative and related practices: In Germany, for example, mission statements are of major importance for the accreditation of private universities (Wissenschaftsrat 2004).

The elaboration of formal structures: Next to the above-mentioned techniques to define self-set goals, the modern university creates formal structures so as to competently and efficiently deal with highly specialised tasks. These structures have two broad purposes: technical elaboration along cause-effect lines, and control. Universities today are equipped with offices and organisational subdivisions for international affairs, personnel development, controlling, gender issues, organisational development, psychological counselling, and technology transfer offices. They both testify to and manage tasks that only recently have become regarded a university's responsibilities in need of offices with specialised areas of competence and staff with special training. Taking technology as an example, informal and personal ties between academic researchers and industry are now explicitly complemented by formal, organised links, while the responsibility for technology transfer shifted from the individual to the organisation.

The rise of the management profession: On the one hand, professors are more and more involved in a variety of rationalised administrative tasks, including personnel management, accounting, and quality control. On the other hand, whole new categories of professionals and related academic management positions are created. An important indicator can be seen in the emergence of specialised journals on higher education management like the "Journal of Higher Education Policy and Management". Further indicators are specialised courses, programs and institutes dealing with higher education management. As Krücken and Meier succinctly put it: "Note that there are obviously two complementary developments taking place: with the management of education comes the education of management" (Krücken and Meier 2006: 256). The default option in much of this is the US system in which market-like institutions govern university forms and behaviour. The curiosity in Europe is that it is the state that artificially creates pseudo-markets in education, health, and other formerly state-regulated policy areas. This move requires organisations to become strategic actors in the new environment.

Summarising, on the operational level the new regime is supposed to enable the universities, at least ideally, to act like corporations, to be responsive and accountable to the public, and to manage themselves. The expectation is that they will act more efficiently with respect to their (self-) assigned tasks than if they continued to be directed by state bureaucracies. For the management of the universities, state controlled budgeting is cast aside in favour of autonomous budgeting and controlling mechanisms are introduced. Recruitment of staff is now in

the responsibility of the university. Departments and interdepartmental institutes become independent cost units. Finally, a new wage system was introduced that allows for bonuses tied to achievement (e.g. 'national' or 'international recognition' in the respective field).[2]

University presidents and deans are granted much more power to hire and fire and to restructure departments. In *Baden-Württemberg* University presidents are officially called CEOs to signal that their universities are being managed like stock companies. To what extent it will change their behaviour is too early to judge. As universities are regarded in this framework of thinking to be like companies they are also expected to act like them, i.e. to position themselves on their respective markets and compete with their rivals. Two such markets are envisaged: one is that of students, the other is that of knowledge. For both good professors have to be attracted in order to teach and do excellent research. Far-reaching visions see the universities (or rather: a few of them) acting on a knowledge market where a major part of the university's income is realised from patents and cooperative ventures with industry, even though this is only realised by very few of the US universities (Slaughter 1993, Mowery et al. 2004).

Two remarks are in place to contextualise these processes for the German case. In Germany, science and higher education policy is an arena the political attraction of which resides with the federalist structure of the country where the *Länder* have jurisdiction over the education system including the universities. It is an arena where political activism can roam freely because the consequences appear only much later when responsibilities can no longer be attributed. Also, it should not be underestimated that the academic community is, or rather was, the last privileged, prestigious estate-like group inducing envy in a mass democratic society which tolerates elite status only on condition of achievement and accountability. Legitimated by democratic egalitarian values and, in the current situation, by imperative budget constraints, the last privileges can be erased without risk.

On the basis of their neo-institutional approach, Krücken and Meier conclude that universities, while being organisations that routinely adapt to external expectations, the spread of global models of modern actorhood, as outlined above, will certainly generate a great variety of realisations, including ritual adaptation and symbolic politics at the level of the individual institution.

> We assume that universities, which also in their past showed a high degree of openness towards their social environments will incorporate new institutional elements easier than those whose institutional history was mainly defined by a concern with purity and a sense of elitism. (Krücken and Meier 2006: 254)

In this vein, they hypothesise technical universities as well as universities founded in an era of mass education to be more receptive to organisational actorhood. We agree: In our view, the presence of both the issue of elite universities and university rankings corroborate this thesis. Both debates give evidence to the fact that German

[2] The drawback is that in order to introduce these measures at no extra cost to the state the income level had to be lowered so that increases can be paid. The lower of two levels – W2 and W3 – is now paid less than a high school teacher which caused the university teacher association to sue Bavaria for the unconstitutionality of this arrangement.

universities have, specific differences notwithstanding, already accepted the idea of organisational actorhood and its concomitant practices and techniques of extended corporate accountability and responsibility.

BECOMING AN ENTREPRENEURIAL ACTOR: ELITE UNIVERSITIES

At the beginning of the German debate on elite universities, there was – not surprisingly – a diagnosis of crisis, a problem to which promoting excellence in academia seemed the obvious answer. When, early in 2004, the then minister of science and technology, E. Bulmahn, announced the launching of a competition among the German universities under a title that had apparently been invented by a PR – agency: "*Brain up – Deutschland sucht die Super-Uni*" (Brain up – Germany looks for the Super University), it had been accompanied by a scenario of brain drain, economic disaster, and dusty intellectuality considered to be hostile towards entrepreneurship. The mix of anglicised nonsense idiom and TV superstar castings was indicative of the short-winded nature of the initiative.

However, while subjected to much criticism, no one denied that universities were badly in need of reform and few denied that creating elite institutions was the wrong move entirely. Whoever expected 'elite' to be the non-issue one could have expected it to be after years of silencing or rather tabooing the very issue, must have been surprised. Although some commentators felt obliged to remind the readers of the 'emotional quality' connected to any mentioning of elite, this cannot be regarded as a major concern. On the contrary, the debate was rather pragmatic and straightforward. Especially, when the Federal government in early 2004 announced its plan to create 'elite universities' after the models of Harvard and Stanford Rectors of the larger universities such as Munich and Bonn rushed into the media with declarations that their schools were already 'elite', and that after 30 years of "mediocrity as the measure of all things" finally the support of excellence was in sight. The "visible development of excellence is something we hold to be urgently needed" said the president of the Rectors' Conference and expressed what had suddenly become a broad consensus.

For the most part, the debate revolved around the question in which sense the notion of elite would make sense at all and that enhancing higher education for some (elite) should not dispense the government from improving the condition for higher education, at large. In so doing, the debate was not only just about elite but rather, thus our reading, it was about universities, *already established as organisational actors*, should now become specifically *entrepreneurial actors*. Notably three topics debated among journalists, academic and political stakeholders give evidence to this 'turn':

- *Bildung*: While some articles pointed to the fact that Germany lacked a shared idea of what *Bildung* is or should be, many called attention to the necessity of keeping up a broad concept. In their view, next to producing high-brow scientists, academia should follow a more sustainable and pragmatic path: A successful concept and practice of *Bildung* is when, in fifteen years from now, we still dispose of well-trained teachers, lawyers and physicians. A major

consensus related to the idea that neither education nor politics have the duty to produce quick economic solutions – they provide but the conditions for them. Those voices insisted on less specified curricula and research agendas. Others emphasised the necessity to develop a university that links *Bildung* more closely to entrepreneurship and creativity. Such opinions were mostly accompanied by more or less ironic comments. One pointed comment should therefore not go unnoticed: "It is weird that only those people talk about elite that themselves do not belong to this category" (Haase 2004). The topic of *Bildung* shows that given the diversity of perceived concepts and tasks related to higher education, this can only be realised by a highly diversified spectrum of universities that – in an entrepreneurial spirit – have to find their respective niche in the market.

- *Modalities of financing*: While it was generally applauded that the government set out to invest some money in a so-called excellence-initiative, almost all comments ridiculed the amount of 1.9 billion Euros only. This was generally not considered a sum that could have any effect at all.[3] All the more so, as this sum was identified as having been saved from other academic investments, be it research or be it infrastructure. Most participants in the debate argued in favour of student fees. They were also eager to point out that in order to maintain equal access to higher education (given equal talents and skills) a complementary system of stipends should be implemented. More generally, authors hinted at the lack of a 'scientific culture' made of a densely woven network of alumni, sponsors, and public services that would contribute extra funds for 'their' respective university. The modalities of financing a costly system of higher education highlights the necessities of universities as entrepreneurial fund raisers who, by way of good performance and consumer relation, acquire additional means.
- *Autonomy*: The universities' quest for autonomy epitomised their demand for strong, entrepreneurial actorhood. The general slogan was: yes, we, the universities want to have more autonomy. This, however, not only means 'more money'. Equally important is 'more room for decisions'. Notably, universities want to select their students themselves, as they attract better professors, and vice versa. In addition, students have to pay tuition, i.e. contribute their share to a society's expenses for excellent education. Also, the salaries for professors, which, as yet, follow the civil service structure, have to be rendered competitive

[3] 1.9 billion Euros were supposed to be distributed among all universities applying for the status of an 'elite university'. When, after more than a year of political in-fighting among the Länder, finally an agreement was struck and the funds could be released the DFG could not even tell potential applicants on the basis of what criteria their proposals would be evaluated. For weeks scientists and university administrations had to follow their intuition. In the end ca. 320 proposals were submitted for three funding programs. After two rounds of evaluations by an international review panel, the winners will receive about 20 million Euros for five years, a lot of money for German universities whose budgets have been cut piecemeal over at least a decade and more. However, it is necessary to recall that the initiative was designed to bring the German universities up to the status of the US 'elite universities': The models are Harvard, Stanford, and MIT. In 2005, Harvard had an endowment with a market value of 25.2 billion US$, followed by Yale University (15.1 billion), Stanford (12.2 billion), Princeton (11.2 billion) and MIT (6.7 billion) (Infoplease 2007). Their annual budgets can be calculated roughly from these figures by assuming a modest 7% interest.

both with respect to the academic and to the extra-academic market. Finally, universities should have the choice to compete freely on the knowledge market with their specific areas of competence, rather than being forced into a reform 'from above', again fixed by many regulations. As if this 'compulsory autonomy' needed explicit reiteration, the new law regulating the universities of Northrhine-Westfalia comes under the title 'Higher Education Freedom Law' (*Hochschulfreiheitsgesetz*).

These topics all clearly rest on the assumption that universities have to be endowed with strong actorhood, if an elite is to come about at all. However, in our reading, the term elite is but a hook on which both politicians and representatives of universities arguing for it hang their case:

As regards politicians, their goal is framed in a straightforward wish for "more Nobel prize winners" (Scholz 2004). Therefore politics should strive to invest notable sums into universities; the growing consensus being that each university should be supported according to its individual strength (Kraft 2004, see also conservative politicians such as Goppel 2004). Even politicians belonging to the Green Party who are generally sceptical about the idea of an elite university proper, advance the idea of fostering already existing strengths by helping to create networks of excellence between universities, extra-university institutes and firms (e.g. Sager 2004). An important task therefore is to differentiate among universities: "Times are gone when each university was able to offer everything" (Wulff 2004). Rather, the state has to provide the means for individual universities to compete with each other by setting their goals for themselves (e.g. Frankenberg 2004). With this set of arguments the university as organisational actor is established. Programmatic differences notwithstanding, all actors in politics as well as education spokespeople in major enterprises adhere to this notion (e.g. Becker 2004).

The same is true for university presidents and other scientific stakeholders. They are united in celebrating the effects it has on the universities. The president of the Science Council (Wissenschaftsrat) is quoted: "I observe a revolution (*Aufbruch*) like never before", Heidelberg university's president sees "a new world", and the science minister of Rheinland-Pfalz is enthusiastic: "The depressive mood is gone" (Spiewak 2005, 45). While all agree that Germany cannot, in earnest, strive for a Harvard university, it should and can advance a diversified choice of excellent institutions in teaching and/or research. If there is any viable role-model in the US, some call for a reorienting their vision from Harvard to either public universities such as the University of Wisconsin or the State University of New York (Weiler 2004). Many scientists call for the implementation of networks of leading institutes for which professional schools would provide a suitable format (Weiler 2004). The diversification of the university system also implies a partial decoupling of the Humboldtian unity of teaching and research (e.g. Bode 2004), as well as the need to adapt the funding system to the pace of scientific development: thus far, it has been far too rigid, sometimes binding research for 10 to 20 years (Käs 2004). Here again, all nuances between the positions notwithstanding, the university as organisational actor is established. Governance, funding, diversification – the university is the major agent. Politicians cannot do more but provide better conditions, either directly

by way of increased financial means, or indirectly, by granting universities the autonomy to select their students and to levy student tuition.

The general message conveyed by the media is that it is impossible to create an elite university by design; it rather evolves – if at all. An evolutionary model would capitalise on strong research centres that developed from third party funds, productive cooperation among institutes of existing research organisations, academic programs and on competition among those centres. The necessary precondition is: Improve the framework for science and research, notable financial resources have to be invested. As some budgets have not been increased for more than 20 years vis-à-vis rising costs for staff and technical infrastructure as well as ongoing inflation some fields (e.g. molecular biology) are doomed to insignificance (Hönig 2004).

The implicit assumption of this discourse is that ultimately the university system has to be reformed so as to create individualised organisational actors endowed with means to govern themselves on their respective markets, orienting themselves towards academia, the economy, *and* the general public: voilà the entrepreneurial university.

The main characteristics of the entrepreneurial university are its efficiency and its responsibility, both intimately tied to each other by way of transparency: Namely, differences in quality between universities, i.e. between their products have to be made visible to their 'customers'. Students (who are misleadingly called that) are supposed to be informed about the quality of teaching. Likewise, the 'customers' of the knowledge production side of the universities, i.e. primarily companies, are supposed to be informed about the quality of research. For this purpose (and, as frank bureaucrats will admit, to get lazy professors spoiled by their civil service lethargy up and running) evaluations have been introduced in order to create differences between universities where similarities and equality has been the principle.[4] Evaluations, in turn, are the tool to create rankings.

While universities, for a long time, insisted on receiving more resources in order to improve their performance, and politicians reacted by asking for more 'value for money', a certain immobility prevailed. However, with the political demand for transparency and the introduction of rankings, things changed. But it was the media, not the universities themselves that responded first to this call for transparency: Rankings of teaching and research (among German, European, and/or international universities) became *news value*, whether the universities liked it or not.

DISPLAYING THE ENTREPRENEURIAL ACTOR: UNIVERSITY RANKINGS

When the first university ranking was published in 1993 by the news magazine *Der Spiegel,* it received harsh criticism for methodological flaws. But it started a competition among the media to publish similar rankings. Since then all kinds or different rankings have been produced both nationally and internationally. They

[4] In 1999 a media campaign about lazy professors *("faule Professoren")* fired up the debate over a new wage system for the universities which would be based on achievement oriented compensation (Kaiser 1999).

have become a fad, and a source of income, for example for the semi-official CHE that, although a private company under the auspices of the Bertelsmann Foundation, has the backing of the Rectors' Conference (HRK). Above all they have become the chief instrument of inducing competition among the German universities and, thus, to provide them with the requisite information of their relative standing, thereby enabling them to devise strategies how to structure themselves, in which areas to invest and which to abandon, which lines of research to pursue, which students to target. The initiative came at a crucial point in time: The universities were under severe stress for financial support after years of coping with deficient funding. Their desperation made them ready to enter into a competition in which they hoped to gain an advantage with their respective strengths. Even before the campaign for the 'elite university' had officially been announced several universities declared on their websites to belong into that category.

If rankings are published, the comparative information is made accessible to the interested public, notably students. Yet, although CHE's primary goal is supposedly to provide information to new students about various disciplines at German universities it also aims 'through comparison to produce transparency of supply and achievement in the higher education sector' and to 'motivate universities to achieve specific profiles' (Berghoff et al. 2005: B1). The most important purpose of rankings is thus to compare similar units or organisations according to specific criteria (indicators) in order to enable them to position themselves and adapt their behaviour if they want to achieve a higher rank.

The crucial condition for rankings to be effective is that they can inform an organisation so as to adapt its behaviour accordingly. As an instrument for efficient, knowledge-based governance it needs to refer to dimensions that the organisation (here: a university) can influence. There are, however, a number of problems associated with the construction of rankings that should be dealt with first before considering their effects. They mostly concern the appropriateness of the chosen units of analysis.

- *Complexity?* First, universities vary considerably in size and internal structure. Large universities are more complex than small ones, their different departments almost inevitably vary in quality, and they are more difficult to direct. As organisations they comprise different functions – teaching, research, knowledge and technology transfer – the respective quality of which may or may not be correlated.
- *Meaningful entity?* Secondly, it is at least problematic to assume that a university as a whole can act like an industrial firm. University presidents will argue that they are in a position to make the requisite decisions but the power of department chairmen and professors is well known (even though it varies between academic systems and over time). For historical and political reasons the universities are more like an organisational shell for professors working in very different disciplines and having very different interests. (This is true at least in continental Europe and post-war Japan). Their role cannot be reduced to that of middle management underlings who can be hired and fired *ad libitum*. For researchers the status of their department is more relevant than that of the

university as the attribution of reputation happens primarily within the discipline and not across disciplinary lines. Loyalty to and interest in promoting the image of the entire university is limited since it contributes to reputation within the relevant community at best in very indirect ways. The reputation of departments clearly rests on the reputation of its prominent members. They gain their reputation predominantly from research, an activity fraught with uncertainty as success depends at least in part on factors outside their control. If funding fads shift, if researchers move or retire that reputation may change quite rapidly. Consequently, the reputation of different departments at one university may vary widely.

- *Relevant dimensions?* Thirdly, the issue is which dimensions are measured and entered into the ranking. Almost all rankings are multidimensional. Ranking entire universities they try to capture their standing as research institutions and as teaching institutions at the same time. Research quality is usually measured in terms of publications and citations or some combination thereof (e.g. CEST). Quality of teaching is sometimes ranked by the student/teacher ratio or by the number of places in computer labs (which are measures of the *conditions* of teaching!), but also by subjective judgments from students, professors and personnel managers (e.g. CHE). It is obvious that a president and his administration cannot influence in a meaningful way the number of Nobel Laureates that their university is to have in the future (this indicator is used in the 'Shanghai ranking', see SJTU 2003). Even the judgments of personnel managers escape their strategies because these opinions are often based on past (personal) experience, coloured by biographical contingencies and rarely on systematic comparison.
- *Experts?* This leads to the problems of 'who ranks'. The use of 'experts' to evaluate universities or even just departments, as is the case in the Times Higher Education Supplement (THES) ranking (THES 2004), has been proved to be fraught with problems. If experts have a large 'cognitive distance' to the fields or units to be evaluated (personnel managers (!) as in CHE's rankings), their judgements do not correlate with the more reliable bibliometric indicators at all (van Raan 2005: 7).

Summarising, the multitude of indicators each of which measures, in highly simplified form, only a particular fraction of the activities that take place within the organisational framework of a university, cannot provide a coherent picture of such a complex institution, let alone rank any number of them along a meaningful scale. Rankings are partly meaningless if they compare units that are incomparable. For these reasons many agencies justifiably shy away from ranking entire universities and, instead, only rank departments.

But evidently, the temptation to rank universities is great because they, rather than departments, are the focus of identities for policymakers and the media. Universities can be identified with cities or regions like soccer teams. They are about to become 'brands' to which their environment (staff, students, alumni, the region, etc.) can relate, provided there is public relation. That is why university rankings have become news value, and the media have embarked on a competition

for rankings. Such sudden public attention has forced the universities to follow suit. After a long time of resistance they now have succumbed to the pressure. Suddenly they strive for a place on the top with an enthusiasm as if great fortunes could be gained from this exercise. With this we now turn to some of the effects that rankings, methodological and/or ideological critique notwithstanding, exert.

To begin with, the Technical University of Berlin (TUB) is an interesting example in that it takes pains to not appearing as mere object of rankings but rather proactively 'sells' the results. To this end, the TUB has published in its press service a list of rankings and listed its respective position. Compiled in a table this not only gives an impression of the plethora of rankings but also of the variety of places that this university has achieved in each of them (table 1).[5]

Table 1. Selection of rankings, indicators and position of the TUB (source: TUB 2005)

Ranking	*Indicators*	*TUB's Ranks*
Times Higher Education Supplement 04	Subjective judgments of researchers, student/teacher ratio, citations, share of international students and professors	18 nat'l 60 world
Shanghai Jiao Tong University 2004	Nobel Laureates, publications in reputed journals, citations	18 nat'l 202-301 world[6]
CEST 'Champions League' 1994-99	Publications (absolute)	246
Focus Magazine 2004	Student/teacher ratio, external funding, reputation; only by subject	5, 6, 10
DFG – funding 2003	Funding in 99-01	20
Humboldt ranking 2003	Attraction of university by number of fellows who chose university 98-02	15
CHE 'Research Universities' 2005	Percentage of 'strong research departments' among total number of departments. 'Strong' = external funds, patents, PhDs, publications	34
'Manager Magazine' 1999	Opinion poll among largest German companies (a) and consultancies (b) which universities they would give preference in computer science	(a): 6 (b): 8
'Capital' Magazine 2003	Poll of personnel managers of the 250 largest German companies, by subject: economic engineering, electrical engineering, mechanical engineering, economics, business administration	3, 5, 9, 17, not ranked

[5] For sources of rankings see the notes to figures 2 to 5. In one case, the CHE ranking, the position of the TU is not mentioned in its press release but instead the ranks in specific subjects are given. This is not accidental. The authors must have found rank 34 not very attractive to report. Instead they boast with results such as a 7th place in PhDs etc. This style of reporting is typical as we shall see below.

[6] Class of ranks 202-301.

It is evident from this selection of indicators that the measured dimensions are divergent, they pertain to a diverse array of activities, and the results of the respective rankings reflect this diversity. The following figures 1 and 2 illustrate the same effect for a number of universities both in national and international rankings.

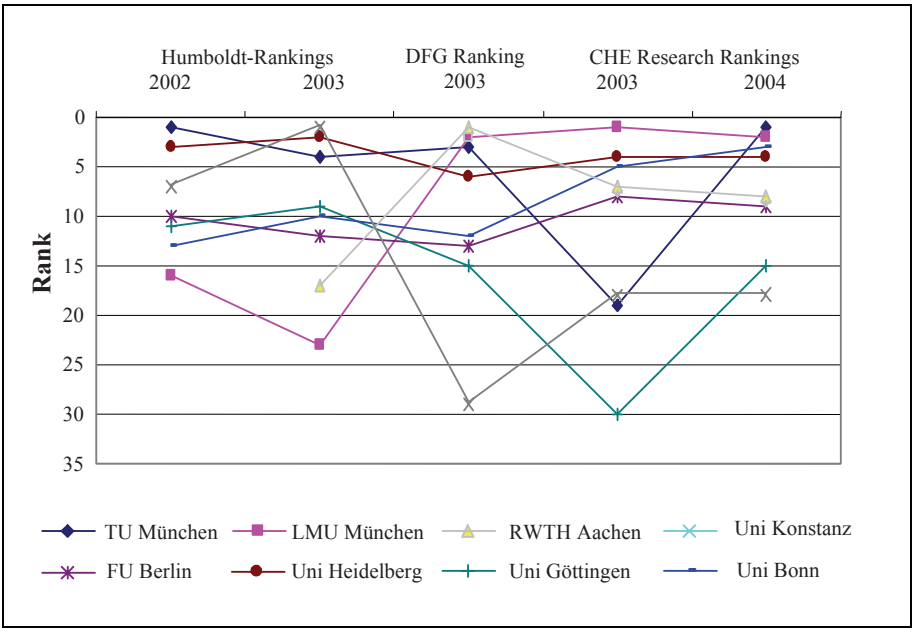

Figure 1. Positions of selected universities in national rankings[7]

[7] Sources: Alexander von Humboldt-Stiftung (2002, 2003), Berghoff et al. (2003, 2004), and DFG (2003a).

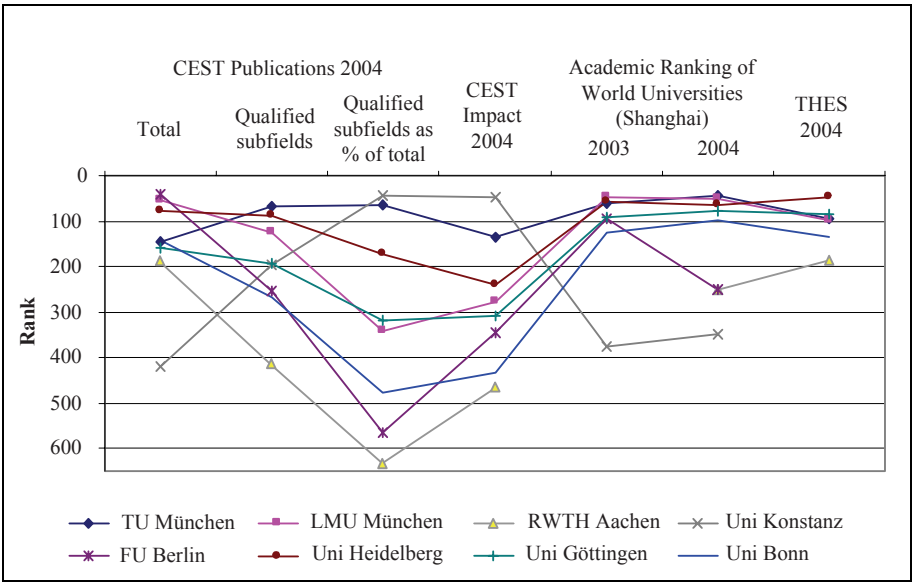

Figure 2: Positions of selected universities in international rankings[8]

Both figures show the differences in ranks for each university individually depending on the ranking and, thus, the indicators used. Where a certain stability can be observed the ranking is the same but in a different year (e.g. Shanghai 2003/2004; CHE 2003/2004 (Berghoff et al. 2005)). For some universities even the *inter-ranking* comparison produces similar ranks. This is typically the case for the very top research universities such as Harvard, Stanford, MIT, Oxford and Cambridge. This seems to indicate that these are institutions of a kind that does not exist in Germany (nor anywhere else in Europe except in the UK).

Finally we look at two German rankings for two consecutive years (2003 and 2004) to see how ranks of particular universities have changed. The expectation is that neither on the dimension of research (figure 3) nor that of the number of Humboldt fellows (figure 4) choosing a particular university would result in dramatic changes within such a short time period. Instead, we observe such dramatic shifts in 2 out of 8 in the case of research, and in 3 out of 8 in the case of Humboldt fellows. In the case of TU Munich and University of Göttingen their upward move can only be an artifact of the ranking method. The shifts in the Humboldt ranking are most likely due to common short-term fluctuations of the number of visitors that make such a ranking highly questionable.

[8] Sources: CEST (2004), SJTU (2003, 2004), and THES (2004).

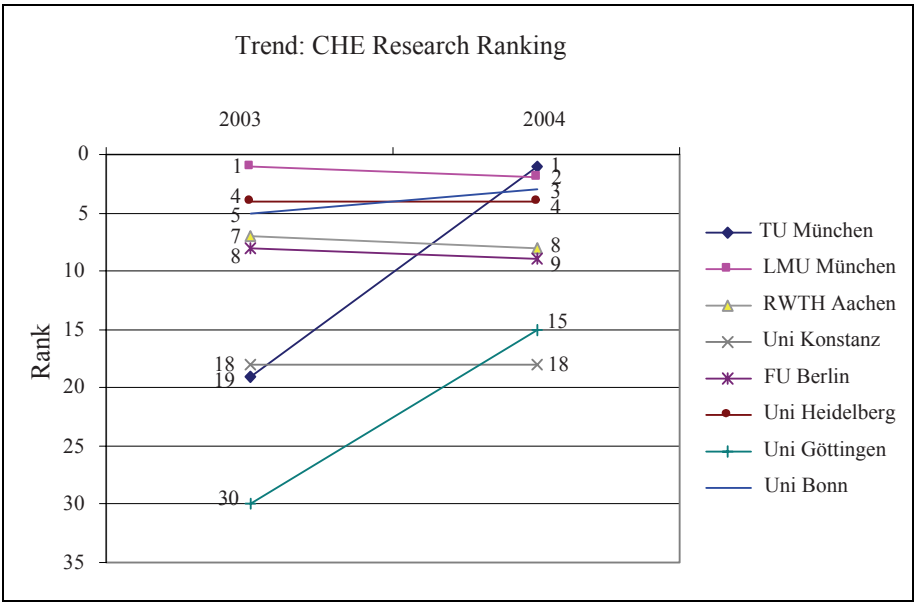

Figure 3: Ranking of research universities (Source: Berghoff et al. 2003, 2005)

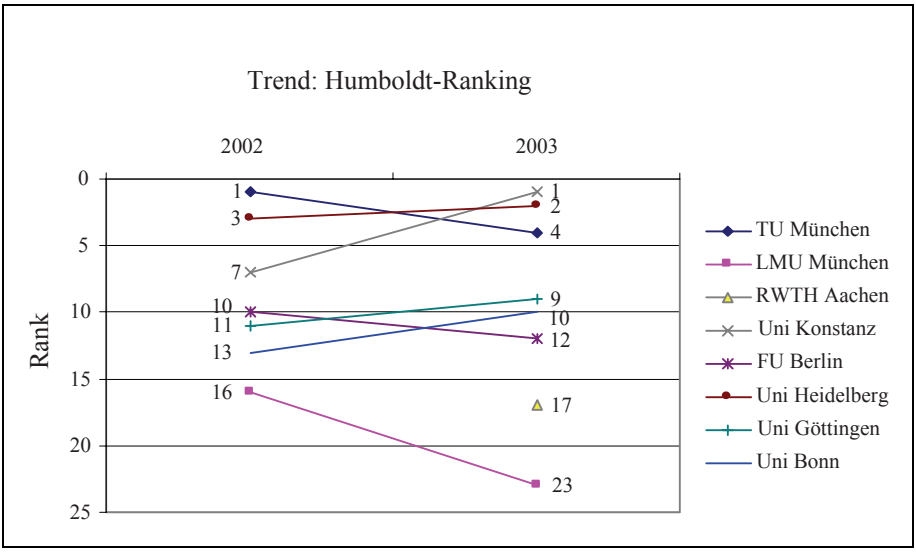

Figure 4: 'Humboldt' Ranking of universities (source: Alexander von Humboldt-Stiftung 2002, 2003)

University[1]	DFG-approvals 1999-2001	Scientific staff 2000: Professors	Scientific staff 2000: Scientists and academics in total	Third party funding in total 1999-2000	Centrality in networks of DFG-funded coordinated programmes 1999-2001	Number of DFG reviewers 1999-2001	Number of AvH visiting researchers 1997-2001	Number of DAAD scientists and academics 2000-2001	Number of DAAD students/ graduates 2000-2001	Participation in the 5th EU-Framework Programme 1998-2002	Publications in international journals (CEST study) 1994-1999
Aachen TH	R 1-10	R 21-30	R 1-10	R 1-10	R 1-10	R 11-20	R 21-30	R 1-10	R 1-10	R 1-10	n/a[2]
München U	R 1-10	R 1-10	R 1-10	R 1-10	R 1-10	R 1-10	R 1-10	R 1-10	R 1-10	R 1-10	R 1-10
München TU	R 1-10	R 11-20	R 1-10	R 1-10	R 1-10	R 1-10	R 1-10	R 11-20	R 21-30	R 1-10	R 1-10
Tübingen U	R 1-10	R 11-20	R 1-10	R 1-10	R 1-10	R 1-10	R 1-10	R 1-10	R 11-20	R 1-10	R 1-10
Erlangen-Nürnberg U	R 1-10	R 1-10	R 1-10	R 1-10	R 11-20	R 11-20	R 1-10	R 21-30	R 21-30	R 1-10	R 11-20
Heidelberg U	R 1-10	R 11-20	R 1-10	R 1-10	R 1-10	R 1-10	R 1-10	R 1-10	R 1-10	R 1-10	R 1-10
Stuttgart U	R 1-10	R 31-40	R 11-20	R 1-10	R 11-20	R 21-30	R 11-20	R 11-20	R 11-20	R 1-10	R 21-30
Würzburg U	R 1-10	R 21-30	R 11-20	R 11-20	R 11-20	R 11-20	R 11-20	R 31-40	R 21-30	R 11-20	R 11-20
Berlin HU	R 1-10	R 1-10	R 1-10	R 1-10	R 11-20	R 1-10	R 1-10	R 1-10	R 1-10	R 21-30	R 11-20
Karlsruhe U	R 1-10	R 31-40	R 21-30	R 1-10	R 1-10	R 21-30	R 11-20	R 11-20	R 1-10	R 1-10	R 21-30
Freiburg U	R 11-20	R 21-30	R 11-20	R 11-20	R 1-10	R 1-10	R 11-20	R 11-20	R 11-20	R 11-20	R 1-10
Bonn U	R 11-20	R 1-10	R 11-20	R 11-20	R 11-20	R 1-10	R 1-10	R 1-10	R 1-10	R 11-20	R 1-10
Berlin FU	R 11-20	R 1-10	R 11-20	R 11-20	R 11-20	R 1-10	R 1-10	R 1-10	R 1-10	R 21-30	R 1-10
Hamburg U	R 11-20	R 1-10	R 1-10	R 11-20	R 1-10	R 1-10	R 21-30	R 1-10	R 11-20	R 11-20	R 1-10
Göttingen U	R 11-20	R 11-20	R 11-20	R 21-30	R 21-30	R 1-10	R 1-10	R 1-10	R 1-10	R 11-20	R 1-10
Köln U	R 11-20	R 1-10	R 11-20	R 11-20	R 21-30	R 11-20	R 1-10	R 11-20	R 21-30	R 21-30	R 11-20
Bochum U	R 11-20	R 11-20	R 21-30	R 21-30	R 1-10	R 11-20	R 11-20	R 21-30	R 11-20	R 21-30	R 21-30
Frankfurt/Main U	R 11-20	R 1-10	R 11-20	R 11-20	R 21-30	R 11-20	R 11-20	R 11-20	R 21-30	R 11-20	R 11-20
Münster U	R 11-20	R 1-10	R 1-10	R 21-30	R 21-30	R 11-20	R 11-20	R 11-20	R 31-40	R 31-40	R 11-20
Berlin TU	R 11-20	R 11-20	R 21-30	R 11-20	R 11-20	R 11-20	R 11-20	R 1-10	R 11-20	R 1-10	R 21-30

Figure 5 DFG Ranking 2003 (ranks 1-20 only, source: DFG 2003b: 133)

A particularly cogent example of how rankings become disinformation is the DFG funding ranking (figure 5). It is not only inaccessible but also combines very different dimensions apart from extra-mural funds received by a university from the DFG (only!).

Although these results are only a selection of the rankings that have been published it is obvious that their diversity reflects, first of all, the media's (and some of the funding organisations') interest in reporting on the exercise. The rankings *do not* contribute to increased transparency of the performance of universities nor have they improved their capacity to act strategically (see Hornbostel 2001: 140).

Given this result, one should, however, not jump to further conclusions: Despite their problems regarding methods, meaningfulness, and potential for (self-) governance, rankings are far more than just window-dressing for those who excel and ignorable asides for those who 'lose'. First of all, they have, to a considerable extent, strengthened the competitive spirit among universities. As van Raan states: "Rankings strengthen the idea of the academic elite, and institutions use the outcomes of rankings, no matter how large the methodological problems are, in their rivalry with other institutions" (van Raan 2005: 5). But without clear criteria about what to compete for, or the power to act accordingly, competition takes a turn into public communication only.

At this point, the university as entrepreneurial actor comes in: In their entrepreneurial capacity, guidance for (self-) governance is what rankings indeed provide due to the fact that *universities respond to the public responses rankings generate*, or rather, to the responses; universities *imagine the public to have*. In this perspective, rankings are an instantiation of *imagined publics* (Gisler et al. 2004). Thus, universities are primarily engaged in reacting to rankings in the media, trying to sell success, to explain failures, and to promise improvement. Their hope is, of course, that reflecting success in the media will ultimately reach their relevant publics: policymakers and students.

It is because, unlike in the UK where the 'research assessment exercise' is firmly institutionalised and operating, in Germany the rankings have not (yet!) been followed up by political measures, the universities have been left to use the results for their own public relations. This is illustrated by the selective interpretations of rankings issued by the university press offices. We take only the CHE – Research ranking of 2004 and reactions to it. CHE ranked the universities according to the share of departments that are ‚strong in research' of all departments. Although the universities are supposedly not ranked as entire units but by subject, the report comes out with a graph and a table ranking the universities as a whole. (Only the lowest group is listed alphabetically). 'Strength in research' is measured in terms of 'extra mural funds,' number of publications, citations per publication (only for the natural sciences) and number of PhDs. Universities are entered on the basis of the number of departments (or rather 'disciplinary units') considered 'strong in research' as percentage of all departments entered. This is how some universities reacted to the ranking 2004.

Table 2: University reactions to rankings[9]

University	*CHE rank 2004 (no. of strong/no. of assessed departments) share of strong departments*	*Subjects*	*Reaction*
TU Munich	1 (4/5) 80%	Chemistry, physics, electrical engineering and information technology, engineering	This is the result of fundamental structural reforms. Our next goal is the international top class. MIT and Stanford are the benchmarks.
LMU Munich	2 (9/12) 75%	English, biology, economics, education, history, law, pharmacology, physics, economics	We can be proud of this top position … we can strengthen it further. We must carry out the necessary structural reforms to be ready for the international competition.
Univ. of Bonn	3 (7/10) 70%	English, history. Law, pharmacology, physics, chemistry, economics	One must always look upon rankings with care but that we are more and more often on one of the first positions is a confirmation of our efforts to develop the University of Bonn into one of the research universities of international rank.
Univ. of Freiburg	7 (7/11) 63.6%	English, biology, history, law, sociology, physics, economics	The university has stabilized its position in the top group …After third positions in the past two years Freiburg has moved to second place in their crucial CHE rankings of 'top position by scientist' (relative indicators).
Univ. of Cologne	16 (4/11) 36,4%	Economics, law, sociology, BWL	University of Cologne makes a very good standing – unfortunately biology and physics have been ranked falsely – as a result of the CHE not having the crucial data for these fields … if one took these data into account the university belonged to the top 10 German universities.

[9] Sources: press releases taken from university websites.

University	*CHE rank 2004 (no. of strong/no. of assessed departments) share of strong departments*	*Subjects*	*Reaction*
Univ. of Rostock	No rank (0/11) 0%, listed in bottom group		This follows inevitably from the methodology as the smaller institutions cannot have the large absolute numbers. One cannot compare a gorilla with diligent ants and declare the gorilla the winner only because it weighs more. The results will be taken seriously …
Univ. of Hohenheim	66 (1/2) 50%	BWL	The ranking proves the university to be one of the strongest research 'lighthouses' in the region together with Stuttgart university. According to CHE definition Hohenheim takes a position in the top group of the 12 particularly strong universities in Germany.

The pattern is anything but surprising. The universities that come out on top declare this as justified and deserved. Those that end up on the lower ranks pick out detailed information that can help to paint a brighter picture and/or they question the methodology of the ranking altogether. (To be sure, the methodology is not quite clear even where it is explained; see Berghoff et al. 2002). In the eyes of CHE these rankings are addressed to established and young scientists and scholars (ibid.). In actual fact they contain little information that could prove to be helpful to them. How are they to translate a university's top or bottom position into concrete action? Instead, the rankings are a media event and are used as a way for universities to position themselves in a race watched by the (imagined) public.

So far (as of early 2006) none of the rankings have been elevated to an official status consented by science policymakers and universities alike.[10] However, in spite of a lack of coercion after years of resistance and critique even on the side of the policymakers and science administrators suddenly the new neo-liberal rhetoric of competition, evaluation and market orientation has taken a firm hold of academia. It has to be noted that the rhetoric initiates and legitimates a sweeping institutional change that goes far beyond the practical, useful, and doable. This can only be explained with the increasing practice and plausibility of the 'entrepreneurial university' that is part and parcel of the all-encompassing audit culture, to which we will now turn.

[10] Efforts are underway to achieve some common measures, and an 'institute of quality control' (IfQ) has been founded under the umbrella of a consortium of the DFG, Humboldt University, the Berlin Science Center, and the Berlin-Brandenburg Academy of Sciences, but it is not clear when and if at all the IfQ will be in a position to create a unified ranking system accepted by the Länder and the scientific community alike.

ENTERPRISING UNIVERSITY IN THE AUDIT SOCIETY

To echo Bill Readings' (1996) *The University in Ruins*, the 'grand narrative' that traditionally defined the function of the university has shifted from the Kantian concept of 'reason' and the Humboldtian idea of 'culture' to the modern techno-bureaucratic idea of 'excellence'. The University of Excellence is less concerned with issues of scholarship or disciplinary knowledge than with 'output productivity', 'Best Practice', 'Quality Assurance' and 'value for money'. The contents of teaching and research matter less than the fact that it be excellently taught or researched (Readings 1996: 13).

Brenneis et al. (2005) rightly maintain that the university of excellence has not simply replaced the earlier models of the university. Rather, as they show by analysing the Draft Strategic Plan of the University of Auckland (a typical document of the accountable organisation), both models co-exist. On the one hand, there is the classical model of a university as a community of scholars, an institution that imparts universal knowledge in the traditions of the cultures and common human values it reflects; and an institution that educates students as responsible and contributing members of, and future leaders of, their societies. On the other hand, there is the more market-oriented and neoliberal model which pictures the same university as "a leader in innovation and the creation of knowledge and development of intellectual and social capital, contributing to the advancement of human condition generally and the increase in wealth and living standards in their local society in particular" (University of Auckland's draft strategic plan 2003, quoted in Brenneis et al. 2005: 4). In other words: Today's university is now expected to succeed with paradoxical goals of fostering the creation of a sort of managerial elite while training a mass of students to underpin the industrial requirements of a nation operating in a competitive global economy dedicated to 'wealth creation' and the exploitation of patents and other forms of intellectual capital (see Smith and Webster 1997: 1).

A striking example for incorporating the classical model into the neo-liberal one in Germany is the 'Zeppelin University' (ZU), a recently founded private university. On its homepage 'about us', we read:

> Zeppelin University defines itself as an individualized, international and interdisciplinary educator of well-rounded decision-makers and creative innovators in the fields of business, culture and politics, as well as a multi-disciplinary research institutions exploring issues relevant to society (….) Our objective: the uncompromising pursuit of academic curiosity and excellence (…) equal emphasis on the development of personality and the acquisition of an excellent academic qualification by balancing today's relevant management and communication tools ('doing tools') with decision-making abilities, methodical thinking and specialized theoretical knowledge... (Zeppelin University 2006)

While the ZU elegantly plays with academic and managerial quality, stylising their skilful combination as ideal education for future managers in 'medialised' knowledge societies, it also testifies to the ease with which the managerialist regime re-orders the prior one. Moreover, it hints at an important reason why it has been introduced and accepted so zealously on a larger scale: The scientific system already abounds with epistemic (e.g., organised scepticism) and non-epistemic values (e.g.,

orientation toward the Common Good) as well as with procedures of self-control (e.g., peer review). Against this backdrop, science is not antithetic, but rather highly receptive to the epistemic value of quality (that is, an allegedly new understanding thereof), to the introduction of more non-epistemic values (e.g., marketability of both university and their students) and to more forms of control (e.g., ratings, rankings) demanded by the techno-bureaucratic model. In particular, the appeals to audit and quality cannot be rejected:

First, the current reform of the university towards an entrepreneurial entity, driven by the imperatives of audit, is designed to subject universities to new regimes of economy and efficiency. The rationale for this is that, as public institutions and recipients of taxpayers' money, universities must be made 'more accountable' to their various stakeholders' and to the public. Most universities thus have adopted a managerial form of accountability geared to measuring and enhancing 'productivity.' Regular external and internal audits of research output and teaching quality have become the norm. We accept these in part because they are couched in the benign language of 'transparency' and 'accountability' that is hard to oppose without appearing Antediluvian. The scientific community cannot deny being accountable to the public, it cannot resist the demand for transparency, because pertaining to its internal procedures of peer review and quality control it lays claim to that value itself, and it cannot, least of all, reject the demand for efficiency and prudence in the use of public money. Audits are, thus, protected by a shield of the uncontested values of democratic governance.

Secondly, in an analogous fashion this protective shield pertains to the norm of quality as well. Readings (1996) made the particularly important observation that it would be anachronistic to think of 'excellence' as an ideology. What makes it so effective as a political instrument is that excellence has no *content:* it is neither true nor false. Like other political technologies, 'excellence' presents itself as emphatically non-ideological – which is precisely why it is so hard to contest or challenge. Could any 'reasonable person' be opposed to raising standards or enhancing quality?

Both appeals to democratic governance and to improving quality seem to immunise the managerial regime of the emerging enterprising university from critique. Indeed, efforts to adapt the university to the exigencies of globalised labour markets, contracting budgets and new forms of governance deserve support. The critique we advance here is not directed against reform of the university nor should a past golden age of the university be implied against which the shortcomings of the present reforms are being judged. The critique is rather directed against the ritualisation of reform, against the mindless application of managerial tools the reasons of whose generalisability are feeble and whose consequences remain un-reflected. It is directed against the irrationalities that indicate the transcendence of the aspirations originally initiating the reform, and it is directed against the all too well known fate of reforms, namely that the reformers don't themselves use the models they insist their targets use. The consequences of the ranking and evaluation exercise craze become glaringly evident when principles of management are shown to be self-contradictory. Here are some examples:

- *Cost.* The managerial reform produces costs. On the one hand, an expanding bureaucracy administering the new regime – both within the universities and outside in the form of the newly established evaluation and accreditation industry – is concerned with implementing the evaluation schemes according to the text book. But often it has no direct experience of the internal organisation of a university nor does it have the time and energy to reflect on secondary effects. Thus, it can be observed already that the most fundamental principle of all evaluative and controlling measures is violated regularly, namely that the cost of these measures must not exceed the gains. Hence, controlling the costs of academic controlling would be a worthwhile endeavour.[11]
- *New academic tasks.* Self-auditing, responding to recurring evaluation exercises, preparing project proposals for grant applications now take up a considerable share of the researchers' time. In some cases, e.g. the EU funding for research, the task of preparing the requirements are so complex that member countries have set up specialised bureaucracies to help researchers. In addition, each university has its own staff overseeing the process. At the same time the success rate of grant applications to major funders has diminished, roughly from 50% in the early 1990s to 30% in 2004, with the tendency to decrease further as more and more applicants join the competition. (In some programs and in some foundations it is said to be as low as 10%). The real success is disputably not the achievement of better research but of increasing the time spent collectively on writing applications. Nowhere are these effects calculated as costs.[12]
- *Contradictory norms of knowledge production.* Growing market pressures on universities continue to exacerbate these imbalances, particularly by eroding disciplinary knowledge practices: With changes in the nature of knowledge production, faculties as well as individual researchers are spending more time working on interdisciplinary teams and workforces, doing community service projects and public scholarship. Interestingly enough, although the call for scholars to produce more 'robust', relevant, and publicly responsible research is tantamount to a novel norm for science, it is never the type of research considered to produce excellence. Transdiscipinarity, in this respect, is a highly time-consuming instance of joint knowledge production that researchers who wish to excel should rather refrain from (see Maasen and Lieven 2006).
- *Controlling what?* Management supports the new regime of audit because it provides a tool not only for measuring productivity but also for 'incentivising' and controlling the academic workforce ('discipline and publish' is how some academics – *pace* Foucault – describe this new regime of 'constant visibility' before the gaze of anonymous officials). The result is that university departments and even individual staff members must now be 'benchmarked'

[11] See the papers by Cozzens and Kneller in this volume discussing the relative merits of input versus output control.

[12] This is at least true for Germany. In the US – several years ahead, as usual – the National Science Board has established a task force to think about devising an alternative to the current review system to allow for radically innovative research. The underlying conviction is, of course, that the present system creates pressures to do mainstream research only.

> against each other and then ranked in competitive national and international league tables – what is more, the criteria for audit-based governance of research are vacuous because they are utterly self-referential: In order to attain the higher score's of one's competitors, the same strategies are being intensified: more publications in higher-impact journals, more extra-mural funds, more international contacts, but this does not answer the political questions which disciplines to favour, which topics to choose, which funders to apply to, in which journals to publish. A research system in which all actors are oriented toward the same external indicators runs the risk of losing its diversity and thus its innovativeness because the indicators are severely limited in their behavioural orientation function, and they trigger the same strategic reactions.

Indeed, perhaps the most perplexing characteristic of the desire to quantify and evaluate everything is that it short-circuits discussion and debate about quality and contents. It acts as a veritable 'anti-politics' machine (see Ferguson 1990; Scott 2001). While advocates of these reforms claim they have weeded out mediocrity and promoted excellence, critics argue that they have induced conflict and stress, undermined collegiality and intellectual freedom, and fuelled a culture of collusion and compliance (Shore and Wright 1999). 'Elite' is created by ranking, i.e. by (quantitative) procedure, not by content. Reflecting on this, Michael Power (1997), too, emphasises that in order to be audited, an organisation must indeed actively transform itself into an auditable commodity, structured to bring their procedures into line with the anticipated standards demanded by external assessors – scholarly, professional or lay. The impact of audit procedures on university culture is therefore to engender a coercive type of accountability.

In the audit society, Power argues that the rise of audit can be explained as a response to the uncertainties of 'risk society'. Taking up ideas of social theorists such as Beck and Giddens, Power proposes that the audit explosion represents "a distinct response to the need to process risk": a process designed to provide "visions of control and transparency which satisfy the self-image of managers, regulators and politicians" (ibid.: 143). What is needed, perhaps, is a more sophisticated understanding of the concept of 'accountability', one that distinguishes between its 'democratic' and 'autocratic' forms. While this call for 'more politics' may seem too idealistic, this need not be the case: We rather suggest applying the principles of accountability, efficiency and transparency to the exercise itself: Universities, as organisational actors, should act responsibly by way of taking account of the costs and gains of accounting first. More deliberations may prove more efficient than hitherto thought of. Finally, ratings, rankings, evaluations and elite, while highly functional for restructuring the system of higher education toward a more diversified and dynamic element of the knowledge society, are badly in need of more transparency themselves. And what could be more appropriate to the spirit of the new governance than self-application?

REFERENCES

Alexander von Humboldt-Stiftung (2002), *Humboldt-Ranking 2002*, Bonn, http://www.avh.de/en/aktuelles/presse/pn_archiv_2002/2002_14_2.htm (accessed 30 May 2007).

Alexander von Humboldt-Stiftung (2003), *Humboldt-Rankings 2003. How Attractive are German Institutions of Higher Education to Top-Flight International Researchers?*, Bonn, http://www.avh.de/en/aktuelles/schwerpunkte/ranking.htm (accessed 30 May 2007).

Becker, F. (2004), 'Mut zu mehr Autonomie', *Rheinischer Merkur*, January 26, 2004.

Berghoff, S., G.Federkeil, P.Giebisch, C.-D. Hachmeister, and D. Müller-Böling (2002), *Das Forschungsranking deutscher Universitäten, Analysen und Daten im Detail*, revised version November 27, 2002, working paper no. 40, http://www.che.de/downloads/AP40.pdf (accessed 30 May 2007).

Berghoff, S., G.Federkeil, P.Giebisch, C.-D. Hachmeister, and D. Müller-Böling (2003), *Das CHE-Forschungsranking deutscher Universitäten 2003*, working paper no. 50, http://www.che.de/downloads/AP50Forschungsranking2003.pdf (accessed 30 May 2007).

Berghoff, S., G.Federkeil, P.Giebisch, C.-D. Hachmeister, and D. Müller-Böling (2005), *Das CHE-Forschungs-Ranking deutscher Universitäten 2004*, working paper no. 62, http://www.che.de/downloads/che_forschungsranking_2004.zip

Bode, C. (2004), 'Wir sollten nicht immer nach Harvard schielen'. *GA-Interview*, January 28, 2004.

Brenneis, D., C. Shore and S. Wright (2005), 'Getting the Measure of Academia: Universities and the Politics of Accountability', *Anthropology in Action*, 12, 1-10.

Bröckling, U., S. Krasmann and T. Lemke (eds.) (2000), *Gouvernementalität der Gegenwart. Studien zur Ökonomisierung des Sozialen,* Frankfurt am Main: Suhrkamp.

CEST [Center for Science and Technology Studies] (2004), *Scientometrics Rankings. Universities and Colleges Participating in the Champions League: Rankings by four Performance Indicators,1998-2002*, Bern: CEST, http://www.cest.ch/Publikationen/2004/four_rankings2002.pdf (accessed 2 May 2005).

Clark, B. R. (1998), *Creating Entrepreneurial Universities. Organizational Pathways of Transformation*, Surrey: Pergamon Press.

DFG [Deutsche Forschungsgemeinschaft] (2003a), *Förder-Ranking 2003. Institutionen – Regionen – Netzwerke. DFG-Bewilligungen und weitere Basisdaten öffentlich geförderter Forschung*, Bonn: DFG, http://www.dfg.de/ranking/download/dfg_foerder_ranking_2003.pdf (accessed 30 May 2007).

DFG [Deutsche Forschungsgemeinschaft] (2003b), *DFG – Funding Ranking. Institutions - Regions - Networks*, Bonn: DFG, http://www.dfg.de/en/ranking/archive/ranking2003/download/dfg_funding_ranking_2003.pdf (accessed 30 May 2007).

EUA [European University Association] (2001), *Salamanca Convention 2001. The Bologna Process and the European Higher Education Area*, Geneva: EUA.

Ferguson, J. (1990), *The Anti-Politics Machine: Development, Depoliticization, and Bureaucratic Power in Lesotho*, Cambridge: Cambridge University Press.

Foucault, M. (2000), 'Die Gouvernementalitat' in Bröckling, U., S. Krasmann and T. Lemke (eds.) *Gouvernementalität der Gegenwart*, Frankfurt am Main: Suhrkamp, pp. 41-67.

Frankenberg, P. (2004), 'Elite-Universitäten müssen sich entwickeln', *Stuttgarter Nachrichten*, January 8, 2004.

Gisler, P., M. Guggenheim, A. Maranto, C. Pohl and H. Nowotny (eds.) (2004), *Imaginierte Laien. Die Macht der Vorstellung in wissenschaftlichen Expertisen*, Weilerswist: Velbrück Wissenschaft.

Goppel, T. (2004), 'Die bayerischen Unis – Spitze trotz Sparzwang', *Süddeutsche Zeitung*, January 8, 2004.

Haase, A. (2004), 'Das Gerede von Elite ist eine Gespensterdiskussion'. *Main Post*, January 10, 2004.

Hönig, M. (2004), 'Notwendige Schritte auf dem Weg zu Elite-Universitäten', *Union der deutschen Akademie der Wissenschaften,* January 9, 2004.

Hornbostel, S. (2001), 'Die Hochschulen auf dem Weg in die Audit Society. Über Forschung, Drittmittel, Wettbewerb und Transparenz', in Stölting, E. and U. Schimank (eds.) *Die Krise der Universitäten*, Sonderheft Leviathan 20/2001, pp. 139-158.

Infoplease (2007), *College and University Endowments 2005*, http://www.infoplease.com/ipa/A0112636.html (accessed 30 May 2007).

Kaiser, G. (1999), *Faule Professoren. Press Release of the Rektor of the University of Düsseldorf*, September 17, 1999.

Käs, J. (2004), 'Eliten gibt es nicht per Dekret', *Leipziger Volks-Zeitung*, January 7, 2004.
Kraft, H. (2004), 'Jede Uni in ihren Stärken fördern', *Süddeutsche Zeitung*, January 7, 2004.
Krücken, G. and F. Meier (2006), 'Turning the University into an Organizational Actor', in Drori, G., J. Meyer and H. Hwang (eds.), *Globalization and Organization: World Society and Organizational Change*, Oxford: Oxford University Press, 2006: 241 – 257.
Maasen, S. and O. Lieven (2006), 'Transdisciplinarity: A New Mode of Governing Science?', in Maasen S., M. Lengwiler and M. Guggenheim (eds.), *Discipline and Research: Practices of Inter-/Transdisciplinary Cooperation in Science,* Special Issue of Science and Public Policy, 33, 399-410.
Marginson, S. and M. Considine (2000), *The Enterprise University. Power, Governance and Reinvention in Australia*, Cambridge: Cambridge University Press.
Miller, P. and N. Rose (1995), 'Production, Identity and Democracy', *Theory and Society*, 24, 427-467.
Mowery, D. C., R. R. Nelson, B. N. Sampat and A. A. Ziedonis (2004), *Ivory Tower and Industrial Innovation: University-Industry technology transfer before and after the Bayh-Dole Act*, Stanford: Stanford University Press.
Power, M. (1997), *The Audit Society. Rituals of Verification,* Oxford: Oxford University Press.
Readings, B. (1996), *The University in Ruins*, Cambridge: Harvard University Press.
Sager, K. (2004), 'Wir sind keine Technikfeinde', *Berliner Zeitung*, January 8, 2004.
Scholz, O. (2004), 'Wir wünschen uns wieder mehr Nobelpreisträger', *Süddeutsche Zeitung*, January 8, 2004.
Scott, J.C. (2001), *Meritocracy and Democracy: The Temptations of Mechanical, "Objective" Impersonal Measures of Quality*, Democratic Vistas, Devane Lecture Series, Yale University, http://www.yale.edu/yale300/democracy/media/apr24.htm (accessed 30 May 2007).
Shore, C. and S. Wright (1999), 'Academic Culture and Anthropology: Neoliberalism in British Higher Education', *The Journal of the Royal Anthropological Institute*, 5, 557-575.
SJTU [Shanghai Jiao Tong University] (2003), Academic Ranking of World Universities - 2003, http://ed.sjtu.edu.cn/rank/2003/2003Main.htm (accessed 30 May 2007).
SJTU [Shanghai Jiao Tong University] (2004), Academic Ranking of World Universities - 2004, http://ed.sjtu.edu.cn/rank/2004/2004Main.htm (accessed 30 May 2007).
Slaughter, E.S. (1993), 'Beyond Basic Science: Research University President's Narratives of Science Policy', *Science, Technology and Human Values*, 18, 278-302.
Smith, A. and F. Webster (1997), *The Postmodern University? Contested Visions of Higher Eduation in Society*, Birmingham: Society for Higher Education and Open University Press.
Spiewak, M. (2005), 'Auf die Spitze getrieben'. *Die ZEIT*, 47, November 17, 2005, 45-46.
THES [The Times Higher Education Supplement] (2004), *World University Rankings: The World's Top 200 Universities*, The Times Higher Education Supplement, http://www.thes.co.uk/search/story.aspx?story_id=2016877 (accessed 30 May 2007).
Trow, M. (1996), 'Trust, Markets and Accountability in Higher Education: A Comparative Perspective', *Higher Education Policy*, 9, 309-324.
TUB [Technische Universität Berlin] (2005), *Hochschulrankings*, http://www.tu-berlin.de/service/ranking.htm (accessed 2 May 2005).
van Raan, A.F.J. (2005), 'Challenges in Ranking of Universities', invited paper for the *First International Conference on World Class Universities, Tong University, Shanghai, June 16-18, 2005*, http://www.cwts.nl/cwts/AvR-ShanghaiConf.pdf (accessed 30 May 2007).
Weiler, H. (2004), 'Harvard lässt sich niemals verwirklichen', *Die Welt*, January 13, 2004.
Weingart, P. (2005), 'Impact of Bibliometrics upon the Science System: Inadvertent Consequences?', *Scientometrics*, 62, 67-85.
Wissenschaftsrat (1985), *Empfehlungen zum Wettbewerb im deutschen Hochschulsystem*, Köln: Wissenschaftsrat.
Wissenschaftsrat (2004), *Leitfaden zur institutionellen Akkreditierung*, July 2004 (Drucksache 6189/04), Köln: Wissenschaftsrat..
Wulff, Carsten (2004), Ein Stück aus dem Tollhaus, *Frankfurter Allgmeine Zeitung*, January 1, 2004.
Zeppelin University (2006), What is ZU, http://www.zeppelin-university.de/index_eng.php (accessed 6 November 2006).

CHAPTER 5

JOCHEN GLÄSER AND GRIT LAUDEL

THE SOCIAL CONSTRUCTION OF BIBLIOMETRIC EVALUATIONS

A CONSTRUCTIVIST APPROACH TO BIBLIOMETRIC EVALUATION TOOLS

While national systems of research evaluation vary in many dimensions, they all need to rely on very few methods of evaluating research performance. These methods constitute a crucial interface between the science system and science policy through which information about research is translated into strategic knowledge for policy decisions. They therefore merit specific attention. This chapter is motivated by the observation that one type of evaluation methods – bibliometric evaluations – is rapidly spreading. Our aim is to identify the processes that drive this growth by creating an image of bibliometric evaluation methods as valid, reliable and therefore legitimate means of performance evaluation. We thus take up Woolgar's (1991) initiative for 'a sociology of measurement technologies and their use in science policies' and apply it to one important type of evaluation techniques, namely bibliometrics. However, we share neither Woolgar's interest in the reasons why bibliometricians have not accepted the fundamental criticisms levelled at citation analyses, nor the *radical* constructivist perspective he applies. Instead, we want to understand why bibliometric methods are being embraced by an ever-increasing number of users with an ever-decreasing regard for validity and reliability. Thus, we are not discussing professional – careful, methodologically sound, state-of-the-art – bibliometric evaluations, which in our opinion can be as valid or invalid as evaluations by peer review. We are interested in the rapid spread of these evaluations beyond the boundaries of this professionalism.

In order to understand this process, we will treat the application of bibliometric methods as a process of knowledge construction and apply insights of the constructivist sociology of scientific knowledge to their analysis. The premise that the interests and powers of actors co-shape construction processes (e.g. Barnes 1977; Pickering 1982) suggests identifying the actors who participate in the construction

R. Whitley and J. Gläser (eds.), The Changing Governance of the Sciences, 101–123.

of bibliometric evaluations, their interests concerning bibliometric evaluations, and their power to influence them. A second important insight concerns the construction of scientific facts. Ethnographic studies (Latour and Woolgar 1986) and citation context analyses (ibid.; Cozzens 1985) have demonstrated that the construction of scientific facts includes a process of decontextualisation. In this process, information surrounding the conditions under which scientific knowledge has originally been produced is gradually omitted. The decontextualisation has been described in the sociology of scientific knowledge as a process of omitting modalities. We apply this heuristic by treating statements about the applicability of bibliometric methods as modalities and by looking at the fate of these modalities in the hands of the different actors. These modalities are omitted (neglected, ignored, or denied) by interested and powerful actors who take part in the application of bibliometric methods for evaluative purposes.

We begin our discussion by clarifying what we regard as the important modalities of bibliometric methods by sketching why they can be regarded as valid under certain conditions, and what statements on limitations of bibliometric methods are embedded in the state of the art of bibliometric research. We then proceed by demonstrating how these modalities disappear because of the interests and powers of the four main actors involved, namely the users in science policy and science management, the commercial owner of the major citation databases, Thomson Scientific's Institute for Scientific Information (ISI), the community of professional bibliometricians, and amateur bibliometricians, i.e. academics, manager and politicians who apply bibliometrics in all sorts of ways without having the necessary professional background.

THE MODALITIES OF BIBLIOMETRIC EVALUATION METHODS

In the following analysis, we treat bibliometric methods as instructions for the collection and manipulation of empirical data. These instructions also contain information about the conditions under which a specific method is applicable, i.e. will yield valid and reliable results, and about the information that is produced by a specific method. This information can be thought about as 'when and how to use' – statements, i.e. as 'modalities' that are part of the instructions.

The concept of 'modalities' was introduced by Latour and Woolgar in their analysis of the construction of scientific facts and has been subsequently expanded by Latour (Latour and Woolgar 1986: 75-90; Latour 1987: 23-29). Modalities are 'statements about statements', i.e. statements that modify the validity and reliability of statements about scientific findings. For example, in the sentence "The structure of GH.RH was *reported to be* X" the modality "reported to be" modifies the statement about the structure of GH.RH. If the modality is dropped, we get "The structure of GH.RH is X", which is a different statement (Latour and Woolgar 1986: 78).

Latour later distinguished between positive modalities "that lead a statement away from its conditions of production, making it solid enough to render some other conditions necessary" and negative modalities "that lead a statement towards its conditions of production and that explain in detail why it is solid or weak" (Latour

1987: 23). A fact is seen here as a statement that is devoid of negative modalities and is linked to positive modalities. It is not accompanied by any information about the persons who produced the knowledge or about the conditions under which the knowledge was produced because this information is regarded as unnecessary. Latour and Woolgar describe the construction of facts as processes in which modalities are removed from statements, and the controversies about facts as processes in which interested parties add or remove modalities to statements in order to weaken them or to make them more solid (Latour and Woolgar 1986: 88; Latour 1987: 25-26).

This general description of fact construction can be applied to methods. Scientific methods contain statements about the conditions under which they can be applied and the way in which they must be applied in order to produce valid and reliable results. These statements are part of the 'prescription of operations' that represents the method. However, they can also be considered as modalities because they modify the methodical statements by limiting their applicability to specified conditions and ways of use. In scientific contexts these modalities are thought of as inseparable from the methods because any application of a method rests on assumptions about applicability and proper procedure. They are implicitly tested in every application of a method, but simply confirmed in most cases. Together with the methodical statements, they are the subject of methodological research, where they are challenged and changed in the attempts to apply a method to new problems. Laboratory studies observed that these modalities can be replaced by locally created ones, a practice which can lead to significant variations of what is thought of as a standard procedure. Such variations, when observed, are subject to serious criticisms by scientists (Knorr-Cetina 1981: 37-40; Latour and Woolgar 1986: 158).

In the remainder of this section, we will identify some of these inherent methodological modalities of bibliometric methods. As mentioned in the introduction, we start from the presupposition that bibliometric methods actually can measure aspects of research quality. Establishing this ability is difficult for the fundamental reason that it is not at all clear what research quality 'really' is. The discussions that circle around this topic usually settle on some statement like 'you know it when you see it' or, more scientifically, that the quality of research is a complex, field-specific property that ultimately can only be judged by members of the scientific community.

This would deny the possibility of any other analysis of research performance than peer review. However, bibliometrics has a strong point in stating that scientists cite their colleagues' work when it is useful for their argument, which means that the cited work has had a certain impact on the citing author's work, and that the usefulness and impact are aspects of research quality. The observation that citations indicate use, and therefore usefulness as well as impact, is the basic argument for using them as an indicator of quality. The reconciliation between the complex and indeterminate concept of research quality and the measurable property of impact has been achieved by regarding impact as *an aspect of* research quality that can be measured by bibliometric indicators, and that bibliometric indicators can therefore serve as 'partial indicators' of quality (Dieks and Chang 1976; Martin and Irvine 1983; Moed et al. 1985a; Weingart et al. 1988; Phillimore 1989; Van Raan 1996).

This interpretation provides the justification of bibliometric evaluations as they are conducted by professional bibliometricians. It also contains the first and most fundamental modality of all citation-based performance measures, namely the principal limitation of this type of evaluative technique that it does not measure quality per se but rather "an important aspect of quality" (Van Raan 1996: 404; see also Moed et al. 1985a: 133-135). Although impact is meanwhile regarded as "[t]he most crucial parameter in the assessment of research performance" (Van Raan 2000: 303), i.e. the modality has been weakened, bibliometricians are aware of the fact that *they do not measure quality*, which is one of the reasons why they regard it as essential that their results need to be interpreted by experts of the field (Moed et al. 1985a: 147; Van Raan and Van Leeuwen 2002: 614-615).

As readers would probably expect from their own experiences with citing and citation, many objections have been raised against the premise that citations indicate use or impact. It is not surprising that one major argument in the debate on bibliometric methods has always been that citations cannot be used as an indicator because they are given and withheld for a great variety of reasons (e.g. Gilbert and Woolgar 1974; MacRoberts and MacRoberts 1989). We refrain from discussing these reasons because we think that there is a convincing rebuttal of the whole argument:

> That is statistically only the case if all researchers refer to earlier work completely arbitrarily. But nobody can seriously maintain that the references in, for instance, this paper are totally unreasonably and completely arbitrary ….valid patterns in citations will be detected if a sufficiently large number of papers is used for analysis. Furthermore, it is statistically very improbable that all researchers in a field share the same distinct reference-biases. (Van Raan 1998: 134-135)

This argument is consistent with all the empirical work about reasons for citations of which we are aware. It is also consistent with both the 'normative' and the 'constructivist' theories of citation. These theories state that the act of citing is governed by compliance with norms ('to give credit where credit is due') respectively by micropolitical interests (to convince – 'enrol' – readers, see Cozzens 1989 for a description of both positions).[1]

The argument that citations *statistically* represent impact contains a second modality. Bibliometric methods must be applied to a larger number of publications for the statistics to become reliable. If enough publications exist and citation data can be accessed, it is possible to measure the one aspect of research quality that is measurable by bibliometrics, namely international impact. This modality has been

[1] Bibliometricians strongly prefer the 'normative' theory because it emphasises the impact of the cited on the citing publication. Their recent attempts to prove the 'normative' theory and to disprove the 'constructivist' theory misconstrue the latter as stating that authors cite highly reputed colleagues (Baldi 1998: 833, 835; White 2004: 93, 96-97) respectively highly cited papers (Moed and Garfield 2004: 295-297) in order to persuade their readers. This is an oversimplification because according to the 'constructivist' theory authors would not restrain themselves to citing authoritative colleagues or papers but would opportunistically cite anything that supports their arguments. Since the most useful tools of 'persuasion' (Gilbert 1977) or 'allies' (Latour 1987) are those that are thematically relevant to the author's research, the 'constructivist' theory would predict the same publications to be cited as the 'normative', namely publications that either have been used in the research leading up to the publication or are used in the publication to support an argument.

explicitly discussed and acknowledged by bibliometricians (Van Raan 2000: 307-309; Butler 2001: 49).

A third important modality of citation studies is technical in nature. For valid conclusions about research performance to be drawn, the whole research output of the evaluated unit must be covered by the analysis. In particular, the citation databases used for bibliometric analyses must enable counting citations to all publications of evaluated units. This is particularly important when bibliometric methods are applied at lower levels of aggregation, i.e. close to the minimal number of publications required to perform reliable statistical analyses. At the level of research groups, 99% completeness of publication data is necessary (Moed et al. 1985a: 139-140), and one missing well-cited publication can create a significant error (Smith 1981: 93; Nederhof 1988: 204).

A fourth modality of the application of bibliometric measures concerns time. Citations that indicate the use of publications and thus their quality occur some time later, and the validity and reliability of citation-based evaluation measures therefore depends on the time-frame chosen for analysis. Since publications reach their highest citation score after three years in many fields of the natural sciences, this time span is considered as the minimal 'citation window' in these fields, and bigger citation windows may be necessary in others (Moed et al. 1985a: 136; Van Raan 1996: 403). This means in turn that the most recent publications for which reliable citation data can be obtained are three years old at the time of the bibliometric evaluation. This trade-off between reliability and timeliness of bibliometric evaluation methods is an inevitable consequence of the modality concerning time.

The fifth basic modality is caused by the specificity of knowledge production in different fields. Since the practices of knowledge production vary between fields, so do publication and citation practices (Moed et al. 1985b). This implies that the results of bibliometric measurements are field specific and can neither be compared nor aggregated without normalising the results with field-specific reference values (Van Raan 1996: 403). Furthermore, the delineation of fields becomes a crucial task on which the validity of bibliometric evaluations depends.

These are the most basic and most important modalities of bibliometric methods. In the following section we will add several others after a closer look at the actual source of citation data. With these modalities taken into account, bibliometric methods can offer information about aspects of research quality. Sociologists of science were the first to use bibliometric methods for evaluative purposes in their studies of the reward system of science and of processes of stratification (Cole and Cole 1967, 1972, 1973). From these early studies, a line of research emerged that develops and uses bibliometric indicators to assess the quality of publications, scientists, organisations, and countries.

MARKETING BY DROPPING MODALITIES

Citation studies require data on the publications of scientists, research organisations, or academic fields, and on the frequencies of their citation. To date, the only databases that offer this sort of data are the Science Citation Index, the Social Science

Citation Index, and the Arts and Humanities Citation Index.[2] These databases are the property of the firm Thomson Scientific that sells online access to the databases, online access to specific indicators, and data derived from these databases to anyone who is interested.

This is not only an absolute monopoly, which is very rare in the economy, it also creates the unusual situation whereby a whole scientific community (the bibliometricians) depends on data that are not a public good but need to be bought. While private property of data occurs in other fields as well (in particular in the life sciences, see e.g. Marshall 1997; Brickley 2002), the problem faced by bibliometrics is much more severe because there is no competing public production of data. The major source of data for the community is privately owned.

This situation suppresses the central mechanism of quality control in scientific communities, namely the use of knowledge in subsequent knowledge production processes. Any such use is also a test of the knowledge that results in explicit or implicit confirmation or in the proposal of changes to that knowledge. Since the bibliometric data is shielded from the community by the property rights of Thomson Scientific, the usual feedback loops of using data, finding and reporting errors, and subsequent improvement of the data do not occur. Any flaws in the data remain there, and there is little improvement over time. This limits the research opportunities for bibliometricians, which was commented on by Glänzel and Schoepflin who wrote that "the databases fall short of the expectations of bibliometricians" (Glänzel and Schoepflin 1994: 380) Similarly, Barre mentions "a fragile and unstabilized situation regarding the supply and quality of data" (Barre 1994: 423; see also Bookstein 1994: 459). Furthermore, peer review of the data generation is impossible. Moed has summed up the results of an extensive analysis of Thomson Scientific's citation indices as follows:

> In our institute's huge analysis of more than 20 million cited references matched to 8 million target articles extracted from the *Science Citation Index* (*SCI*) and related ISI citation indexes, we found that when data are derived from 'simple' or 'standard' citation-matching procedures, citation statistics at the level of individuals, research groups, journals and countries are strongly affected by sloppy referencing, editorial characteristics of scientific journals, referencing conventions in scholarly subfields, language problems, author-identification problems, unfamiliarity with foreign author names and ISI data-capturing conventions. (Moed 2002: 731)

The many criticisms of products sold by Thomson Scientific do not lead to changes in the data. Meanwhile, many bibliometricians are convinced that criticising the owner of the database is a risky business as long as one still needs to buy data from them. While there is no evidence that Thomson Scientific exerts any pressure on the bibliometrics community, the belief that it is better not to contradict them is powerful enough, and certainly shapes actions according to the famous Thomas theorem: "If man believe a situation to be real, it is real in its consequences" (Thomas and Thomas 1928: 572). Since only one source of data is available, database-related

[2] A recently created competitive product (Elsevier's *Scopus*) has the potential to support bibliometric analyses as well.

modalities enter the description of bibliometric methods. The most crucial of these modalities are:[3]

- The databases' coverage of the literature and the changes over time in coverage must be taken into account. The literature of the social science, arts and humanities, but also of some of the engineering sciences and for example mathematics is not well covered by the databases. This undermines the validity of results that are obtained by the use of these databases for bibliometric evaluations in those fields.
- Publications must be unambiguously assigned to authors, which can be difficult because of spelling errors, or homonyms, i.e. different authors having the same last name and initials.
- Publications must be unambiguously assigned to organisations, which is difficult because of incomplete and erroneous information.

Thomson Scientific drops these and many other modalities when offering its own products for research evaluation such as *Essential Science Indicators*, *Highly-Cited.com*, *National Science Indicators* that contain evaluative information about specific countries; *Citation Laureates* (the most highly cited academics) who have been named in award ceremonies in several countries (e.g. Japan, Australia, and Denmark); and evaluations of all sorts published in the bimonthly newsletter *Science Watch*. Examples of dropped modalities include:[4]

- The journal impact factor which is calculated as the average number of times articles from the journal published in the past two years have been cited in the current year. The serious flaws of this factor have been criticised for twenty years now without success (for a recent review, see Glänzel and Moed 2002). Here it suffices to notice that the modality concerning time is dropped and an extremely short citation window (less than two years for one half of the publications and less than one year for the other half) is used. Thomson Scientific has implicitly acknowledged this modality by adding a guide for the calculation of a five year impact factor to its website.
- *Essential Science Indicators'* citation ranking of scientists ignores homonyms. The helpfile indicates that each name in the list refers to one scientist. However, it also adds the interesting comment that "[s]cientists having the same last name and initials may represent multiple individuals". Thus, the modality of citation analysis – that unambiguously assigning publications to scientists is a prerequisite for valid analyses – is acknowledged, but the data and ranking lists remain unchanged.
- In 1995, bibliometricians published the following critique:

> In 1992, several publications appeared in the journal *Science Watch* – published by the Institute for Scientific Information – in which impact indicators have been calculated for universities or even research departments in the field of chemistry. […] We have strong indications that ISI has made several severe errors in their assignment of papers to universities. ISI has not taken into account all variations under which the name of a

[3] An extensive list of modalities that includes the modalities discussed by us and many more has been provided by van Raan (1996: 402-404).

[4] Examples are taken from the websites of Thomson Scientific's "Web of Knowledge" (Homepage: http://www.isiknowledge.com), which was accessed on April 28th, 2006.

> university appears in the addresses. For instance, ISI seems to have missed important variations in the name of the University of Leiden. As a consequence, the number of papers assigned to this university […] is much too low. (Moed et al. 1995: 411-412).

Four years later, the same criticism – that error in addresses produce invalid rankings of universities – was published with regard to the then-marketed product *University Indicators on Diskette*, which proved to be inaccurate in the case of Australia (Butler 1999). Today, *Essential Science Indicators'* citation ranking of institutions still largely ignores variations in addresses. The helpfile states under "Name variations" that "[i]nstitutions may appear with different name abbreviations, in which case more than one entry may need to be consulted". However, some articles are not counted at all due to name variations. Using the 'advanced search' feature of the *Web of Science*, we found 212 Papers that contained only 'ANU' (the common acronym of 'Australian National University', that is not listed in the *Essential Science Indicators*) and checked for the first ten papers and for the most highly cited paper (196 citations) that the addresses in question indeed referred to the Australian National University. Again, the modality of unambiguous assignment is acknowledged, but not acted upon.

The marketing efforts of Thomson Scientific and their impact on science policy have been recently dealt with by Weingart (2005). For our purposes it is important to notice that the products that are marketed are largely devoid of modalities, which means that the validity of the products is questionable. Naturally, the impact of dropped modalities is not mentioned by the seller. Even the information about modalities mentioned above is well hidden. It can be found on the 'Help' sites, i.e. on sites one usually turns to when encountering problems with the intended search. For unsuspicious users, the bibliometric information marketed by Thomson Scientific looks simple and straightforward because most modalities have been removed.

THE DEMAND FOR MODALITY-FREE BIBLIOMETRIC EVALUATIONS

As early as 1975, science policy and management recognised the potential of bibliometric evaluations. In what is probably the first report on this use of bibliometric methods outside the bibliometrics community, Wade (1975: 429) mentioned among other examples the use of bibliometric information by US universities in their decisions on promotion and tenure and by the US National Science Foundation in its assessment of the funding of chemistry departments.

Since then, bibliometric evaluations have diffused into many policy and management processes. Over the last two decades, this diffusion has been driven by the adoption of the paradigm by science policy. This paradigm has at its core the belief that market competition and market exchange are the best way of conducting *any* public task regardless of its content, and that there is no better way of solving allocation problems or producing efficiently. Market competition is supposed to improve efficiency because only the most efficient producers survive, and the market pressure on efficiency forces producers to adopt the most efficient internal structures. Thus, in order to achieve the most efficient conduct of public tasks, these tasks need to be assigned to autonomous units that compete for the resources needed to produce the required outcome. The application of the market paradigm to public

administration has led to the conviction that its organisations should adopt corporate structures and practices, and that competition between them needs to be introduced in order to make sure that public money is only given to the best performers, and is used efficiently by them (James 2001: 233).

The application of 'new public management' to science has resulted in a change of the major governance instruments (e.g. Slaughter and Leslie 1997; Henkel 2000; Marginson and Considine 2000). Governments are consciously or subconsciously introducing market structures into their science systems (see Engwall and Nybom, this volume). The autonomy of research organisations, albeit always limited by their dependence on public funding, is increasing. Competition between research organisations and between researchers is increasing because more and more money is distributed in the form of competitive funding. In many countries, performance-based funding of universities has been introduced over the last three decades or is currently being introduced (see Whitley, this volume). Universities employ their newly gained autonomy by managing research performance and research conditions, a practice which creates a strong trend towards hierarchical corporate structures that weaken the traditional academic self-governance (Morris 2002; Schimank 2005; Gläser and Laudel, this volume).

Both the market constellations and the corporate structures that are evolving in the science system require comparative assessments. The functioning of markets rests on the comparability of products and prices because without comparability a competition for exchange opportunities would be impossible. The management of organisations requires comparisons of the performance of a subunit with that of comparable subunits of competitors, other subunits within the organisation, and the past performance of the same subunit. These needs for comparable performance assessments are the reason why evaluations have been spreading throughout many science systems. Evaluations are the only way to render the idiosyncratic processes of knowledge production comparable, and to make them manageable to outsiders who lack the scientific knowledge to 'manage by content'.

The two basic methods of research evaluation – peer review and quantitative indicators – can both be used in comparative evaluations. In order to achieve comparability with peer review processes, reviewers are usually asked to rank the subjects of their evaluation, thus providing a relationship between the applications that abstracts from their idiosyncratic content and can be used by non-scientists. The now classical example of this use of peer review is the UK's Research Assessment Exercise, in which more than 60 panels of assessors evaluate the outcomes of university research.

The other, more recent evaluation strategy that is being applied more and more widely uses quantitative indicators of research performance. There are three main reasons for the recent rapid growth of the demand for 'metrics', i.e. for quantitative and among them bibliometric methods (see Weingart 2005: 122 for a similar argument). Firstly, the increasing demand for evaluations cannot be met by peer review. The ability and willingness of academics to serve as reviewers of their colleagues is limited, and is already strained by the increasing demand for everyday review activities due to the rise of competitive funding and publications. It is unlikely that peer review alone could meet the demand for evaluation. A related reason for preferring

bibliometric evaluations over peer review is that the latter appear to be cheaper. While it is not entirely clear that carefully conducted professional bibliometric evaluations are necessarily cheaper, the costs of a congregation of assessors are obviously significant.

Secondly, bibliometric evaluations appear to be legitimated by scientific practices as well as being objective and therefore more trustworthy than peer review. Their legitimacy stems from the fact that they rest on cumulated peer judgements. Most publications must pass peer review, and citations indicate peer judgements on usability and impact. Being based on a multitude of those judgements made by a large number of researchers, bibliometric methods appear to overcome the idiosyncrasies and biases of small, all-powerful groups of ad-hoc assessors. Their objective character makes them an ideal tool for routine intraorganisational evaluations, where any kind of peer review would either be very costly (if external assessors are employed) or would put an enormous strain on intraorganisational personal relations if scientists from the same organisation acted as assessors. After all, the peer review of colleagues violates norms of the scientific profession (Schimank 2004).

Thirdly, bibliometric evaluations give the impression of being accessible by politicians and managers without the involvement of scientists. In peer review assessments, recommendations are inextricably linked to idiosyncratic judgements of quality and potential. The numbers produced by bibliometric evaluations appear to be decontextualised and thus can be more easily processed than qualitative judgements by assessors who are outsiders to the policy and management processes.

These perceptions, which make bibliometric evaluations appealing to science policy and management, already miss some of the previously described modalities. They are convincing only as long as it is ignored that the results of bibliometric measurements need to be interpreted by scientists from the field, that they only provide information about past performance, and that they are not equally applicable at all levels of aggregation or to all fields. These and other modalities are readily dropped by science policy and management when requesting or conducting bibliometric evaluations. A leading bibliometrician described this demand:

> Quite often I am confronted with the situation that responsible science administrators in national governments and in institutions request the application of bibliometric indicators that are not advanced enough. They are aware of this insufficient quality level, but the want to have it 'fast', in 'main lines', and not 'too expensive'. (Van Raan 2005a: 140)

While bibliometricians can stand firm in these situations and decline providing 'cheap and dirty' evaluations, science policy and management can often use their discretion to initiate modality-free bibliometrics. We provide examples from three different countries, which cover the policy, intraorganisational collective and individual levels (for more examples, see Adam 2002; Cameron 2005: 112-115).

(1) The following quote from a report on the use of performance indicators in the evaluation of the humanities and social sciences in the Netherlands describes a radical change of mind of Dutch science policy concerning the modalities of bibliometric methods. After observing that in the 2001 Observatory, bibliometric methods

were not applied to the humanities because of their divergent publication culture, the report continues:

> In its 2003 Observatory, however, the Ministry presents tables of 'relative citation-impact scores of Dutch universities, by discipline', as related to the world average by discipline. In these calculations, literary studies in the Netherlands, for example, score much higher than the world average, whereas this indicator is for example much lower for law. Apparently, the Ministry's officials are already applying their own research evaluation methods, while the researchers are still reflecting on the adequacy of such methods, thereby also seemingly neglecting the cautionary notes that are contained in the methodological appendix to the 2000 Observatory. (KNAW 2005: 9)

The initial refusal to apply bibliometric methods to the humanities was based on an acknowledgement of modalities concerning the coverage of the Dutch humanities by Thomson Scientific's databases, which is generally low in the humanities, even lower for journals in languages other than English, and does not include books. Two years later, the modalities were dropped.

(2) Germany is currently introducing performance-based funding in universities and research organisations. A Committee of the Deutsche Forschungsgemeinschaft, Germany's major Research Council, issued recommendations on the performance-based resource distribution in medical university departments (DFG 2004). The recommendations acknowledge the modalities of quantitative performance measurement in general and Thomson Scientific's journal impact factor in particular, and propose the development of procedures and criteria for the assessment of the content of publications. Observing that the development and trial of such procedures is a research project in itself and requires time, the report proposes to use the impact factor as a substitute in the meantime because it is an "albeit imprecise, but relatively inexpensive fallback solution" (ibid.: 15, our translation). The conclusion to this section reads as follows:

> For the calculation of the performance-based funding … the evaluation of original publications can be conducted in terms of a stepwise introduction of quality criteria by using the unweighted impact factors of the respective journals. (ibid., our translation)

Thus, the most important modality of all – that bibliometric methods do not measure quality – and the numerous modalities of the impact factor are dropped for administrative convenience.

(3) In an ongoing project on the evaluation-based funding of Australian university research (see Gläser and Laudel, this volume), an historian described aspects of the promotion procedure (note the shocked responses of the interviewer, who knew about the modalities):

> Historian: … In fact, I have now - for promotion, I've had to look at my citations.
>
> Interviewer: What?
>
> Historian: What you have to do - - -
>
> Interviewer: You go into these citation indices?
>
> Historian: More than that. More than that, because that citation index is not very useful. It's limited. I mean, so I found about 330 citations of my work, but only probably 50 off that. I've found them in other places.

Since the literature in history is not at all well covered by Thomson Scientific's *Arts and Humanities Citation Index*, the historian collected the publications of his field in which citations to his work were likely to occur, and counted these citations manually. The interviewee's promotion was at stake, and he had no choice but to comply with the bureaucratic requirement that ignored all modalities about the validity of citation counts at the individual level, the coverage of the literature by existing databases, the field-specificity of publication and citation practices, etc.

WORKING ON MODALITIES

To conduct a valid bibliometric analysis requires extracting data from the databases of Thomson Scientific, cleaning them, and analysing them carefully by applying state-of-the-art methods. This is what professional bibliometricians do. Many bibliometricians conduct this kind of service as consultancies for political actors or research organisations. However, this is not their major interest. Bibliometrics is an academic field that is aimed at the advancement of knowledge about quantitative analyses of the published output of science (in the widest sense). In the context of our analysis this means that the bibliometrics community works on the modalities of bibliometric methods. Bibliometricians change modalities of existing methods by refining the methods, or they overcome modalities by inventing new methods. Their designing and testing of methods and data establishes, reshapes and erases modalities. The bibliometrics community is the gatekeeper of bibliometric methods and their modalities.

Unfortunately, the bibliometrics community suffers from structural problems that severely limit its opportunities to act as a watchdog for the correct applications of bibliometric methods. The following account is mainly based on the published contributions in a discussion about a 'crisis' of bibliometrics, which occurred in 1993 and 1994.[5] While many discussants disagreed with the diagnosis of a crisis, the structural problems were more or less confirmed in this discussion and in later contributions. The major problems that have an impact on the work with modalities of bibliometric evaluations are:

Epistemic fragmentation and isolation. Glänzel and Schoepflin noticed "failing communication" leading to "parallel studies on the same issue but with completely divergent conclusions and a Babylonian chaos in terminology". They also observed "a lack of consensus in some fundamental questions" and the "drifting apart" of the sub disciplines of scientometrics (Glänzel and Schoepflin 1994: 377). The authors later confirmed the centrifugal trend by a bibliometric study of the journal *Scientometrics* (Schoepflin and Glänzel 2001).[6] The fragmentation is partly driven

[5] Glänzel and Schoepflin presented a paper on the crisis theme in the closing session of a bibliometrics conference in Berlin 1993. The paper and responses from many bibliometricians were subsequently published in a dedicated issue (No. 2-3 of Volume 30) of the journal *Scientometrics* in 1994.

[6] The 2001 article lists six "categories", namely "(1) Bibliometric theory, mathematical models and formalisation of bibliometric laws; (2) Case studies and empirical papers; (3) Methodological papers including applications; (4) Indicator engineering and data presentation; (5) Sociological approach to bibliometrics, sociology of science; and (6) Science policy, science management and general or technical discussions" (Schoepflin and Glänzel 2001: 305). These categories are later referred to as

by the unequal access to data. We already commented on the problematic situation created by the private ownership of the community's major data source (problematic for both the data and the community). Glänzel and Schoepflin hinted at a resulting split of the community:

> On the economic side, there are in fact two classes of bibliometric research groups by now: the ones who can afford to buy expensive data sets, process complex data analyses and plan long-term bibliometric research programs, and the others who cannot do all this. (Glänzel and Schoepflin 1994: 379)

This situation contributes to epistemic fragmentation because some groups develop advanced bibliometric methods that cannot be applied by anyone else, and analyses conducted with these methods cannot be replicated, as the following quote indicates:

> Our study shows that, even for an extensive analysis covering many institutions and thousands of publications, it is possible to extract citation data for non-source items. One major caveat applies to this. For any such analysis to be undertaken efficiently and effectively, direct access to the raw data behind the ISI (or similar) indexes is an essential pre-requisite. (Butler and Visser 2006: 340)

This access only exists in a very few places in the world. We know of no bibliometric research group that has this access and uses it to advance bibliometric methods – except for the bibliometrics group at Leiden University, to which one of the authors of the article (Visser) belongs. The epistemic fragmentation has reached a level where the most advanced findings cannot be replicated by members of the bibliometrics community.

The epistemic fragmentation is accompanied by an epistemic isolation of bibliometrics from other fields of science studies. This isolation has been implicitly acknowledged by Glänzel and Schoepflin (1994: 377, 381), and has later been the subject of several bibliometric analyses of science studies journals (Leydesdorff and Van den Besselaar 1997; Van den Besselaar 2000, 2001). Two attempts to renew the links between quantitative and qualitative science studies (see Leydesdorff 1989; Leydesdorff and Wouters 1996) have not led to changes in the situation. While this isolation of bibliometrics from more qualitatively and theoretically oriented science studies is lamentable, it can by no means be attributed to epistemic changes of bibliometrics alone. The microsociological turn of the sociology of science plays a major role in the separation because the current sociology of science largely neglects aggregate units of analysis such as scientific specialties, to which the statistical approach of bibliometrics applies (Gläser 2001; see also Van Raan 1998).

Weak institutionalisation. Glänzel and Schoepflin wrote that there are only "very few … educational programs for informetrics/ bibliometrics/ scientometrics/ technometrics at universities and colleges" (Glänzel and Schoepflin 1994: 378) and mentioned a split between research groups with and without access to bibliometric raw data (see above). This assessment was confirmed by two respondents to their paper for the situation in Latin America and in the USA.

> Because of the multi- and interdisciplinary nature of our field, it does not easily fit into the hierarchical structure of traditional knowledge fields, and even less so, when

"sub-disciplines" (ibid.: 311). The authors observe that by 1997, the field has become dominated by the sub-disciplines (2) and (3).

> administratively and functionally universities, such a those in Latin America, are bifurcated into science on the one side and humanities on the other. (Russell 1994: 408)

> [In the United States there] are, at the bottom line, few people and virtually no resources for scientometrics; it is a small field with little perceived relevance to major societal goals. (Griffith 1994: 490)

This situation has not changed, and has even become worse in some countries because of the financial pressures on universities.

Lack of standards. The epistemic fragmentation described above already suggests that there might be a problem with missing standards. Glänzel and Schoepflin reinforce this issue by stating that the lack of consensus among the different research groups renders them unable to actively defend their scientific standards (Glänzel and Schoepflin 1994: 381). The bibliometrics community has indeed been troubled by a lack of standards for some time, which becomes clear from the responses to the Glänzel and Schoepflin paper by Vinkler (1994: 499), van Raan (1994: 531), and Luukkonen (1994), and from a workshop dedicated to the topic (see Glänzel et al. 1996 for an overview ; Glänzel 1996; Vinkler 1996). This lack of standards becomes apparent in the current discussion of citation-based performance measures for the social sciences. The most important modality concerns the coverage of the published output in the social sciences by Thomson Scientific's databases, which is regarded as insufficient for valid evaluations by most bibliometricians. However, the position of the bibliometrics community is by no means unambiguous, as the following statements indicate:

- In a report for the UK's Economic and Social Research Council, Katz did not conduct citation impact analysis for most of the social science fields. He stated that "[s]ocial science research is published in a wider variety of publication types and addresses more national issues than natural science research. This makes the construction of internationally comparable bibliometric indicators somewhat problematic." (Katz 1999: i)
- The contrary position is taken by Godin: "one of our main conclusions is that *bibliometrics is as good a tool for measuring the social sciences as it is for the natural, biomedical and engineering sciences*" (Godin 2002: 4, emphasis in original). Godin acknowledges the problem of coverage but counters that "the internationalization of the social fields is changing researchers' publication practices" (ibid.: 11). He backs his argument by referring to Hicks (1999). Hicks indeed observes a trend towards internationalisation of social science (ibid.: 206-208) but states in the summary of her review that "[n]one of the authors discussed here believe SSCI-based bibliometric indicators alone can form a basis for evaluation" (ibid.: 212), which is exactly what Godin has done (Godin 2002: 4).
- Van Leeuwen uses the report by Katz and the paper by Godin as references to back the following statement: "The last couple of years, studies for example in the UK and Canada have indicated the possibilities and advantages of applying bibliometrics in the evaluation of social sciences." (Van Leeuwen 2006: 133) Shortly thereafter, he states that "[t]he main argument against applying bibliometric techniques in the evaluation of social sciences research has always been

(and still is) the (poor) coverage of the social sciences by ISI's SSCI …." (ibid.: 133-134).

- When challenged by one of us (Gläser 2006) for letting a bibliometric evaluation of sociology departments pass regardless of its ignorance of numerous modalities (Sternberg and Litzenberger 2005), the reviewer for the journal scientometrics responded by stating *inter alia* "Without a doubt, the research contributions of economists and social scientists can be gauged from the attention their work receives in journals indexed by the SSCI and EconLit." (Daniel 2006: 332).
- A review of the literature on bibliometric evaluations of the social sciences reaches the conclusion that in the social sciences, "the same bibliometric methods can be applied as in science, but with several extensions. In particular, a broader range of both publications (including non-ISI journals and monographs) and indicators is needed in many social sciences and humanities." (Nederhof 2006: 96)

These contradicting statements indicate that there is no consensus on bibliometric evaluations of the social sciences and humanities. Given the epistemic fragmentation and weak institutionalisation of the community, it is difficult to see how a consensus could emerge. Until it exists we can expect some bibliometricians to conduct bibliometric evaluations which are considered invalid by others.

Commercialisation. The perception of bibliometrics being dominated by science policy and business interests was a major concern of Glänzel and Schoepflin, who considered it to be one of the causes of fragmentation and diminishing quality of some bibliometric studies. The increasing commercialisation of bibliometrics has been confirmed by several respondents, not all of whom, however, interpreted commercialisation as causing a crisis.

> Scientometric research has become indirectly dominated by the interest group "science policy and business". Its interest is clearly focused on "prompt" and "comprehensible" indicators, while the state of knowledge would allow the application of more sophisticated methods. Moreover, such research-reports tend to be only partially published and without the necessary methodological enhancements, which reduces its value for the bibliometrics community. As a consequence a clear shift away from basic and methodological research towards applied bibliometrics can be observed. (Glänzel and Schoepflin 1994: 380)

> When "commercialization" entered the field of scientometrics, it is certain that the commercial value helped develop the discipline. However it is also certain that the field which had originally been "supply oriented" (or "hobby like" affair) has been transformed to "demand-pull" affair. Analysis has to meet the clients' (decision makers) needs, in some cases must even limit observations to what the clients want to see, "science" thus becomes "a good" which must sell. (Miquel 1994: 444)

> Yes, too rapid commercialization is probably one of the biggest threats for our research field. Quick and dirty citation analyses, rainy-Sunday-afternoon publication analyses, stupid hit-lists, completely irresponsible applications of specific methods like co-citation analysis, already did a lot of harm to our research community. (Van Raan 1994: 531)

> Latin American bibliometric research has often been of the kind sponsored by government and scientific bodies interested in raw (and often unqualified) data on

> national scientific achievement for feeding science policy. The production of data without sound theoretical and methodological foundations which is then used to back-up policy decisions implicates the bibliometricians and scientometricians supplying the original data. (Russell 1994: 409)

Bibliometrics has some features of a research technology as described by Joerges and Shinn (Joerges and Shinn 2001), particularly a "generic quality" and the development of a "metrology". However, bibliometric methods are scarcely used as research methods in fields other than bibliometrics itself. The technology does not diffuse into fields of application, either. The bibliometrics community is not an 'interstitial' community but homogeneously institutionalised in the academic sector. Therefore it is better seen as a specialty whose common body of knowledge consists of specific data and methods. The problems of epistemic fragmentation and isolation, weak institutionalisation, lack of standards, and commercialisation, albeit not necessarily symptoms of a crisis, contribute to the inability of the bibliometrics community to guard bibliometric methods. They are partly due to the variety of audiences and funding sources (see Whitley, this volume). However, because of its weak institutionalisation the bibliometrics community strongly depends on the 'soft money' from a variety of sources rather than being able to use this money to shield itself from adverse effects, as has been predicted by Whitley. The diverse audiences and funding sources 'reach through' and fragment the community. Modalities of methods and data are *local* constructs by bibliometric research groups rather than a *global* state of the art of the community. For many bibliometricians, commercial clients are a more influential reference group than the bibliometrics research community. Interests are heterogeneous, too. While there is a widespread epistemic interest in the advancement of bibliometric methods, the dependence of many groups on income from clients creates a parallel commercial interest in satisfying these clients.

Without a consistent interest in defending the modalities or even an agreement on what they are, the bibliometrics community is in no position to fulfil the role usually ascribed to professional experts – at least not beyond the local spheres of influence of a few advanced research groups.

AMATEUR BIBLIOMETRICS

The image of bibliometric methods has undergone a significant change. Initially, they were met with 'hostility' and the threat of legal action (Garfield 1979: 360; Weingart 2005: 117-118). Ten years later Glänzel, Schoepflin and others worried about damage to an otherwise good image of bibliometric methods. Today, according to Weingart, we observe "a dramatic shift away from the well founded skepticism to an uncritical embrace of bibliometric numbers" (ibid.: 119). One of the reasons for this climate change that has not yet been sufficiently acknowledged is 'grass roots bibliometrics' or, more precisely, 'amateur bibliometrics'. We use the latter term to denote the practice of producing bibliometric analyses of an evaluative character by actors with little or no professional background in the field, and with little or no knowledge or regard for the modalities involved.

This practice is more widespread than even the most pessimistic outsider would assume. The use of individual-level bibliometrics in organisational decisions about tenure and promotion has already been mentioned. Today, many organisations routinely use bibliometric indicators for the performance evaluation of their academics (for the US American Library and Information Science, see Meho and Spurgin 2005). Funding agencies in the biomedical sciences often request the submission of journal impact factors or citation counts for the publications contained in the investigators' CVs, thus forcing applicants to conduct amateur bibliometrics. At higher levels of aggregation, ranking exercises that are partly or entirely based on bibliometric indicators proliferate at all levels of aggregation.

Being amateur bibliometrics, all these practices have in common the ignorance of the modalities of bibliometric analyses. They represent extreme cases in which the perceived need for evaluations has overridden any concerns about validity. A characteristic case in point is the 'Shanghai ranking', whose gross misfit of validity and international political impact had moved one of the leading bibliometricians to expose the serious methodological flaws of the exercise (Van Raan 2005a). The authors of the ranking answered that any technical flaws in the database (ISI's citation databases) are the sole responsibility of the authors of the articles contained in the database, and refused to acknowledge any responsibility of the owner of the database or the producer of the evaluation (Liu et al. 2005). Van Raan responded by emphasising the responsibility of the evaluator:

> Of course, the use of a certain description of an affiliation is and will always remain the sole responsibility of authors. But this author-related responsibility is for articles only. As soon as somebody wants to make 'constructions' (i.e., indicators) on a higher aggregation level that constitute an entirely new added value in order to transform 'the world of individual authors and publications' into 'the world of evaluation', it is the responsibility of this person or institute to define an affiliation as good as possible on the basis of any available information that must go beyond the 'local' responsibility of an author for his or her article. (Van Raan 2005b: 111)

Amateur bibliometrics reflects the rapidly increasing demand for evaluations, which it promotes at the same time. It is supported by two interwoven trends. Firstly, the growing competition for funding drives actors to use any resource that might improve their position. Thus, any evaluation of questionable validity will be actively marketed by those organisations that find themselves on top or at least above their major competitors (see Weingart and Maasen, this volume). Similarly, scientists who perform well become increasingly interested in opportunities to 'objectively' demonstrate that they are better than their colleagues. Suppose the historian who collected citations to his publications as part of his application for promotion is successful, and on a later occasion perceives himself as competing with colleagues who he thinks are less well cited than him. There is at least a chance that he will advocate citation analysis (possibly by conducting it for his own work) in order to win the competition for funding or to advance his career. Secondly, the increasing specialisation of research forces scientists to use second-order criteria when judging research performance. In many fields, specialisation has reached a stage where scientists must rely on proxy criteria for performance such as the reputation of colleagues, organisations, and journals. Apart from the informal communication in

which reputation is discussed, numbers of citations, ranks of journals and external funding are adopted as surrogates for the judgement of content. When asked what he would regard as an extraordinary scientific contribution in his field, a scientist answered:

> Our measure comes from impact factors of journals. You know, if you are looking for a global measure, that's how you do it: is your work published in *Science* or *Nature*. They are the gold standard. And they are the gold standard because they accept even for review a very small number of the submitted materials. I think it is much more difficult to apply another standard because standards vary within certain branches of disciplines. So, I am an experimental plant biologist, essentially a plant physiologist and you know there are twenty or thirty branches.

As the quote indicates, scientists are left with the second-order criteria because there is no other way of assessing the quality of contributions from areas which are too remote to understand them but are nevertheless necessary to know about. This occurs in peer review evaluations when assessors cannot judge the content of research (as they should) because it lies outside their area of competence. A common response is resorting to properties of the publication such as the rank of the journal, citations, etc. (Gläser and Laudel 2005).

The need to use crude bibliometric indicators to judge research performance as part of their everyday work and the need to constantly market performance jointly contribute to the acceptance and even active introduction of some person counting another person's or an organisation's citations, that is amateur bibliometrics. An unwanted side effect of amateur bibliometrics is that it lowers the methodological standards of bibliometrics in general, because for amateur bibliometricians, none are needed. The widespread acceptance of bibliometric evaluations resulting from amateur bibliometrics is due to the personal experience of the numerous amateur bibliometricians (anybody can count citations) and of those who do not believe in the validity of these exercises but feel overwhelmed by the ubiquity of amateur bibliometrics. The major danger of this development is that peer review might loose its character as an independent counterpoint to bibliometric evaluations when the assessors – being amateur bibliometricians themselves – simply trust the numbers.

CONCLUSIONS

The discussion of the actors who jointly promote bibliometric evaluations has demonstrated that their interests align well. Science policy and management want quick and easy-to-handle evaluations, and get them because (a) they distribute resources in the science system in a way that requires this sort of evaluation and (b) they are powerful patrons of bibliometrics, which in turn is not coherent and independent enough to create and defend standards of bibliometric conduct. Since neither the clients, nor the owner of the data source, nor the background chorus of amateur bibliometricians care for the modalities of bibliometric methods, and since there is no strong counterweight, modalities are easily dropped. This is not to say that there are no considerate, well constructed bibliometric evaluations. There are many of

those, but they are not as cheap as many clients need them, and the few groups who can produce them cannot meet the demand.

The strong demand for bibliometric evaluations that contribute to the disappearance of modalities is at least partly caused by bibliometrics itself. In a dynamics that has been described by Rip as the "promise-requirement cycle" (Rip 1997: 628-632), bibliometrics has promised methods that can be used to evaluate research performance (e.g. Garfield 1979), a promise that has been turned into a performance expectation by science policy. Bibliometrics thus confronts the ghosts it called.

Our discussion of the conditions needed for evaluations explains why bibliometric evaluations in general are growing so popular.[7] They are valid in certain areas and under certain conditions; they give the impression of being easy to understand because they come as numbers; they are actively marketed by a commercial enterprise; professional bibliometricians depend on conducting them; and amateur bibliometricians are spreading the message without caring about risks and side-effects. We have also made it clear why in this process modalities of the methods are dropped. All the actors involved implicitly contribute to the construction of bibliometric evaluations as a universal tool that measures research quality, even though none of them would describe them as such. The whole – the use of bibliometric techniques as a universal, modality-free evaluation tool – is certainly more than the sum of its parts, i.e. of the numerous individual applications of bibliometric methods and discussions about them. The 'subtracted value' in form of missing modalities is the result of a complex construction process whose elements we have traced in this article.

What can be done? It is difficult to imagine a way in which the urge to evaluate could be contained. Therefore, it becomes important to make clear to everyone that appropriate bibliometric evaluations are not cheap, and to discuss the use that is made of them. Apart from that, two major steps suggest themselves. First, science policy should support the *professionalisation of bibliometric evaluations* by securing the independence of the profession from both its source of data and its customers, and by strengthening 'disinterested' academic work on the qualification of bibliometric methods. Systematic training of bibliometricians as well as some dedicated 'watchdog activities' could be expected in return. The latter should include the introduction of a code of ethics and quality control guidelines and mechanisms. Secondly, the creation of a public citation database would enable the quality control of data and thus help to overcome many of the described problems. Both steps require major policy efforts, international coordination, and investments. They would therefore only be undertaken if detrimental effects of 'quick and dirty' bibliometric evaluations make themselves felt, e.g. when powerful actors in academia or the public lose their trust in science policy because of invalid evaluations. Since the validity of all research evaluation exercises is inherently limited, and since excellent research is 'robust' enough to come out top in most

[7] In this article, we have focused on bibliometric methods. The situation concerning the second major quantitative evaluation method – calculating external grant income – is even worse. No systematic research on this indicator has been conducted, and we are just learning about its modalities (Laudel 2005). For very much the same reasons as described in section 4, the indicator is nevertheless widely applied by science policy and management.

evaluations, it is unlikely that the problems of misconstrued bibliometric evaluations will ever become pressing enough to warrant major investments.

This means that we will continue to live with the status quo for a longer time. But we should not despair! In all of its history, the science system as a whole has proven to be very adaptable to all sorts of external changes. It will be interesting to see how it adapts to bibliometric evaluations.

REFERENCES

Adam, David (2002), 'The counting house', *Nature*, 415, 726-729.

Baldi, Stephane (1998), 'Normative Versus Social Constructivist Processes in the Allocation of Citations: A Network-Analytic Model', *American Sociological Review*, 63, 829-846.

Barnes, Barry (1977), *Interests and the Growth of Knowledge*, London: Routledge & Kegan Paul.

Barre, R. (1994), 'Do Not Look for Scapegoats - Link Bibliometrics to Social Sciences and Address Societal Needs', *Scientometrics*, 30, 419-424.

Bookstein, A. (1994), 'Scientometrics: new opportunities', *Scientometrics*, 30, 455-460.

Brickley, Peg (2002), 'A Scrap over Sequences, Take Two', *The Scientist*, 16, 13 May 2002.

Butler, Linda (1999), 'Who 'Owns' this Publication? Problems with assigning research publications on the basis of addresses', *Proceedings of the Seventh Conference of the International Society for Scientometrics and Informetrics, Colima, México, July 5-8, 1999,* pp. 87-96.

Butler, Linda (2001), *Monitoring Australia's Scientific Research: Partial indicators of Australia's research performance,* Canberra: Australian Academy of Science.

Butler, Linda and Martijn S. Visser (2006), 'Extending citation analysis to non-source items', *Scientometrics*, 66, 327-343.

Cameron, Brian D. (2005), 'Trends in the Usage of ISI Bibliometric Data: Uses, Abuses, and Implications', *portal: Libraries and the Academy*, 5, 105-125.

Cole, Jonathan R. and Stephen Cole (1972), 'The Ortega Hypothesis', *Science*, 178, 368-375.

Cole, Jonathan R. and Stephen Cole (1973), *Social Stratification in Science,* Chicago: The University of Chicago Press.

Cole, Stephen and Jonathan R. Cole (1967), 'Scientific Output and Recognition, a Study in the Operation of the Reward System in Science', *American Sociological Review*, 32, 377-390.

Cozzens, Susan E. (1985), 'Comparing the Sciences: Citation Context Analysis of Papers from Neuropharmacology and the Sociology of Science', *Social Studies of Science*, 15, 127-153.

Cozzens, Susan E. (1989), 'What Do Citations Count? The Rhetoric-First Model', *Scientometrics*, 15, 437-447.

Daniel, Hans-Dieter (2006), 'The referee's comments on Gläser's criticism', *Scientometrics*, 67, 331-333.

DFG [Deutsche Forschungsgemeinschaft] (2004), *Empfehlungen zu einer 'Leistungsorientierten Mittelvergabe' (LOM) an den Medizinischen Fakultäten - Stellungnahme der Senatskommission für Klinische Forschung der Deutschen Forschungsgemeinschaft*, Bonn: DFG.

Dieks, Dennis and Hans Chang (1976), 'Differences in impact of scientific publications: Some indices derived from a citation analysis', *Social Studies of Science*, 6, 247-267.

Garfield, Eugene (1979), 'Is Citation Analysis a Legitimate Evaluation Tool?', *Scientometrics*, 1, 359-375.

Gilbert, G. Nigel and Steve Woolgar (1974), 'Essay Review The Quantitative Study of Science : An Examination of the Literature', *Science Studies*, 4, 279-294.

Gilbert, G. Nigel (1977), 'Referencing as Persuasion', *Social Studies of Science*, 7, 113-122.

Glänzel, Wolfgang (1996), 'The need for standards in bibliometric research and technology', *Scientometrics*, 35, 167-176.

Glänzel, Wolfgang, Sylvan Katz, Henk Moed and Urs Schoepflin (1996), 'Preface', *Scientometrics*, 35, 165-166.

Glänzel, Wolfgang and Henk F. Moed (2002), 'Journal impact measures in bibliometric research', *Scientometrics*, 53, 171-193.

Glänzel, Wolfgang and Urs Schoepflin (1994), 'Little Scientometrics, Big Scientometrics ... And Beyond', *Scientometrics*, 30, 375-384.

Gläser, Jochen (2001), 'Scientific specialties as the (currently missing) link between scientometrics and the sociology of science', in Mari Davis and Concepción S. Wilson (eds.), *Proceedings of the 8th International Conference on Scientometrics & Informetrics,* Sydney, Australia, July 16-20th, BIRG, UNSW, pp. 191-210.

Gläser, Jochen (2006), 'Letter to the Editor', *Scientometrics*, 67, 327-329.

Gläser, Jochen and Grit Laudel (2005), 'Advantages and dangers of 'remote' peer evaluation', *Research Evaluation*, 14, 186-198.

Godin, Benoit (2002), *The Social Sciences in Canada: What Can We Learn From Bibliometrics?' Project on the Measurement of the Social Sciences*, Working Paper no. 1, Quebec: INRS.

Griffith, Belver C. (1994), 'Little Scientometrics, Little Scientometrics, Little Scientometrics, Little Scientometrics ... And So on and So On', *Scientometrics*, 30, 487-493.

Henkel, Mary (2000), *Academic Identities and Policy Change in Higher Education*, London: Jessica Kingsley.

Hicks, Diana (1999), 'The difficulty of achieving full coverage of international social science literature and the bibliometric consequences', *Scientometrics*, 44, 193-215.

James, Oliver (2001), 'Business Models and the Transfer of Businesslike Central Government Agencies', *Governance: An International Journal of Governance and Administration*, 14, 233-252.

Joerges, Bernward and Terry Shinn (2001), 'A Fresh Look at Instrumentation: An Introduction', in Bernward Joerges and Terry Shinn (eds.), *Instrumentation Between Science, State and Industry*, Dordrecht: Kluwer, pp. 1-13.

Katz, J. Sylvan (1999), *Bibliometric Indicators and the Social Sciences*, Sussex: SPRU.

KNAW [Koninklijke Nederlandse Akademie van Wetenschappen] (2005), *Judging research on its merits: An advisory report by the council for the Humanities and the Social Sciences Council,* Amsterdam: Royal Netherlands Academy of Arts and Sciences.

Knorr-Cetina, Karin (1981), *The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science,* Oxford: Pergamon Press.

Latour, Bruno (1987), *Science in Action*. Cambridge, Mass: Harvard University Press.

Latour, Bruno and Steve Woolgar (1986 [1979]), *Laboratory Life: The Construction of Scientific Facts*, Princeton: Princeton University Press.

Laudel, Grit (2005), 'Is external funding a valid indicator for research performance?' *Research Evaluation*, 14, 27-34.

Leydesdorff, Loet (1989) 'The Relations between Qualitative Theory and Scientometric Methods in Science and Technology Studies: Introduction to the Topical Issue', *Scientometrics*, 15, 333-347.

Leydesdorff, Loet and Peter van den Besselaar (1997), 'Scientometrics and communication theory: Towards theoretically informed indicators', *Scientometrics*, 38, 155-174.

Leydesdorff, Loet and Paul Wouters (1996), 'Quantitative Measuring or Qualitative Understanding: Is it possible to bridge the divide in STS?' *EASST Review*, 15, 20-24.

Liu, Nian Cai, Ying Cheng and Li Liu (2005), 'Academic ranking of world universities using scientometrics - A comment to the "Fatal Attraction"', *Scientometrics*, 64, 101-109.

Luukkonen, Terttu (1994), 'Are We Longing for the Golden Era Lost or for the One to Come', *Scientometrics*, 30, 481-485.

MacRoberts, Michael H. and Barbara R. MacRoberts (1989), 'Problems of Citation Analysis: A Critical Review', *Journal of the American Society for Information Science*, 40, 342-349.

Marginson, Simon and Mark Considine (2000), *The Enterprise University: Power, Governance and Reinvention in Australia,* Cambridge, UK: Cambridge University Press.

Marshall, Eliot (1997), 'Snipping Away at Genome Patenting', *Science*, 277, 1752-1753.

Martin, Ben R. and John Irvine (1983), 'Assessing Basic Research. Some Partial Indicators of Scientific Progress in Radio Astronomy', *Research Policy*, 12, 61-90.

Meho, Lokman I. and Kristina M. Spurgin (2005), 'Ranking the Productivity of LIS Faculty and Schools: An Evaluation of Data Sources and Research Methods', *Journal of the American Society for Information Science and Technology*, 56, 1314-1331.

Miquel, J.F. (1994), 'Little scientometrics, big scientometrics ... and beyond', *Scientometrics*, 30, 443-445.

Moed, Henk F. (2002), 'The impact-factors debate: the ISI's uses and limits', *Nature*, 415, 731-732.

Moed, Henk F., W.J.M. Burger, J.G. Frankfort and A.F.J. Van Raan (1985a), 'The Use of Bibliometric Data for the Measurement of University Research Performance', *Research Policy*, 14, 131-149.

Moed, Henk F., J.M. Burger, J.G. Frankfort and Anthony F. J. Van Raan (1985b), 'The application of bibliometric indicators: important field- and time-dependent factors to be considered', *Scientometrics*, 8, 177-203.

Moed, Henk F., R.E. De Bruin and Thed N. Van Leeuwen (1995), 'New bibliometric tools for the assessment of national research performance: Database description, overview of indicators and first applications', *Scientometrics*, 33, 381-422.

Moed, Henk F. and Eugene Garfield (2004), 'In basic science the percentage of 'authoritative' references decreases as bibliographies become shorter', *Scientometrics*, 60, 295-303.

Morris, Norma (2002), 'The developing role of departments', *Research Policy*, 31, 817-833.

Nederhof, A. J. (1988), 'The validity and reliability of evaluation of scholarly performance', in Anthony J. F. van Raan (ed.), *Handbook of Quantitative Studies of Science and Technology,* North-Holland: Elsevier, pp.193-228.

Nederhof, Anton J. (2006), 'Bibliometric monitoring of research performance in the Social Sciences and the Humanities: A Review', *Scientometrics*, 66, 81-100.

Phillimore, A.J. (1989), 'University Research Performance Indicators in Practice: The University Grants Committee's Evaluation of British Universitites, 1985-86', *Research Policy*, 18, 255-271.

Pickering, Andrew (1982), 'Interests and analogies', in Barry Barnes and David Edge (eds.), *Science in Context. Readings in the Sociology of Science,* Milton Keynes: The Open University Press, pp. 125-146.

Rip, Arie (1997), 'A Cognitive Approach to Relevance of Science', *Social Science Information*, 36, 615-640.

Russell, Jane M. (1994), 'Back to the Future for Informetrics', *Scientometrics*, 30, 407-410.

Schimank, Uwe (2004), 'Leistungsbeurteilung von Kollegen als Politikberatung', in Thomas Brüsemeister (ed.), *Die beratene Gesellschaft: Zur gesellschaftlichen Bedeutung von Beratung*, Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 39-55.

Schimank, Uwe (2005), ''New Public Management' and the academic profession: Reflections on the German situation', *Minerva*, 43, 361-376.

Schoepflin, Urs and Wolfgang Glänzel (2001), 'Two decades of "Scientometrics": An interdisciplinary field represented by its leading journal', *Scientometrics*, 50, 301-312.

Slaughter, Sheila and Larry L. Leslie (1997), *Academic Capitalism: Politics, Policies, and the Entrepeneurial University,* Baltimore: The Johns Hopkins University Press.

Smith, Linda C. (1981), 'Citation Analysis', *Library Trends*, 30, 83-106.

Sternberg, R. and T. Litzenberger (2005), 'The publication and citation output of German faculties and disciplines based upon SSCI data', *Scientometrics*, 65, 29-53.

Thomas, William I. and Dorothy Swaine Thomas (1928), *The Child in America: Behavior Problems and Programs*, New York: Knopf.

Van den Besselaar, Peter (2000), 'Communication between science and technology studies journals: A case study in differentiation and integration in scientific fields', *Scientometrics*, 47, 169-193.

Van den Besselaar, Peter (2001), 'The cognitive and social structure of STS', *Scientometrics*, 51, 441-460.

Van Leeuwen, Thed N. (2006), 'The application of bibliometric analyses in the evaluation of social science research. Who benefits from it, and why it is still feasible', *Scientometrics*, 66, 133-154.

Van Raan, Anthony F. J. (1994), 'Little scientometrics, big scientometrics ... and beyond', *Scientometrics*, 30, 529-531.

Van Raan, Anthony F. J. (1996), 'Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises', *Scientometrics*, 36, 397-420.

Van Raan, Anthony F. J. (1998), 'In matters of quantitative studies of science the fault of theorists is offering too little and asking too much - Comments on: Theories of citation?', *Scientometrics*, 43, 129-139.

Van Raan, Anthony F. J.(2000) 'The Pandora's Box of Citation Analysis: Measuring Scientific Excellence - The Last Evil?' in Blaise Cronin and Helen Barsky Atkins (eds.) *The Web of Knowledge,* Medford, New Jersey: Information Today Inc., pp. 301-319.

Van Raan, Anthony F. J. (2005a), 'Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods', *Scientometrics*, 62, 133-143.

Van Raan, Anthony F. J. (2005b), 'Reply to the comments of Liu et al', *Scientometrics*, 64, 111-112.

Van Raan, Anthony F. J. and Thed N. Van Leeuwen (2002), 'Assessment of the scientific basis of interdisciplinary, applied research - Application of bibliometric methods in Nutrition and Food Research', *Research Policy*, 31, 611-632.

Vinkler, P. (1994), 'Words and indicators. As scientometrics stands', *Scientometrics*, 30.

Vinkler, P. (1996), 'Some practical aspects of the standardization of scientometric indicators', *Scientometrics*, 35, 237-245.

Wade, Nicholas (1975), 'Citation Analysis: A New Tool for Science Administrators', *Science*, 188, 429-432.

Weingart, Peter (2005), 'Impact of bibliometrics upon the science system: Inadvertent consequences?', *Scientometrics*, 62, 117-131.

Weingart, Peter, R. Sehringer and M. Winterhager (1988), 'Bibliometric indicators for assessing strength and weaknesses of West German science', in Anthony J. F. van Raan (ed.), *Handbook of Quantitative Studies of Science and Technology,* North-Holland: Elsevier, pp. 391-430.

White, Howard D. (2004), 'Reward, persuasion, and the Sokal Hoax: A study in citation identities', *Scientometrics*, 60, 93-120.

Woolgar, Steve (1991), 'Beyond the Citation Debate: Towards a Sociology of Measurement Technologies and Their Use in Science Policy', *Science and Public Policy*, 18, 319-326.

Weak and Strong Research Evaluation Systems

CHAPTER 6

JOCHEN GLÄSER AND GRIT LAUDEL

EVALUATION WITHOUT EVALUATORS

The Impact of Funding Formulae on Australian University Research

THE ELUSIVE EFFECTS OF RESEARCH EVALUATION SYSTEMS

The Australian research evaluation system (RES) is unique in its exclusive reliance on a funding formula. For each university, statistics on income from competitive research grants, numbers of publications, numbers of current research students (Masters and PhD students), and timely completions of Masters and PhD studies are collected and used to calculate the allocation of state funds without any further consideration. The research component of university funding has developed into an allocation of three national competitive grants, including the Institutional Grants Scheme (IGS), Research Training Scheme (RTS), and Research Infrastructure Block Grant (RIBG), which use the indicators with different weightings as shown in table 1.

Table 1. Money distributed and indicators used in the three competitive grants in 2004
Source: Nelson 2005: 38, 75-76

Indicators used in funding formulae	*Competitive research block grants (Mio AU$) and weight of indicators*			*Share of research funding controlled by the indicator Mio AU$ (%)*
	RTS (552.153)	*IGS (290.591)*	*RIBG (182.982)*	
Research income from competitive grants	40%	60%	100%	547.774 (55.5)
Successful research student completions	50%			270.490 (27.4)
Number of higher degree research students places		30%		85.384 (8.7)
Research publications	10%	10%		82.559 (8.4)
Total amount of research block funding				986.207

R. Whitley and J. Gläser (eds.), The Changing Governance of the Sciences, 127–151.

The Australian system is a strong (standardised, public, transparent, and consequential) RES (see Whitley, this volume). It is highly standardised in that it is applied in exactly the same way to all sciences, social sciences, arts, and humanities. The same unified lists of grants that count as 'research income' and the same categorisation of 'research publications' (articles in refereed journals, books, book chapters and full refereed conference papers) are applied to all fields. The funding mechanism is also highly transparent because most of the information is on public record. The amount of money received by each university – and therefore the actual performance according to the measures – is on public record, too. Because of the formula, academics knew that each of their articles in a peer-reviewed journal in 2004 contributed about 2058 AU$ to their university's income, and that a book contributed five times as much. The consequences of the scheme for funding are significant. In 2004, the three research block grants amounted to 24% of the block funding for teaching and research allocated to universities.

This system has been in place, with variations and additions, for about 15 years.[1] Universities, university departments, and academics can be expected to have adapted to it, which inevitably raises the question about the effects of the system: has it achieved its purpose, namely increasing good performance by rewarding it and by automatically channelling money to the best performers? Has the adaptation of the affected actors produced non-intended side effects that are detrimental to the science system?

These questions are difficult to answer, and indeed have not been satisfyingly answered for any RES. While the political discussion about RES is laden with statements about their achievement of intended effects - improved quality - and about their negative side effects, none of these claims is backed by reliable empirical evidence. Attempts to establish positive or negative effects of RES face the two problems of identifying these effects and of causally attributing them to the RES. Identification is difficult because the intended and non-intended effects are changes in epistemic features of research processes such as quality, uncertainty, or interdisciplinarity, which poses serious measurement problems. Causal attribution is hindered by the overlap of RES by a multitude of other institutional influences on scientific research. Arguments that try to establish positive effects of RES can easily be countered by references to countries that show similar improvements without having any RES in place. Conversely, purported negative effects of RES can be ascribed to other conditions such as a general lack of money, decreasing success rates of grant funding, or government policies tying resource allocation to application-oriented research.

[1] Up to 2008. The formula-based RES will be replaced by a different procedure in the near future.

The aim of this chapter is to demonstrate that, in spite of the methodological problems, epistemic effects of RES can be identified and ascribed. We will use data from a current empirical study of Australian university research to demonstrate that the Australian RES has little direct steering effect in itself but does contribute to a general shortage of recurrent funding, and to a strong dependence on a small number of external sources, a situation that indeed does change research.[2]

APPROACH

Research Strategy – the Challenge of Causal Attribution

In order to solve the problem of causal ascription, we must identify the social mechanisms that link RES to changes in the conduct and content of research. Following Mayntz (2004: 241) we define a social mechanism as *a sequence of causally linked events that occur repeatedly in reality if certain conditions are given and link specified initial conditions to a specific outcome* (for similar but less precise definitions see Merton 1968: 42-43 and Hedström 2005: 11). Identifying social mechanisms means that we can demonstrate *how* a specific cause – an RES - produces changes in university research and thus causally attribute these changes to RES. The concept of mechanisms has first been introduced to science studies in Whitley's (1972) criticism of the Mertonian sociology of science, but has never been taken up.

The identification of the social mechanisms that link RES to changes in the production of scientific knowledge must bridge a rift that has opened in science studies during the last three decades. The initial conditions and specified outcomes of these mechanisms are investigated by different strands of science studies that have grown so far apart that today they seem almost incommensurable. Much science policy research has focused on changes in the governance of science, including funding policies (Braun 1993; Ruivo 1994; Guston 1996; van der Meulen 1998; Silvani, Sirilli, and Tuzi 2005). While political actor constellations and policies are identified as independent variables, the studies are less clear about the effects they investigate. For example, it is sometimes hypothesised that the changing role of funding councils will affect the cognitive content of science, but these effects are either not described at all or else only at a very general level without empirical backing (e.g. Rip 1994; Braun 1998). Mayntz and Schimank have argued that in order to understand the mechanisms that channel external expectations towards science, the "performance level of the science system" needs to be included in the analysis (Mayntz and Schimank 1998, p.753). So far, this has rarely been done (notable exceptions are Van der Meulen and Leydesdorff 1991; Morris 2000).

This disregard for changes in scientific knowledge by much work on science policy is complemented by a tendency to ignore the role of institutions by constructivist sociology of science. After its constructivist turn, the sociology of science developed a strong interest in scientific knowledge as a dependent variable and the

[2] Funding of this project by the German Federal Ministry of Education and Research in its programme 'science policy studies' and by the Australian Research Council is gratefully acknowledged.

micro-focus in this tradition produced many accounts of researchers' adaptations to local circumstances and of the consequences of these adaptations for the practices of knowledge production and the content of knowledge (e.g. Knorr-Cetina 1981; Lynch 1985; Latour and Woolgar 1986). However, the microscopic focus of these accounts favoured the production of individual single-case studies and hindered more comparative approaches. Additionally, the detail with which many single instances of knowledge production were studied resulted in the neglect of macro-structures and dominant institutions (Knorr-Cetina 1995: 160-163; Kleinman 1998: 285-291; Mayntz and Schimank 1998: 751).

For the causal mechanisms linking changes in state-science relationships to knowledge production to be identified, it is necessary to integrate science policy studies of RES as initial conditions with the constructivist studies of the conduct and content of research as specified outcomes. This synthesis can build on the constructivist insight that researchers opportunistically adapt their practices of knowledge production to their local situation, and the insight from science policy studies that the institutions of the national science system co-produce these situations by determining power relations and access to resources.

Systems of evaluation are specific institutions, i.e. systems of formal and informal rules that govern actions.[3] They merge political institutions (of analysing, reporting, and decision making) with institutions governing evaluative practices of scientific communities. The use of elements of the knowledge production process in RES links political actions to the conduct and content of research. Since the features of research that are used to measure quality are inextricably linked to other epistemic characteristics of that research, the adaptation of research strategies and approaches to the 'quality expectation' is likely to change more than research 'quality' as measured by the system. This is why unintended epistemic effects of RES are observed. In order to identify the social mechanisms that change the quality and other epistemic features of research, we must ascertain the impact of the Australian RES on the situations in which academics at Australian universities conduct research, the ways in which academics adapt to these situations, and the resulting changes in knowledge.

The main channel through which RES influence research is the money distributed to universities on the basis of performance measurements. This money makes RES an important element of the resource environment of universities. Universities can be expected to adapt to, as well as attempt to influence, their environments. While they respond not only to the RES but also to their whole environment, we can expect them to develop strategies for maximising their income from the RES by increasing performance as measured by the RES. This adaptation of universities to

[3] Contrary to the mainstream of the sociological 'new institutionalism' (Powell and DiMaggio 1991) we maintain that the concept 'institution' should not be extended to collective beliefs and frames, but is better reserved for systems of formal and informal rules (North 1990; Mayntz and Scharpf 1995; Scharpf 1997). This restricted concept has the theoretical advantage of enabling a separate treatment of qualitatively different social phenomena (rules and belief systems), thus supporting the distinction between systems of rules and actors' perceptions of these rules. The methodological advantage of a restricted concept is that it can be better linked to specific empirical strategies for investigating phenomena. The search for rules requires other empirical strategies than the search for belief systems.

RES is a first set of mechanisms that needs to be identified. The mechanisms of adaptation at the university level are 'remote' insofar they not affect research directly, but co-produce the situations of academics in the universities.

The direction and conduct of research are affected directly by a second set of mechanisms, namely the adaptation of researchers to their situation by making decisions on research. Laboratory studies have established that knowledge production consists of a stream of decisions in which researchers adapt to the local contingencies of their laboratories, rules and actor constellations in their organisation, funding opportunities, and societal expectations (e.g. Knorr-Cetina 1981). We may add that they also adapt to existing knowledge and to the dominant institutions of their scientific communities, a fact that is often neglected in constructivist studies. Some of the decisions of researchers are strategic in nature because they constrain further choices, thus creating a path dependency of research. The most important strategic decisions are the selection of research problems, objects and methods, and of partners for collaborations. They are often, but not always tied to explicit decisions on new projects or grant applications. The mechanisms that are at work in these strategic decisions are the 'proxy' mechanisms mediating the impact of RES on the conduct and content of research.

Thus, RES can be expected to influence research by triggering adaptive mechanisms at the university level, which in turn change the conditions for individual

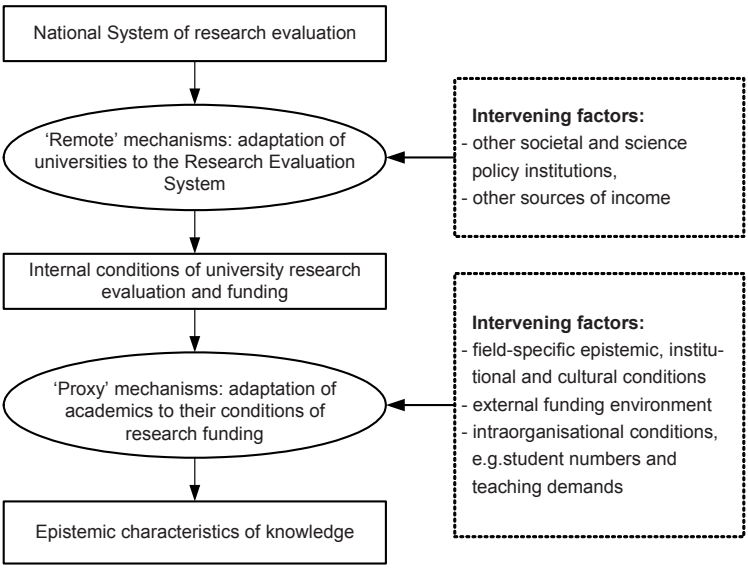

Figure 1. Assumptions about Causal Relationships between the National Research Evaluation System and Knowledge Production

academics and trigger adaptive behaviour at the individual level as summarised in figure 1. A second, more direct but simultaneously weaker impact of RES occurs because institutions may "bypass" the organisational level and directly affect members of the organisation (Scott 1991: 180-181). Academics perceive the performance expectations built into RES and interpret them as signals of what society values about their research. They may adapt to these perceptions regardless of changes in their material conditions for research.

Establishing the possible effects of these mechanisms is the most difficult part of the whole investigation because the sociology of science provides little theoretical guidance as to the nature of the key epistemic properties of research, and equally little methodological support for the empirical investigation of these properties. The following list of properties of research to be included derives from discussions of theory structures (Nagi and Corwin 1972; Whitley 1977; Rip 1982), theoretical attempts to conceptualise paradigmatic maturity (Böhme et al. 1973, 1983) or to apply the contingency approach of organisational sociology to science (Whitley 1984), and from the various warnings about negative effects of RES (Gläser et al. 2002):

- Type of research (for example, methodological, theoretical, experimental or field research) and its dominant orientation (basic, strategic, or applied);
- Relationship to the community's majority opinion (non-conformist versus mainstream);
- Time characteristics of research (long-term versus short-term processes);
- The degree of heterogeneity of knowledge combined in the research (usually referred to as 'interdisciplinarity');
- The degree of intellectual risk taken in the research; and
- Reliability of results.

Since the responses of academics to their institutional conditions of their work can be assumed to depend on characteristics of their field, epistemic properties are not only dependent variables but also intervene in the decisions of researchers. For example, the intention to conduct a project that requires long-term observations might affect the sources of funding addressed, and the fact that a certain line of research is capital-intensive might move researchers to look for collaborators.

Methodology and Methods – the Challenge of Empirical Identification

In order to identify causal mechanisms, variations of causes and effects need to be observed. In order to obtain such variations, we compare Australian universities that are subject to the same national RES but have implemented specific internal systems of evaluation and funding. The overlap of a uniform RES with varying intraorganisational institutions creates differences that can be used to analyse both variations in local conditions and commonalities of universities. The commonalities are likely to reveal national institutional conditions that 'reach through', i.e. affect academics regardless of university-specific adaptations.

The selection of cases for this study was prepared by a comparative analysis of Australian universities, which took into account their position in the highly stratified Australian university system, their research intensity, trends in publication behaviour, and the fields represented in the universities' research profiles. We followed Marginson's (2006: 11) distinction of five groups of universities as shown in table 2.[4]

Table 2. Stratification of Australian universities in terms of research (Source: Marginson 2006: 11 and own calculations based on Nelson 2005: 76-77)

Segment	*Number of universities*	*Share in research block grants in 2005 (%)*
'Sandstones' or 'Group of 8'	8	63.6
'Gumtrees'	11	21.2
'Unitechs'	5	8.3
'New Universities'	12	6.4
Other	3	0.5

A total of seven universities were selected, three from the 'Group of 8' and two each from the 'Gumtrees' and 'Unitechs'. The universities from the other groups did not show a sufficient amount of research and lacked too many of the fields included in our investigation to enable comparisons.

In a second step, the Australian RES will be compared to the German system, which is only now introducing evaluations and therefore can still be regarded as representing the 'ground state' of research that is not influenced by RES (see Lange, this volume). The case of the German state of Lower Saxony, which has conducted peer reviews of all of its university research (Schiene and Schimank, this volume), is an exception in the German context.

The expectation that the impact of RES varies with field specific needs for funding, time characteristics of research, publication practices, etc. (see Whitley, this volume) suggests epistemic differences between scientific fields as a second dimension for comparison. The selection of fields was more difficult because no empirically confirmed systematic description of fields by epistemic characteristics exists. In order to achieve sufficient variation of key epistemic properties listed above, we selected four fields from the natural sciences and one each from the social sciences and humanities. The fields investigated in our study are mathematics; biochemistry; physics; geology; history; and political science.

In this paper we report on the preliminary analysis of data comparing changes in research direction and conduct in two research-intensive universities (U1 and U2)

[4] The 'Sandstones', or 'Group of 8' are most of the older foundations except the Universities of Tasmania and New England. The 'Gumtrees' include all other universities established in each state prior to the higher education reforms that began in 1987. 'Unitechs' are former Institutes of Technology that became Universities in the higher education reform of 1987. The 'New Universities' were created from 'Colleges of Advanced Education' in the same reform. Others include private universities and a few small higher education providers. See Marginson and Considine (2000: 175-232) for a discussion of the segments.

and five fields, namely history, political science, biochemistry, mathematics, and geology. Our data collection combined analysis of documents and internet sites, bibliometric analyses, and qualitative interviews as the core method of case studies. Interviews with university managers were conducted in order to collect data on the perception of funding conditions by universities and their responses. The interviews focused on:

- managers' perception of the funding conditions for their university,
- the research funding schemes currently in place within the university,
- the impact of the RES and internal funding schemes on the core functions of the university (teaching and research), and
- university strategies for the internal governance of research, with special emphasis on performance evaluation schemes for organisational units and academic staff that are currently in place.

We prepared for the interviews with analyses of financial data and strategies of universities, and of internal funding schemes for research, which could be downloaded from the internet.

The interviews with academics were designed to produce information on the interviewees' perceptions of their research biographies and plans as well as their working conditions, with a particular emphasis on resources for research and on performance evaluation. We conducted detailed bibliometric analyses of the interviewees' publications in order to identify topical changes, trends in publication behaviour, and the researchers' international visibility. 'Bibliometric research trails', such as those portrayed in figure 2, were constructed and presented as a basis for discussion in the interview. In the fields with insufficient coverage of publications by the databases of Thomson Scientific (see Gläser and Laudel in this volume on bibliometrics), publication lists were retrieved by internet search, and the network was constructed on the basis of similarities in title keywords.

Based on this reconstruction of previous research, the first part of an interview covered the following main aspects of the research and its epistemic characteristics:

- Research projects conducted since the interviewee joined the university (the researchers' research trails, Chubin and Connolly 1982) and their epistemic characteristics;
- Reasons for abandoning certain topics and following others, especially the extent to which this behaviour is triggered by funding considerations;
- National versus international character of the subject area; and
- Practices of national and international collaboration.
- Publication strategies, especially audiences that are targeted by publications and criteria for selection of journals.

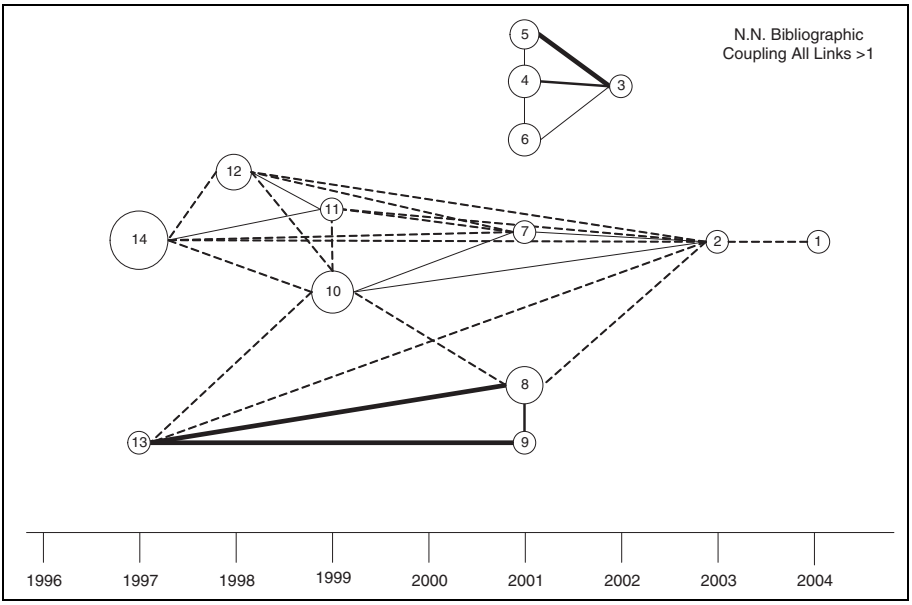

Figure 2. Example of a bibliometric research trail (strength of lines indicates topical proximity, size of circles indicates numbers of citations)

In a second part of the interview, one project that had recently been started was selected for a more detailed analysis. In this part, questions about strategic decisions were asked, namely about the selection of the problem, object, methods, collaborators, and communication channels, and about the reasons for making these selections. A further set of questions referred to funding needs and the way in which funding was obtained for this project including application procedures and selection criteria applied to the researcher's proposals. A third part of the interview was devoted to the interviewee's working conditions. We explored the availability of time for research, teaching loads, performance evaluations, the general degree of autonomy, and external pressures perceived by the researcher.

The interviews took sixty to ninety minutes in the case of researchers and around sixty minutes in the case of managers. With few exceptions, they were tape-recorded and subsequently transcribed. This paper is based on 32 interviews with academics from the five fields and 21 interviews with managers from all levels of the university hierarchy. Data was analysed by computer-guided qualitative content analysis (Gläser and Laudel 2006) that used the variables described in the previous section to extract information from the interview transcription.

ADAPTATIONS TO A STRONG RESEARCH EVALUATION SYSTEM

The Situation of Universities

Australian universities are formally autonomous. They have authority over the content and conduct of teaching and research, their internal structure and internal allocation of money. Creating positions and filling them is at the discretion of the university. Salaries are regulated by local workplace bargaining between the university and the union, and can be topped up by universities in individual cases.

The internal governance of Australian universities is characterised by strong hierarchies. Principal strategic decisions on the academic development of the university are made by the senate, which includes major stakeholders from industry and the community, and is headed by the university chancellor. The vice chancellor is the university's chief executive officer and responsible for university governance. He usually has several Deputy Vice Chancellors (DVCs), one of whom is responsible for all of the research of the university. The major subunits of the bigger universities are faculties and schools within faculties. The hierarchy is complemented by a system of committees that support decision making on all major issues within the university – the academic board advising the vice chancellor, research committees at university, faculty, and school levels, central and faculty promotion committees, and so on. Some of these committees are granted control over funds (e.g. over research grants, see below). The degree of academic self-governance is quite limited. Faculty meetings and staff meetings of schools still take place, but their influence on decisions of managers depends on the discretion of the latter.

The autonomy of Australian universities is limited by the fact that they financially depend on the federal government, whose financial support does not match increases in student numbers.[5] Therefore, universities have little choice but to access all existing sources of funding. Exploiting this dependence, the federal government offers additional funding for universities that fulfil specific demands. For example, the provision of an increase in the operating grants of 5% in 2006 and 7.5% in later years depends on universities' compliance with the "Higher Education Workplace Requirements", which mainly prescribe the offer of individual workplace agreements beside the collective agreements with unions (DEST 2005b). Thus the visible hand 'cannot let go' in Australia either (see Engwall and Nybom, this volume, for the case of Sweden) and interferes with internal matters of universities.

The Australian research funding environment is neither rich nor diverse and does not provide significant alternative funding sources for universities. A major problem is the structure of the Australian economy, especially the weakness of science-based industries; accordingly contributions by industry to the funding of research are low. In 2002, for instance, the share of business in the gross domestic expenditure for research and development was only 51.2% (compared to 64.3% for the EU-15 and 67.8% for the OECD, DEST 2005a: 25). Many researchers cannot find industry

[5] The share of federal government funding of higher education decreased from 57% in 1996 to 41% in 2004. The absolute amount of funding increased by 16% (in actual prices, AVCC 2005), while the number of domestic students rose by 23% in the same period (Nelson 2005: 18).

partners in Australia in fields where this would be possible in most industrialised countries. Furthermore, support from the states is limited to ad hoc-funding in designated fields.

Consequently, the national research councils are the only significant sources of research funding for many fields. These are the National Health and Medical Research Council (NHMRC) that funds health science, biomedical, and clinical research; and the Australian Research Council (ARC) that funds research in all other fields including basic biosciences. Both agencies administer several funding programmes for individual projects in the responsive mode, for collaborative research with industry or between research groups, and for international collaboration.

The chances of getting a grant from a research council significantly vary between funding schemes. In 2005, 'Discovery grants' (individual project grants funded by the ARC in the responsive mode) had a success rate of about 30%, while 'Linkage grants' (collaborative projects with industry) had a success rate of 46% (ARC 2005). In decisions about discovery grants the applicant's track record is the single most important criterion and is weighted at 40%. Hence the applicant's publication list and record of prior grants are decisive for the success of an application. In some fields, applicants are expected to submit numbers of citations or the impact factors of journals in which they publish as additional information on the 'track record' (see Gläser and Laudel, this volume, on the use of these measures).

The project funding in the responsive mode is inflexible because there is only one round of decisions each year, and it takes nine months from the application to the start of funding. If they get an idea for a project at the wrong time, researchers have to wait up to one and a half years for funding. In some funding schemes, the councils provide only part of the necessary money, and industry partners or universities must co-fund projects or equipment.

The ARC and NHMRC are subordinated to the Ministry of Education, Science and Training and to the Ministry of Health, respectively and ministers need to approve the funding decisions made by the councils. The Minister of education, science and training recently exercised this right by preventing several projects that had passed the ARC's peer review from being funded. Apart from this direct political intervention, there are several ways in which a political direction of research is implemented through funding by research councils. An orientation towards applied research is achieved by the significantly higher success rate of collaborative grants with industry partners. Discovery grants do not depend on collaborations with industry but still require applicants to describe the 'national benefit' of the research, which is weighted 10% in the evaluation process. The government has also defined 'national research priorities' and expects the research councils to give grant applications in these areas priority. At a more subtle level, there is a widespread perception that government and research councils prefer applied and 'hot' topics. This perception has been voiced by researchers from all five disciplines. It cannot be dismissed as a rationalisation of failure because holders of grants and scientists who worked as reviewers held the same view.[6]

[6] For a more extensive discussion of the Australian grant funding system, see Laudel (2006).

Adaptation to the Research Funding Environment

The universities respond to their conditions of research funding by trying to maximise their income from the funding formula in general and from external grants in particular. We could identify three mechanisms at the university level.

A first mechanism is *mimesis*. As in all other Australian universities, the two universities analysed here allocate a significant amount of the resources according to an internal formula that mirrors the formula according to which teaching and research resources are allocated to universities by the government. This was first observed by Marginson and Considine (2000: 149) for the 17 universities in their sample. The analysis of documents from all universities conducted by us as part of the case selection process confirmed this observation. Mirroring the research funding formula internally can thus be considered a "standard operating procedure" (March and Olsen 1984) of Australian universities.

University managers told us that although they do not like the funding formula because it does not appropriately measure research quality, the same formula is applied internally because the universities want to maximise their income. The management believe that income maximisation can be best achieved by adopting the external funding criteria and thus rewarding faculties and schools for contributing to the university's income. This mechanism is well known to organisational sociology as "mimetic isomorphism" (DiMaggio and Powell 1991). In our cases, organisations adopt institutional structures applied to them by powerful external actors rather than copying successful organisations, as has been described by DiMaggio and Powell (ibid.: 151-52). However, they clearly respond to the uncertainties of the environment and the poorly understood relationship between internal conditions and research output.

While the internal formulae use the same indicators as the government formulae, the weighting of the indicators is modified. According to the university managers we spoke to, this is necessary because the disciplines differ significantly in their reliance on external grants, on which the most consequential indicator is based. Assigning the same weight to this indicator as in the external formulae would seriously disadvantage the social sciences and humanities, in which research can be conducted with smaller grants or even without grants. To accommodate these disciplines, the weight assigned to external funding was reduced in both universities. U1 increased the weight assigned to research students, while U2 increased the weight assigned to publications.

A second mechanism is *strategic investment in grant acquisition*. The two universities reported on here specifically responded to the importance of external grants for both their income from the funding formula and their researchers' opportunities to conduct research at all by investing the money they had strategically in order to obtain as many external grants as possible. Strategic investment at the individual level was practiced by both universities in similar ways. Specifically, internal grants are made available to academics:

- whose external application nearly succeeded, in order to bridge the year to the next application round and to provide them with the means to further strengthen their application by conducting additional research ('near miss' grants);

- who need seed money for preparing an application, e.g. for beginning new collaborations with industry partners; or
- who have yet to build a track record that can win them grants (newly appointed staff and early career researchers).

Most internal grants in Australian universities are competitive in that academics need to apply for them and may fail. The criteria are similar to those of the funding councils, i.e. quality of the proposal, track record etc. However, no specific weighting is assigned to the criteria. Success rates have been described to us as being 'much higher' than those of the research councils.

All these grants have the sole purpose of improving an academic's chance to win external funding. In addition to this strategic investment at the individual level, U1 employed a mechanism of *strategic investment in critical mass*. At each level of the hierarchy from the Vice Chancellor to the Heads of Schools, 3-4% of the operating grant is allocated to strategic funds that are at the discretion of the respective manager. Additionally, a significant proportion of the RIBG is used to build a strategic fund at the discretion of the DVC (Research). The strategic funds are used to improve the conditions for research in selected areas by providing advanced equipment and time for research (via teaching relief and research-only personnel). The dominant organisational forms of these initiatives are research centres at the school, faculty, or university levels. Decision-making on centres and other strategic investments is hierarchical and applies a leverage principle. Managers at each level of the hierarchy must contribute some of their strategic funds for higher-level managers to contribute some of theirs. Thus, managers at several levels of the hierarchy must agree to a strategic investment at the faculty or university levels.

The centres are created in priority areas (i.e. areas which either are already seen as research strengths of the university or will be developed as such). Establishing a centre requires a pre-existing critical mass of research in that area (see Schiene and Schimank in this volume for similar recommendations of German assessors of university research). This creates problems for the social sciences and humanities whose research is more diverse and individualistic. Centres are expected to employ the competitive advantage created by the strategic investment to secure external grant funding. The grant funding must make centres self-sufficient after three to five years because the strategic investment ceases after that time regardless of the quality of the centre's research.

The role of the priority areas was interpreted differently by the managers at different levels of the hierarchy. A manager at the university level emphasised that belonging to a priority area does not affect strategic funding within the university, which would be allocated on the basis of scientific quality alone. Some of the Deans and Heads of Schools said that proposals from non-priority areas had to be at least slightly better in order to get funded. Academics were under the clear impression that priority areas receive more funding. Not surprisingly, the priority areas largely mirrored national research priorities because obtaining external grants was considered easier in these areas.

While U1 uses all of its research money strategically, U2 also uses a mechanism of *selective grant funding* that resembled the responsive mode rather than the hierarchical decision-making of U1's strategic investment in critical masses. Academics can apply for small research grants (for one year, up to 20.000 AU$), which are administered by faculties. These grants are used to compensate for lack of external funding. One school had introduced a rule according to which academics who have more than 100,000 Dollars in external grants are not eligible for internal grants. Decisions are made by faculty committees. One faculty has further devolved the decision process by letting the schools decide and just confirming these decisions.

A final mechanism that is applied by both universities on various occasions is *individual performance evaluation*. While the yearly performance appraisals are conducted rather informally within the schools and are inconsequential even where salary increments are concerned, evaluations of individual research performance inform decisions about promotions (along the ladder from lecturer to senior lecturer, associate professor, and full professor). Since U1 hires its academic staff on a tenure-track basis, i.e. beginning with five-year fixed term appointments and a decision about tenure after that period, evaluations of individual research performance are also conducted in the contexts of those decisions.

Decisions about promotion and tenure at U1 and about promotion at U2 are made by central university committees. Schools and faculties try to discourage weak applications but routinely agree to applications by 'their' staff, which leaves the actual decision to the university committees. Evaluations are based on the presupposition that each academic should be active in teaching, research, and administration. The higher the level an academic applies for, the more emphasis is laid on research. Interviewees commented that excellent teachers with little research could not become full professor, while excellent researchers who are bad teachers could. Decision criteria include indicators used in the funding formula, i.e. the committees take into account the numbers of publications, external grants awarded, and research student supervision. The procedures at U1 include an additional strong element of quality assessment because applicants for both promotion and tenure are asked to submit three research publications, which are externally assessed.

Another context in which individual research performance is assessed is the allocation of workloads. Academics report on all their activities in the areas of teaching, research, and administration in the previous year. Teaching loads for the next year are assigned on the basis of prior activities. While all academics except the 'research only' staff are expected to contribute to teaching, the teaching load can slightly vary depending on the amount of research points an academic gets for the previous year. However, the distribution of teaching loads is essentially egalitarian.

Consequences for Academics

The emphasis on research affects the teaching-research relationship. In U1, the general perception of academics and many managers was that teaching is of lower priority for the university than research. The investment in research only staff, the reduction of teaching loads for academics who were perceived as key researchers,

and the provision of key infrastructure for research centres was perceived to be cross-subsidised from teaching funds. A very high proportion of the academic staff at U1 is in 'research only' positions. Naturally, the teaching loads of the academics in standard 'teaching and research' positions increase. Grants from the ARC contribute to the separation of teaching from research because academics can apply for teaching relief, i.e. for funds for substitute teachers. Because of the growing teaching loads, this opportunity is used by an increasing number of academics, in particular from the social sciences, arts and humanities. Thus a vicious circle is emerging because the increasing teaching loads force academics to 'buy out' of teaching with their external grants, which in turn increases teaching loads for their colleagues.

Academics in U2 must also cope with increasing teaching loads because of the general scarcity of funding. Heads of School and academics observed the disappearance of the funds for casual teaching, a process that refers a large proportion of teaching back to the academics themselves. Apart from insufficient government funding of teaching, the necessary cross-subsidisation of research by teaching funds (which is due to the insufficient government funding for research) was mentioned as the cause for increasing teaching loads.

However, U2 is different in that teaching is still clearly regarded as the major task of the university. While concerns about increasing teaching loads and the decreasing quality of teaching have been voiced, none of the interviewees questioned the priority of teaching. The proportion of 'research only' staff and the student-to-teacher ratio of U2 are significantly lower than those of U1.

While the two universities applied partly different strategies in their allocations of resources for research, no significant differences in the financial situation of the academics could be observed. At the time of the interviews, only the academics from one School at U1 received recurrent funding for research (a few thousand dollars), which was possible because of their School's exceptionally high income from teaching. The other School budgets in U1 covered salaries and only the most basic infrastructure. Some 'travel grants' for conference attendance were allocated on the basis of applications and selection at school levels. The situation of members of strategically funded centres was somewhat better. However, strategic funding was tied to the creation of centres and to collaboration between faculties or at least schools. This requirement was not equally easily met by all disciplines. Historians and geologists complained that they either had to warp their research to make it fit the centre or would not get strategic funds. But even if they got 'centre money', academics still lacked research funding. The investments concentrated on key equipment and staff rather than money needed to conduct research projects. Centres had more researchers, research time and advanced equipment, but no basic supplies.

Although U2 allocates less money to strategic funds and more to faculties and schools, it is not able to provide recurrent funding for research to all academics. Interviewees reported that some recurrent funding was available in previous years when there was a surplus in the school budget. Nowadays there is not enough money to fund basic supplies.

Thus, research funding in both universities is intermittent, short-term funding that does not cover all costs of research, and is available only to a limited number of

selected academics. Whenever significant funding is required, the only way of obtaining it is to apply for external grants. Most of our interviewees were holding grants from one of the research councils at the time of the interview.[7] One mathematician from U1 and two political scientists from U2 reported that they had no money for research at all. Several academics from both Universities had only internal grants. Others supplemented their grant income by consultancies and other industry-funded projects.

Adaptation by Academics

The academics responded to three elements of their situations, namely to the individual performance evaluations, to the difficulties in obtaining research funding, and to insufficient amounts of resources available for their research. Three mechanisms that operated under these conditions could be observed.

Academics responded to both the individual performance evaluations and the conditions for getting grants by *adapting to indicators*. Academics were conscious of the need to have an impressive publication record and external grants in order to become promoted. Five interviewees reported that they changed their publication strategy by publishing more, publishing alone, and publishing in higher reputed (international) journals. A historian, a political scientist and a geologist named the promotion criteria as causing the changes, while two biochemists stated that they need to boost their track record in order to be successful with their grant applications. None of them reported changes in the content of research resulting from the changed publication behaviour. The descriptions gave the impression that these adaptations occurred ex post in decisions about how to publish finished research. However, changes in later research resulting from the new publication strategies cannot be excluded beyond doubt. A historian who did not need external funding (the expensive part of the project was completed) applied for an ARC discovery grant because having grants was a promotion criterion:

> You are encouraged to apply for grants. The university actually measures the input, not the output, in a sense, so that they reward you - they want to see grant money coming in. [...] I could, probably for the rest of my career, publish material from the files which I already have. I don't really need to ever get any more research material. No need. I have enough. But that's not how universities work. And that's quite distasteful really. I mean, they are pushing us along to apply for money to do new things rather than to say, "Well, for the next 20 years I'm perfectly happy using the files that I have amassed to write more work". They don't want you to do that. They press you to go and get more money.
>
> (Historian, Associate Professor)

An adaptation to indicators by a biochemist significantly affected his research. He was asked by an industry partner to conduct a research project outside his main

[7] This should not lead to the conclusion that most university academics hold external grants. The universities and academics analysed here are not 'representative' because (a) we present results from two research intensive universities, and (b) we selected academics who visibly conducted research because we are interested in the adaptation of research to funding conditions.

research interest and agreed in order to enhance his publication record, which he deemed necessary for securing ARC funding in the future.

A second mechanism is *accommodating patrons*. This mechanism is very widespread because all research we investigated depends on decisions about grants, which are made by university committees and managers (about internal grants or centres) or by funding agencies. Academics are convinced that they need to anticipate and accommodate the interests of the decision-makers, and interpret both success and failure in obtaining funding in these categories. The adaptation mechanism operated at two different levels. At the level of the 'research portfolio' of an academic who followed several lines of research simultaneously, lines of research that were deemed 'unfundable' because they are 'basic' or 'not hot' were given up. Several of our interviewees ended whole lines of research for this reason.

> And then I […] came here [from the US] and ran into a wall and just absolutely could not continue that. And the wall was two things actually: one, there's less of a focus here on basic science, whereas in the US there would be just no question that that would be fundable science. Here it had trouble passing the "significance" test. […] And here, you have to convince the particular reviewer that gets your proposal that this is significant. And that reviewer might work on something quite unrelated to what you're working on, and they might just think, "This is absurd". […] And that caused a lot of problems, yeah. So that was just impossible to continue.
>
> (Biochemist, Senior Lecturer)

Applied or fashionable lines were continued and enhanced when they already existed, or were created. In its stronger version, this move included the search for industry partners and applications for linkage grants. Application-oriented and otherwise fashionable topics were also selected for discovery grant applications. This happened across all five fields.

> So we became interested in it because we were working in [that] area […]. And to be quite frank about it, we developed the project because we knew that that was the sort of thing that could be attractive to the funding agencies, to the ARC. So, you know, the ARC has priority areas of what they consider special priorities. So we thought, well, okay, let's try to develop a project that fits into their priorities for funding. That's how it came about.
>
> (Mathematician, Research Fellow)

The mechanism of accommodating patrons occurred not only at the level of lines of research and research portfolios. It was also applied in the fine-tuning of projects, where applied or fashionable aspects of the research were emphasised and enhanced. For example, one of the historians switched to "politically relevant history" in order to improve the chances of his project proposal.

> I think it's true too at the ARC Discovery level that - I say this because I've recently applied for [a grant] and I'm very conscious of the topic - the research that I'm going to do and the project that I'm going to put up is one that I think is going to be sellable as a humanities person with some sense of national interest. Even though it's in the [earlier] centuries, it's got to have some 21st century bite. So that's going to drive the content of my research in a particular direction rather than another. I'm not going to spend my time whipping up [a grant] on, say, ghosts, fairies, goblins and elves in [earlier period]. It would be a good topic, a great topic. But it's not going to sell to the ARC.
>
> (Historian, Associate Professor)

A third mechanism that could be identified is *opportunistic fundraising*. Under special circumstances, academics conducted scientifically uninteresting projects or consultancies in order to get funding for their research. This mechanism operates if the need for money cannot be satisfied by other sources and an industry partner with matching needs exist. Opportunistic fundraising was practised even by holders of ARC grants because not all necessary expenses can be funded with these grants.

> But you do things that are unrelated. One of the ways we funded the lab is, I do consultancy for [a government agency] […] I did it partly because it was interesting, but more because of money, you know, very well paid work. And I mean […] technically, it's quite interesting, it's quite satisfying technically, but it's not in any sense related to my core research interests, and we basically do it so we've got money for the lab.
>
> (Geologist, Associate Professor)

Having obtained research funding, academics need to match their research projects to the actually available amount of money. This is obviously necessary for academics who are forced to fund their research exclusively by the small internal grants. Interestingly, many of their colleagues who received external grants experienced significant cuts of their budgets and thus faced discrepancies between project design and available funding. All these academics responded by *downsizing or stretching* their projects.

Downsizing occurred with respect to all major features of the project. In mathematics there is no empirical research that could be downsized. Therefore, the only available strategy was narrowing the problem by limiting the range of topics that were addressed in a project. The empirical disciplines could go further by narrowing the research object by either reducing the number of objects investigated (reduce the number of sites for fieldwork, reduce time spent in archives) or using less suitable objects (e.g. sites for fieldwork closer to the university). Similarly, academics could narrow the methodology by reducing the variety of methods used to investigate an object or applying 'cheaper' methods that used less expensive equipment.

> Under ideal conditions the project would look very differently because of two things really. One, I would have an extended stay in B. [...] If you can come down for six weeks it is much more benefit than if you were jetting in for two weeks or ten days or something like that. Because of the connections you have to make and sometimes it takes time to do things. Two ways it would look different, one; under ideal funding conditions it would have enabled me, to spend an extended period of time in B. And I think that would be very relevant and necessary to what I'm trying to do with the project. And secondly, I would have had funds to attend a couple of [certain] meetings and have access to delegates.
>
> Interviewer: You would just interview more people in B.?
>
> Yes, I would interview more people in B. .. and also when you go to these [political events] you sit on the edge of things as well. But sitting on the edge of things you learn quite a lot about what outside of the [event happens] - just by observation and talking to people ... So, ideal conditions really in a sense would get me closer to the source of what I'm trying to understand and write about.
>
> (Political Scientist, Professor)

Another mechanism – stretching the project – is not linked to any strategic design of the project but is rather passive. It occurs when academics do as much as is possible at the time and do less if there is less funding. One historian and two mathematicians told us that their research slowed down due to a lack of funding.

While situations of insufficient funding occurred frequently across all disciplines and triggered the described adaptive behaviour, there were also academics who did not report any adaptations. Ten of our 32 interviewees described no adaptations whatsoever. With one exception (a geologist whose entire work was industry-related), all academics who did not report adaptations work in fields that are not resource-intensive (four in mathematics, three in political science, two in history).

Having described the mechanisms at the individual level, we would like to draw the reader's attention to two mechanisms that we expected to occur but didn't find. Firstly, academics did not adapt to the external or internal funding formulae for the simple reason that they did not receive any money according to this formula. The 'research money' distributed to faculties and schools covered part of the salaries and basic infrastructure. The few remaining funds were used for internal grants and to support some travel. None of this money was distributed according to any formula. The academics at the two universities did not perceive the funding formulae as consequential for their research conditions.

Secondly, an adaptation process that could have been expected under conditions where only the very best research gets external funding is the attempt to improve the quality of one's own research by turning to central problems of the field and achieving solutions that can be published in leading journals. This mechanism – the *improvement of research by choosing important problems at the research frontier* - is one that the various performance-based funding procedures are supposed to trigger. However, we observed no attempts by academics to turn their research into 'world-class research', which is one of the stated aims of the RES (Kemp 1999: 6). The academics we interviewed did not extend the space in which they sought research problems but rather adapted within the limits of that space according to the chances of funding they perceived.

Changes in Knowledge Production

The adaptation of universities to the formula-based funding of their research resulted in an increased emphasis on and better support of research. As a result of the formula-based funding, there is more research in these universities than there has been before. However, this is not a result of the specific procedure applied in the Australian RES. Any strong RES is likely to increase the emphasis on research because it turns research into a source of income for the university. As long as teaching activities are funded without taking their quality into account, the observed shifts in the relationship between teaching and research are also likely to occur under any strong RES.

Apart from this general redistribution of attention and resources from teaching to research, mechanisms at the university level had two major consequences. Strategic investment in critical mass (as observed in U1) provides a competitive advantage to collaborative, interdisciplinary research. By strategically investing in grant applications,

universities significantly provided a competitive advantage to research that is likely to be approved by external sources of funding, thus reinforcing the latters' thematic preferences.

While the epistemic consequences of the mechanism 'adaptation to indicators' are difficult to assess, the other mechanisms at the individual levels can be unambiguously linked to changes in the content of research. As a result of the described adaptive behaviour, research becomes more applied, approximates the mainstream, narrows, and its results become less reliable (less rigorously tested). The increasing orientation towards applications is produced by the internal priority setting of universities and better chances of grant funding for such research. Even the responsive mode of ARC grant funding that is supposed to be thematically neutral is biased in favour of applied topics. The same holds for 'hot' topics that represent the current focus of international scientific communities.

By following these fashions, the Australian grant funding system favours the mainstream against nonconformist perspectives. Researchers drop lines of research that are 'too basic' or 'unfashionable' and advance the remaining research lines towards more applied and 'hot' topics. This implies that their research trails also narrow, i.e. academics investigate fewer topics and observe a narrower field of knowledge production. Since the recombination of knowledge and the creation of links between different fields is a major mechanism of innovation in the production of scientific knowledge (Gläser 2006), the narrowing of research trails reduces the potential for such innovations. It also limits the potential for collaborative research because narrower research trails provide fewer 'docking points' for researchers from other fields. A systematic trend towards narrower research trails could also lead to a reduced diversity at the level of scientific fields in Australia. However, we cannot identify this effect with the methods used in this project.

The reduced scope and reliability of research is caused by the necessity to adapt the project design to the funding that is actually available. Fewer empirical objects, less suitable empirical objects, fewer experiments, measurements, or methods all mean that the knowledge claims offered to the scientific community in publications are less well grounded then they could be. Since these features are directly linked to the quality standards of the scientific communities, we can say that reduced scope and reliability also mean reduced quality of research.

INTENDED AND UNINTENDED EFFECTS OF STRONG RES

When we compare the results of our analysis with Whitley's sketch of effects of strong RES (Whitley, this volume), some differences become apparent. The Australian RES affects the stratification of disciplines only insofar as disciplines that rely heavily on external grants are more important for the universities. Changes in the social structures of disciplines or practices of knowledge production cannot be ascribed to the RES. The reason for this is that the formula-based RES relies on the most basic indicators of academic behaviour and is therefore 'egalitarian'.

Although egalitarian with regard to individual contributions, the Australian RES achieves a highly skewed distribution of research funds, with eight universities

receiving about 64% of the funding. Its effects are limited because in its attempts to economise the government has weakened the steering instrument, i.e. the amount of money. Major effects of the RES are an increasing general support for research and attempts to concentrate resources on the best performers at the expense of others. The content of research is changing mainly because academics respond to the meagre and biased funding environment, of which the RES is the least important part.

The strong pressure of the funding environment has been the most significant influence that has forced most of the interviewed academics to adapt their research strategies. As a result of this adaptation, their research is becoming less diverse, less fundamental, and less reliable. We did not observe moves towards 'better' research by addressing more fundamental problems and providing surprising solutions to them. Except for the abandonment of whole lines of research, changes were topical and incremental.

There might be a deeper reason why academics do not simply 'improve' their research in spite of the mounting pressure. The collective production of knowledge by scientific communities applies a self-selection of tasks because only the scientists themselves are able to formulate tasks they can solve (Polanyi 1962; Benkler 2002; Gläser 2006). If formulating a task that makes sense to the academic is a necessary prerequisite of successful research, than RES can change neither the way in which tasks are selected nor the tasks themselves. At the current stage of our investigation, we would hypothesise that the adaptive behaviour follows a bell curve that is produced by the overlap of two exponential curves (figure 3). The ability to adapt to external conditions highly depends on the capabilities of an academic. Only excellent scientists are able to move across a wide problem area and are able to move between minor and fundamental problems. This ability decreases when we move to the majority. However, the pressure to actually change their research is highest for the academics who are least successful. The overlap of these two characteristics leads to the bell curve which means that significant adaptation will be found only in the middle field where a recognisable pressure towards adaptation coincides with limited capabilities to adapt.

If our hypothesis is correct and many researchers are unable to adapt, then RES in this context can only lead to either increasing discrepancies between institutionalised expectations about the 'right' research and the practices of task definition or to a redistribution of funds to researchers whose practices of tasks definition meet the demands of RES. Thus, in order to improve their research performance, universities should hire better researchers rather than 'work on' the ones they currently have. However, this strategy has already been in place for a long time and cannot significantly change the income of a university because all its competitors are doing the same. Whatever strategy will be applied, will be another case of 'running as fast as you can in order to stay where you are'.

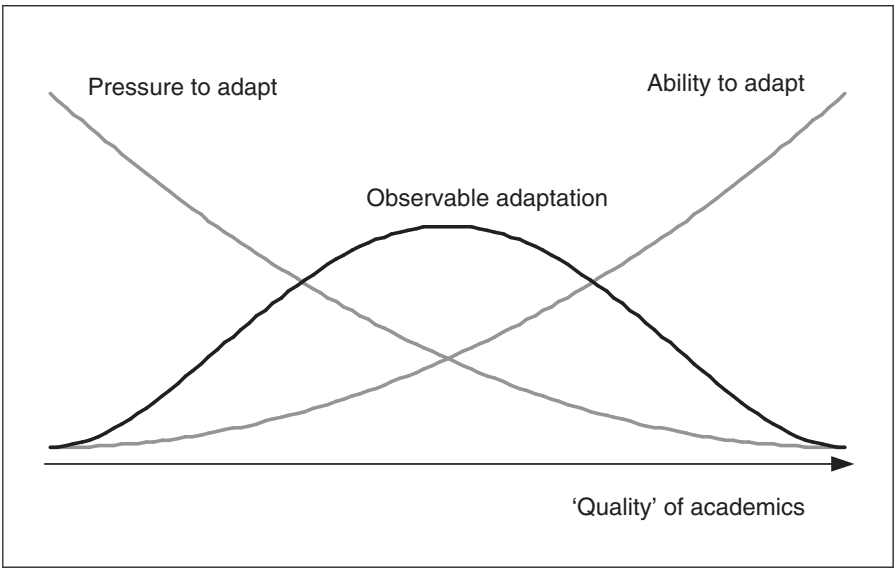

Figure 3. Observable adaptive behaviours resulting from the interaction of pressure and abilities to adapt

Our study has also revealed an important limitation in studying the impact of RES on the content of research. A possibly significant cognitive change is occurring because certain kinds of research are no longer conducted. While we could identify the reasons why certain lines of research are discontinued, we were unable to ascertain features of research that might be systematically suppressed by the research funding system. To obtain these features, we would need to investigate research that has not been conducted, which is obviously impossible.

This dilemma leads us to two methodological conclusions. Firstly, we need to attempt an even more detailed investigation of research processes, namely comparative participant observations of different research settings. With such an approach, a more systematic comparison of conducted and abandoned research processes would become possible. Secondly, it might well be the case that these changes cannot be observed at the micro-level of individual research processes at all. An approach to the measurement of cognitive changes at the meso-level of scientific fields might be necessary. Such an approach could include comparisons of epistemic features of national scientific fields that are subject to different RES. We will try to develop such an approach in further projects (Schmidt et al. 2006).

REFERENCES

ARC [Australian Research Council] (2005), *Australian Research Council Report 2004-2005,* Canberra: Australian Research Council, http://www.arc.gov.au/pdf/annual_report_04-05.pdf (accessed 18 August 2006).

AVCC [Australian Vice Chancellors' Committee] (2005), *Key Statistics on Higher Education*, http://www.avcc.edu.au/documents/publications/stats/2005Edition.pdf (accessed 18. August 2006).

Benkler, Yochai (2002), 'Coase's Penguin, or, Linux and The Nature of the Firm', *Yale Law Journal*, 112, 369-446.

Böhme, Gernot, Wolfgang van den Daele, Rainer Hohlfeld, Wolfgang Krohn and Wolf Schäfer (1983), *Finalization in Science. The Social Orientation of Scientific Progress*, Dordrecht: Reidel.

Böhme, Gernot, Wolfgang van den Daele and Wolfgang Krohn (1973), 'Die Finalisierung der Wissenschaft', *Zeitschrift für Soziologie*, 2, 128-144.

Braun, Dietmar (1993), 'Who Governs Intermediary Agencies? Principal-Agent Relations in Research Policy-Making', *Journal of Public Policy*, 13, 135-162.

Braun, Dietmar (1998), 'The role of funding agencies in the cognitive development of science', *Research Policy*, 27, 807-821.

Chubin, Daryl E. and Terence Connolly (1982), 'Research Trails and Science Policies' in Norbert Elias, Herminio Martins and Richard Whitley (eds.), *Scientific Establishments and Hierarchies*, Dordrecht: Reidel, pp. 293-311.

DEST [Department of Education, Science, and Training] (2005a), *Australian Science and Technology at a glance 2005*, Canberra: Department of Education, Science, and Training, http://www.dest.gov.au/NR/rdonlyres/841985DC-E046-4E82-B07D-478E70D0A5A2/8649/at_a_glance_2006i.pdf (accessed 18 August 2006)

DEST [Department of Education, Science, and Training] (2005b), *Higher Education Workplace Relations Requirements (HEWRRs) under the Commonwealth Grants Scheme (CGS),* Canberra: Department of Education, Science, and Training, http://www.dest.gov.au/sectors/higher_education/programmes_funding/programme_categories/professional_skills/hewrrs/ (accessed 18 August 2006).

DiMaggio, Paul J. and Walter W. Powell (1991 [1983]), 'The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields', in Walter W. Powell and Paul J. DiMaggio (eds.), *The New Institutionalism in Organizational Analysis*, Chicago: University of Chicago Press, pp. 147-160.

Gläser, Jochen (2006), *Wissenschaftlich Produktionsgemeinschaften. Die soziale Ordnung der Forschung*, Frankfurt am Main: Campus.

Gläser, Jochen and Grit Laudel (2006), *Experteninterviews und qualitative Inhaltsanalyse als Instrumente rekonstruierender Untersuchungen*, second edition, Wiesbaden: VS Verlag für Sozialwissenschaften.

Gläser, Jochen, Grit Laudel, Sybille Hinze and Linda Butler (2002), *Impact of evaluation-based funding on the production of scientific knowledge: What to worry about, and how to find out,* Report to the German Federal Ministry for Education and Research, http://www.sciencepolicystudies.de/dok/expertise-glae-lau-hin-but.pdf (accessed 3 November 2006).

Guston, David H. (1996), 'Principal-Agent Theory and the Structure of Science Policy', *Science and Public Policy*, 23, 229-240.

Hedström, Peter (2005), *Dissecting the Social: On the Principles of Analytical Sociology*, Cambridge: Cambridge University Press.

Kemp, David A. (1999), *Knowledge and Innovation: A policy statement on research and research training,* Canberra: Minister for Education, Training and Youth Affairs, http://www.dest.gov.au/archive/highered/whitepaper/report.pdf (accessed 3 August 2006).

Kleinman, Daniel Lee (1998), 'Untangling Context: Understanding a University Laboratory in the Commercial World', *Science, Technology & Human Values*, 23, 285-314.

Knorr-Cetina, Karin (1981), *The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science*, Oxford: Pergamon Press.

Knorr-Cetina, Karin (1995), 'Laboratory Studies. The Cultural Approach to the Study of Science', in Sheila Jasanoff, Gerald E. Markle, James C. Petersen and Trevor Pinch (eds.), *Handbook of Science and Technology Studies,* London: SAGE, pp. 140-166.

Latour, Bruno and Steve Woolgar (1986 [1979]), *Laboratory Life: The Construction of Scientific Facts*, Princeton: Princeton University Press.

Laudel, Grit (2006), 'The art of getting funded: How Scientists adapt to their funding conditions', *Science and Public Policy*, 33, 489-504.
Lynch, Michael (1985), *Art and Artifact in Laboratory Science: A Study of Shop Work and Shop Talk in a Research Laboratory*, London: Routledge & Kegan Paul.
March, James G. and Johan P. Olsen (1984), 'The New Institutionalism: Organizational Factors in Political Life', *The American Political Science Review*, 78, 734-749.
Marginson, Simon (2006), 'Dynamics of national and global competition in higher education', *Higher Education*, 52, 1-39.
Marginson, Simon and Mark Considine (2000), *The Enterprise University: Power, Governance and Reinvention in Australia,* Cambridge, UK: Cambridge University Press.
Mayntz, Renate (2004), 'Mechanisms in the Analysis of Social Macro-Phenomena', *Philosophy of the Social Sciences*, 34, 237-259.
Mayntz, Renate, and Fritz W. Scharpf (1995) 'Der Ansatz des akteurzentrierten Institutionalismus' in Renate Mayntz and Fritz W. Scharpf (eds.), *Gesellschaftliche Selbstregelung und politische Steuerung*, Frankfurt am Main: Campus, pp. 39-72.
Mayntz, Renate, and Uwe Schimank (1998), 'Linking theory and practice: Introduction', *Research Policy*, 27, 747-755.
Merton, Robert K. (1968), 'On Sociological Theories of the Middle Range' in Robert K. Merton (ed.) *Social Theory and Social Structure*, London: The Free Press, pp. 39-72.
Morris, Norma (2000), 'Science policy in action: Policy and the researcher', *Minerva*, 38, 425-451.
Nagi, Saad Z. and Ronald G. Corwin (1972) 'The Research Enterprise: An Overview' in Saad Z. Nagi and Ronald G. Corwin (eds.) *The Social context of Research*. London: Wiley, pp. 1-27.
Nelson, Brendan (2005), *Higher Education Report 2004-05*, Canberra: Minister for Education, Science and Training, http://www.dest.gov.au/sectors/higher_education/publications_resources/profiles/highered_annual_report_2004_05.htm (accessed 8 August 2006).
North, Douglass C. (1990), *Institutions, Institutional Change and Economic Performance,* Cambridge: Cambridge University Press.
Polanyi, Michael (1962), 'The Republic of Science', in Edward Shils (ed.), *Criteria for Scientific Development: Public Policy and National Goals*, Cambridge: Massachusetts Institute of Technology, pp. 1-20.
Powell, Walter W. and Paul J. DiMaggio (eds.) (1991), *The New Institutionalism in Organizational Analysis*, Chicago: University of Chicago Press.
Rip, Arie (1982), 'The Development of Restrictedness in the Sciences', in Norbert Elias, Herminio Martins and Richard Whitley (eds.), *Scientific Establishments and Hierarchies*, Dordrecht: Reidel, pp. 219-238.
Rip, Arie (1994), 'The Republic of Science in the 1990s', *Higher Education*, 28, 3-32.
Ruivo, Beatriz (1994), ''Phases' or 'paradigms' of science policy?', *Science and Public Policy*, 21, 157-164.
Scharpf, Fritz W. (1997), *Games Real Actors Play. Actor-Centered Institutionalism in Policy Research,* Boulder: Westview Press.
Schmidt, Marion, Jochen Gläser, Frank Havemann and Michael Heinz (2006), 'A Methodological Study for Measuring the Diversity of Science', in Jean-Charles Lamirel, Claire Francois and Hildrun Kretschmer (eds.), *Proceedings International Workshop on Webometrics, Informetrics and Scientometrics & Seventh COLLNET Meeting*, Nancy (France).
Scott, W. Richard (1991), 'Unpacking Institutional Arguments', in Walter W. Powell and Paul J. DiMaggio (eds.), *The New Institutionalism in Organizational Analysis*, Chicago: University of Chicago Press, pp. 164-182.
Silvani, Alberto, Giorgio Sirilli and Fabrizio Tuzi (2005), 'R&D Evaluation in Italy: more needs to be done', *Research Evaluation*, 14, 207-215.
Van der Meulen, Barend (1998), 'Science policies as principal-agent games: Institutionalization and path dependency in the relation between government and science', *Research Policy*, 27, 397-414.
Van der Meulen, Barend and Loet Leydesdorff (1991), 'Has the Study of Philosophy at Dutch Universities Changed under Economic and Political Pressures?', *Science, Technology, and Human Values*, 16, 288-321.
Whitley, Richard (1972), 'Black Boxism and the Sociology of Science: A Discussion of the Major Developments in the Field', in Paul Halmos (ed.), *The Sociology of Science,* Sociological Review Monograph 18, Keele: University of Keele, pp. 61-92.

Whitley, Richard (1977), 'Changes in the Social and Intellectual Organisation of the Sciences: Professionalisation and the Arithmetic Ideal', in Everett Mendelsohn, Peter Weingart and Richard Whitley (eds.), *The Social Production of Scientific Knowledge*, Dordrecht: Reidel, pp. 143-169.

Whitley, Richard (1984), *The Intellectual and Social Organization of the Sciences*. Oxford: Clarendon Press.

CHAPTER 7

STEFAN LANGE

THE BASIC STATE OF RESEARCH IN GERMANY: CONDITIONS OF KNOWLEDGE PRODUCTION PRE-EVALUATION

INTRODUCTION

In terms of systematic performance evaluation, research in most German universities still represents what might be termed a 'basic' state because the recent attempts at evaluating research outputs are so new and lacking in direct consequences that they have not yet affected research performance. Analysing this basic state can provide useful information about the situation that research evaluation systems can be expected to change. In this chapter I report the preliminary results from interviews with academics in five different disciplines in one university in North Rhine-Westphalia (NRW) that can be deemed to represent the basic state of research that is not yet affected by retrospective research evaluations of any kind.[1] This analysis is linked to that conducted by Gläser and Laudel in Australia, and the same methods are applied in the German case study as in the Australian one (see Gläser and Laudel, this volume).

The five disciplines discussed are: political science, history, mathematics, biology, and geology. I interviewed five academics from each discipline – three full professors and two associates at the postdoctoral level. Additionally I conducted three interviews at the level of the central university leadership and three interviews with the deans of the faculties where the disciplines are located. This provided us with information about local organisational research strategies and the evaluation tools that academics could be supposed to adapt to.

In order to understand the changing overall context in which performance evaluation of academic research is being developed in Germany, I shall summarise the key features of the shifting governance structures at the national and regional

[1] Funding of this project by the German Federal Ministry of Education and Research in its programme 'science policy studies' is gratefully acknowledged. For reasons of privacy protection, university documents are not listed as sources in the reference section. Furthermore, figures indicating the size of the university are changed without distorting the major information, and names of faculties have been changed.

R. Whitley and J. Gläser (eds.), The Changing Governance of the Sciences, 153–170.

levels in the next section. I then describe how the central university administration and faculty deans are dealing with this changing context, before summarising the major ways in which academics in different fields have adapted to new demands, and their limited responsiveness to them.

THE CHANGING GOVERNANCE OF ACADEMIC RESEARCH IN GERMANY

The systematic evaluation of academic research performance became widely discussed in the aftermath of German Reunification when the research of the institutes of the former East-German Academy of Sciences was evaluated in order to integrate the parts that met international standards into the German research landscape and to close down the rest (Mayntz 1994; Wolf 1995). In the wake of this process the first German 'system evaluation' of associations of non-university research institutes – namely of the research institutes of the *Max-Planck-Gesellschaft*, the *Fraunhofer-Gesellschaft*, the *Helmholtz-Gemeinschaft* and the *Leibniz-Gemeinschaft* – took place (Röbbecke and Simon 2001; Krull and Sommer 2006). Today all institutions in the German non-university research sector except for those directly subordinated to the federal and state governments are subject to regular internal and external evaluation by peer review.

However, this large-scale evaluation exercise did not spill over to university research except for the new evaluations in Lower Saxony (Schimank and Schiene, this volume). Unlike non-university research, which is under the joint authority of the federal government and the regional states, German universities are almost entirely governed and funded by the latter. The federal government provides only the framework law that today regulates access to higher education, graduation formalities, and some general aspects of employment.[2] The states have traditionally funded universities by line-item budgets based on professors and their contracts.

This changed in the last decade as many states struggled to introduce New Public Management (NPM) instruments in the steering of their higher education (Lange and Schimank 2007: 538-541). Lump-sum budgeting was introduced, and most states developed funding formulae based on performance indicators for part of the resource allocation. The amount of the overall university budgets allocated by formula, the construction of the formulae and the kind of performance indicators used are all highly diverse across the 16 states (Leszczensky and Orr 2004). The major indicator of research performance, which is used in nearly all states, is the amount of external research funding. Some states additionally use the numbers of successful PhD graduations and *habilitations* (Jaeger 2005: 8-9). To prevent weakly performing universities from getting into financial trouble, the overall gains or losses for individual universities are usually limited to 1-5% with regard to the preceding budget (Jaeger et al. 2005: 5).

[2] As a consequence of the recent reform of German federalism the federal government's power to act on the higher education system via framework law will be abolished in 2008. Universities and *Fachhochschulen* will then be completely under the legal and financial authority of the sixteen states.

Another innovation in German Higher Education policy that is related to evaluations is the negotiation of individual 'performance agreements' between a state and each of its public universities since the early 2000s (Kehm and Lanzendorf 2006: 159-161). In these quasi-contracts, a university and the state agree on necessary improvements of performance in teaching and research, and on measures by which the university will create a unique 'profile'. Typical measures by which 'research profiles' are built include cooperation with local non-university research institutions, the creation of (preferably interdisciplinary) research centres and graduate schools, and internationalisation activities. Performance agreements must be considered a very weak research evaluation system (RES, see Whitley in this volume) because most of them define only vague performance goals and do not mention any sanctions for non-compliance. In none of the states have the consequences of a university's failure to reach the defined goals been clarified.

The trend towards profile building in universities has recently been reinforced by the 'Excellence Competition' that has been jointly introduced by the federal government and the states (Schimank and Lange 2006; Weingart and Maasen, this volume). In this first national-level initiative promoting university research the German universities compete for shares of a 1.9 billion € budget by submitting proposals for graduate schools, 'clusters of excellence'[3] and outstanding concepts for the future development of a university. Additional funding is provided selectively for good performers.

These recent initiatives do not, though, mitigate the significant reduction in financial support for German universities that has taken place over the last three decades and which has made an increasing proportion of German university research dependent on external funding. Compared to the Australian situation as reported by Gläser and Laudel in this book, however, the German research landscape is rich and diverse. University professors can get external research funding on a competitive base from a large number of different kinds of funding sources. The most prominent source is the DFG, which provides more than 40% of the external funding for university research (Kuhlmann and Heinze 2004: 53). A significant proportion of that funding is allocated in a competitive responsive mode. During the last twenty years DFG funding has grown increasingly competitive and the approval rate of the individual funding has declined from 59,5% in 1986 to 36,6% in 2005 (HRK 1993; DFG 2006: 112).

The second pillar of the German funding landscape consists of the thematically selective and often mission-oriented funding programmes of the German Federal Ministry of Education and Research (BMBF), other ministries at the federal and state levels, and the framework programmes of the European Commission. Funding by these sources is competitive, but eligibility is linked to research on predefined topics or applied research. In an evaluation of its funding processes in 1998 the DFG was criticised for its funding of small individual projects. It was recommended that the DFG should concentrate more on thematically focused programmes: "*In principle, resources should be concentrated on a few thematic fields and on fewer more visible projects*." (Kehm and Lanzendorf 2006: 165-166) In addition to the

[3] This means close collaboration with local or regional non-university research institutes.

already existing Large Collaborative Research Projects the DFG introduced funding for large decentralized researcher groups, thematically focused programmes and research centres.

Thus, the German funding landscape is developing a bias towards thematically focused, collaborative or otherwise 'big' research in both the institutional funding stream (performance agreements and excellence competition) and external funding. One important corollary of this trend is that the influence of politicians, funding administrators, and the scientific elite involved in funding bodies is increasing while individual researchers are loosing their agenda setting power in the long run.

Turning now to consider the particular regional context of the university we are investigating here, NRW has the most universities in Germany (15 in total) and more than one quarter of all German students (27%) are enrolled in NRW higher education institutions. In this state, a steadily increasing proportion of the allocation for casual staff and equipment has been distributed since 1994 according to a formula that uses performance indicators. The share of indicator-based funding now constitutes 20% of the universities' lump-sum budget (Hartwig 2006: 19). The formula uses a mix of input and output indicators with specific weightings (Leszczensky and Orr 2004: 35). Most of the formula-based funding is allocated on the basis of teaching performance and size. Only 5% of a university's total budget is allocated according to research indicators (4% based on the amount of external funding and 1% based on the number of PhD graduates). The varying capital intensity of disciplines is taken into account by assigning different weightings to three groups of disciplines (ibid.).[4]

In 2001, the whole NRW higher education landscape was evaluated for the first time. An expert committee evaluated all disciplines in the 15 universities with regard to student demand, graduation rates and research activity. This evaluation led to the closure of some courses and institutes, to the merger of the universities of Duisburg and Essen, to several recommendations for the contents of performance agreements (Hartwig 2006: 20-21), and to plans for a monitoring system in which universities report on the achievement of performance goals (Fangmann 2006: 60). The performance agreements between the state of NRW and its universities are supported by an 'innovation pool' of 1,000 from formerly 2,000 positions that have been taken from universities. These positions are reallocated to universities in order to support teaching and research in areas defined in the performance agreements (Hartwig 2006: 18-19).

As a result of these measures, a significant part of the universities' basic funding is now based on performance criteria of some sort. In return the universities obtained the ministry's promise to be exempted from overall budget cuts until 2006 (Fangmann 2006: 54-55).[5] However, the consequences of these evaluation measures

[4] Since 1999, funding for the medical sciences is allocated according to a separate formula that allocated first 10% and now 20% of the funding according to performance indicators. The two research related indicators measure the amount of external funding (weighted 42.75% in the formula) and the number of publications (weighted 28.5%) (Leszczensky and Orr 2004: 35-36).

[5] Since the 1.1.2007 the third generation of performance agreements became effective, with the state of NRW promising a steady annual allocation of a total sum of 3.7 billion € to the whole system of universities and Fachhochschulen until the year 2010.

for research have been rather weak. While both the federal framework law and the higher education laws of NRW require evaluations of teaching and research in all higher education institutions, attention is still very much focused on teaching. It is thus not surprising that while all charters of NRW's universities now contain chapters on evaluation, these are only concerned with teaching and do not prescribe or regulate the evaluation of research.

UNIVERSITY AND FACULTY EVALUATION AND STEERING OF RESEARCH

The university under investigation is large by German standards. A rector and his vice rectors lead the university and represent it as an academic corporation. The university consists of several large faculties headed by deans. Rector, vice rectors and deans are elected by the traditional bodies of academic self-governance for a fixed term. They are all full professors and will return to their institutes and chairs after having served their term. Administrative matters and finances are the responsibility of the university's chancellor who is a public servant appointed by the state government. He is the head of the administration and represents the state's concern for the correct use of resources in the central executive of the university.

The university has an evaluation system for teaching but not for research. While the rectorate collects information about research projects and publications from all faculties for a biannual university research report and the central administration has additional information about the amount of external funding, none of this information is used for internal evaluations. The deputy chancellor explained that the many simultaneous pressures for reforms make it impossible for the university leadership to push for research evaluation at this time. However, attempts to establish procedures of self-evaluation by certain institutes and disciplines – e.g. biology and medicine – are financially supported by the rectorate. There are also some instruments at the central level for building the research profile of the university according to the performance agreement with the state of NRW. The rector has recently established a board that advises faculties that have to fill vacant chairs on research profiles. He has also created a central fund for supporting the preparation of proposals for large cooperative research grants.

Neither of these instruments reduces the autonomy of the faculties or chairs. The performance agreement has been largely developed in a bottom-up process, i.e. based on proposals made by the faculties. The project proposals supported by the central funds are also bottom-up initiatives. The interviewed vice rectors shared the opinion that it should not be the task of the central university leaders to steer research in certain directions. Excellent research should grow from the chairs, whose initiatives could then be amplified and supported by the rectorate. Both vice rectors stated that it would be a dangerous illusion of some politicians to think that research could be directed from above. The rectorate sees its role in profile building and strategic research management mainly lying in recruitment policy.

Considering next the changing role of faculties and their deans, it is important to note that while all three of the faculties analysed here are rather large, student

enrolment and teaching loads vary. The Faculties of Humanities and of the Social Sciences have overall high student enrolment and therefore high teaching loads, while teaching loads in the Faculty of Natural Sciences are lower on average and vary between the disciplines, as summarised in table 1.

Table 1. Faculty Characteristics and Evaluation Schemes

Faculty	*Social Sciences*	*Humanities*	*Natural Sciences*
Students per chair	170	124	68
External funding per chair	43,000 €	70,000 €	220,000 €
Profile building	Recruitment of new professors emphasizes research and collaboration		
	Building research centres in some areas		
Research evaluation schemes	Voluntary (small amounts of money affected)	None (and none planned)	None (but under discussion)

All interviewed deans explained that although they have now by law much more discretion to make strategic moves and to put more pressure on bad performers, they still prefer the principle of decision-making by consensus. All deans claimed that top down decision making with a harsh managerial attitude would lead nowhere in very big faculties – according to one dean "this is unthinkable".[6]

While the funding of universities has changed significantly, internal fund allocation still follows the traditional pattern and does not apply performance indicators. This means that 80% of a faculty's budget is allocated directly to chairs and only 20% is at the discretion of the faculty. The deans can use up to half of the discretionary money to build a strategic fund for 'profile building'. All deans reported that they make use of this opportunity (see below). However, the allocation is too small to achieve any significant effect. The remaining 10% of the 'free' money can be distributed to chairs either as incremental increases or according to performance indicators at the discretion of the faculties.

The three faculties in our sample significantly differ in their use of these opportunities. The Faculty of Social Sciences introduced a performance-based funding mechanism some time ago, which uses 10% of its recurrent funding. Chairs can receive money from this pool according to their teaching, administrative and research performance. The latter is measured by

- profile building activities;
- presentations at high ranking conferences;[7]
- the amount of external funding; and

[6] However, the deans of the Faculty of Social Sciences and of the Faculty of Natural Sciences were described by their colleagues as leaders and "agenda setters" who make well-defined use of their power when it comes to the re-dedication of vacant chairs and the introduction of incentives for successful research.

[7] Disciplines agree on the 20 most outstanding events in the following year and report them to the dean.

- the publication of journal articles according to a clearly defined hierarchy of research journals for every discipline.[8]

The rewards for research performance vary between 750 € for presenting a paper at an internationally recognised conference and 8,000 € for leading a successful bid for a large collaborative research grant. Most interesting in the German context is the attempt to steer the publication behaviour in the faculty by rewarding publication in an international top journal with 6,000 €, in a high ranking journal with 4,000 €, in a good journal with 2,000 € and in an applied journal with 1,000 €.

This scheme is a very weak evaluation system because rewards are paid on demand and non-participation has no consequences. The possible loss of 10% of the recurrent funding is not ideal but also not too punitive for weak performers or fundamental opponents of such systems. Nevertheless, this incentive system makes the Faculty of Social Sciences the leader of the introduction of performance-based funding schemes in the university.

The Faculty of Natural Sciences distributes 10% of its budget to the institutes and chairs according to the performance criteria applied by the ministry to the university (see above). The dean is currently trying to convince his colleagues to accept an incentive system similar to that of the Faculty of Social Sciences.[9] The fund allocation in the Faculty of Humanities has no evaluative component at all. The faculty uses 5% of its budgets for projects in teaching and research without applying performance indicators. An evaluation or incentive system to reward excellent research is neither in existence nor in the making.

Strategic funds for supporting 'profile building' are built at the Faculty of Humanities from 5% of the budget and at the Faculty of Natural Sciences from the maximum possible 10%. The money is used to support special appointments and bonus packages with newly appointed professors whose recurrent funding depends on performance, to support bids for large collaborative grants, and to create research centres that strengthen the faculty's profile. The latter strategy is of particular importance to the Faculty of Humanities because of its size and diversity. In order to save small 'orchid disciplines' with lower enrolment from being closed down by the state government, the faculty founded some research centres that link small disciplines to larger ones, pool resources, and boost research collaboration. These centres are intended to contribute to the profile of the faculty as it is described in the performance agreement with the ministry. However, diversity is declared to be a value of its own in these agreements, and the centres are intermediary umbrella institutions promising diversity *and* profile.

Since strategic funds are too small to boost research profiles, the main instrument of profile building in all three faculties is the recruitment policy. While the deans of all three faculties consider teaching and research to be equally important, increasing

[8] The ranking of journals is undertaken by the disciplines and is then submitted to the faculty. The final hierarchy of journals is discussed and determined by the entire faculty each year.

[9] In addition to this, a peer review of the entire faculty is currently being planned. It will include internal reviews and self-reports as well as external reviews by international peers. This evaluation shall help identifying weakly performing disciplines and institutes, and provide incentives to strengthen enrolment quota or to find back to the cutting edge of research.

attention is being paid to the research profile of candidates. In response to current political expectations, the Faculty of Humanities looks for younger professors with emerging research agendas and the ability to initiate and maintain international collaborative research. The traditional humanities academic, working in solitude and freedom, seems to be a disappearing role model even in a classical German humanities faculty. The Faculty of Natural sciences looks primarily at the candidate's publication record, external funding and 'distinction'. The latter means international excellence in a specialty that later on could become some kind of 'trademark' for the entire faculty in the international competition for students and funding.

ACADEMIC RESPONSES TO THE CHANGING CONTEXT OF RESEARCH

In this section I explore the key changes in the conditions for conducting academic research and how academics in different fields are responding to them. The most important ones are the decreasing time available for research and reductions in research funding. I then discuss how academics are being affected by research evaluations and by the profile-building activities in their faculties.

Time

All interviewed professors said that the most limiting condition for their research is the availability of time. While teaching loads varied between faculties and were not seen as a problem by all interviewees, all professors consistently complained about high and still growing administrative duties. The teaching load of a university professor is fixed at nine contact hours per week. The actual time professors need to spend on teaching depends on student numbers in courses and the resulting time that must be spent on consultations, marking, exams, and teaching-related committee work. Teaching loads are therefore much higher for most academics at the Social Sciences and the Humanities Faculties than for their colleagues in the Natural Sciences.

All interviewed professors adhered to the Humboldtian model of a teaching-research nexus, the political scientists and historians even in spite of their high teaching loads. The reasoning for integrating teaching and research invoked the traditional arguments about research benefiting from teaching, e.g. by advanced natural science students making valuable contributions to research.[10] Much more than teaching all professors saw rising bureaucratisation and the resulting high amount of administrative work as the most significant restriction of their research activities. The bureaucratisation is partly due to the size of this particular university and the resulting number of administrative levels and structures of academic self-governance. Additional administrative loads are produced by:

[10] This is the reason why many of the natural science professors are strongly opposed to the study reform that introduces a two tier Bachelor/Master system. Bachelor students will be useless in research. Many professors expect them to occupy laboratory space and instruction time without ever producing scientifically valuable results.

- the many reform agendas simultaneously pushed by German higher education politics such as the 'Bologna process', 'Excellence Competition', and others which require extensive additional planning, decision-making, and reporting;
- the devolving of previously centralized administrative tasks to institutes and chairs in the context of 'New Public Management'; and
- the bureaucratic drift of research itself. The interviewed professors estimated that today an enormous amount of work needs to be invested to guarantee the success of a funding proposal.[11] All professors who had extensive external funding declared that they hire assistants whose only task is to deal with research bureaucracy and to prepare proposals.

Funding

German university professors are allocated recurrent funding that is determined by their rank. Chairs usually receive funding for one secretary, two full-time research and teaching associates, and some casual staff. They also get a certain amount of start-up money when they are newly appointed. They can save part of that money for times when their recurrent funding is only sufficient to maintain equipment but not to invest into something new. Lower-rank professors may receive money for one associate but in most cases have money only for casual staff and a much smaller amount for equipment than a chair holder. The level of recurrent funding considerably varies between professors of the same rank. They can negotiate increases with the university each time they receive an offer from another university. The recurrent funding details are usually kept a strict secret.

The initial funding of research infrastructure for newly appointed professors usually does not entitle them to later replacements or modernisation of their research utilities. All biologists declared that their research would be in jeopardy were their equipment to fail. The university has no money for replacing the equipment the professors received when they were appointed, and the funding agencies usually do not pay for this equipment because it is considered basic infrastructure that the university should provide.

The extent to which academics could conduct research with their recurrent funding or were dependent on external funding varied between the disciplines. All three history professors declared that it is hard but not impossible to conduct historical research projects in the traditional lone manner without grants. Only one of them held an external grant at the time of the interviews. A second professor had received external funding several times previously but not at the time when we spoke to him, while the third professor never had any external funding at all. Since the recurrent funding does not cover research expenses, two history professors and both interviewed associates reported that they invest private money to get the books, copies, or microfilms they need, or to finance stays in national and foreign archives and libraries. All historians (and the professor of political theory) claimed that the

[11] Projects funded by the European Union are burdened not only with a fierce competition but also with time-consuming contract negotiations and enormous reporting duties during the whole project phase.

scarcity of money spent on the university libraries in this university and in Germany generally is one of the biggest obstacles to efficient research.

The situation of mathematicians was very similar to that of the historians. Two of three interviewed professors were classical 'solitary and free' researchers who conduct low-cost research by solving mathematical problems with 'paper and pencil' and only use a PC for calculations and simulations. They do not have external funds because their basic funding is sufficient to supply them with the few things they need including a small amount of travel money for attending international conferences. Both claimed that doing mathematics is a solitary activity where "... results are found on an intellectual level where you can't delegate tasks", so additional personnel are not necessary for research.

The third professor of mathematics leads an institute for applied mathematics with an unusually high amount of external funding from several sources and many research staff. Most of his projects are funded by thematically focused funding programmes of the federal and state governments. Some others are funded by industry partners. One project evolved from a network initiative of the university's rectorate and received university funding. The professor claimed that he could do research solely with his recurrent funding but only on a much smaller scale. He has built a group that now depends on his success in fund acquisition.

While recurrent funding appears to be sufficient for much historical and mathematical research, the majority of academics from the political sciences, geology and biology depend heavily on external funding. All interviewed political science professors hold external grants, which were deemed a necessary condition for research by two of them. The biologists in our sample conduct the most costly and staff-intensive research. They use laboratories with expensive equipment and need staff to conduct experiments, observe research processes and maintain machinery. Two professors of biology declared that they could not perform effectively without much external funding. The two associates and the recurrent funding of a chair holder suffice to prepare research proposals or to perform one project at a time. However, the professors claimed that they can only perform successful and internationally visible research in the highly competitive environments of their specialties if they can conduct several related projects at the same time.

The third biology professor has external funding through collaborations with industry but none from scientific funding agencies. He is heavily engaged in teaching and uses diploma candidates for his basic and applied research work. For a very limited number of doctoral students he tries to get stipends from renowned industrial foundations. His equipment is less expensive than that of his colleagues but he can generate income by selling side products of his research work – biomaterials which he synthesises – to industry: "This money is much easier to get than funding from the DFG or others."

While geologists depend as strongly on external funding as their colleagues in biology, their access to grants depends on the specialty in which they work. The geologist who conducts research on a universal topic that enjoys increasing international attention had no problems in acquiring external funding. Like the biologists he claims that his basic funds are not sufficient to perform internationally

competitive and visible research. A second geology professor works on an applied topic and attracts significant funding from industry. He is thus able to compensate for the decline in recurrent funding for his specialty from the university, which occurs because the funding is channelled to other specialties in the name of 'profile building'.

The third geology professor we interviewed is in the unhappy situation that none of these reasons applies to his specialty because he works on a regional topic which has no potential for applications. While he conducts low-cost research with "hammer, pocket lens and back pack", he nevertheless needs more research funding than is available from the recurrent funding because he needs to travel to the research sites, to employ doctoral students, and to prepare and analyse his research objects. To cover these costs he applies for grants from the DFG. He complained that this is increasingly difficult and time-consuming because of the increasing competition. When he has no external funding he cuts costs by looking for interesting research objects in the vicinity of the university.

Evaluations and Incentive Schemes

None of the interviewed professors – not even the political scientists whose faculty has introduced an incentive scheme – was under the impression that their research is evaluated by the university or the faculties. With the exception of the professor of political theory, all political scientists said that they adapt to their faculty's incentive scheme by trying to place their publications in the most highly rewarded journals. They saw the incentive scheme as reinforcing existing research behaviour rather than changing it. Two of the three interviewed professors have earned more from the scheme than they had to contribute. The professor who conducts mostly theoretical research is strongly opposed to the reward system for publication behaviour because he considers monographs (which are not rewarded) as the primary medium of academic communication. He uses articles only to publish side products of his research. In his opinion incentive systems like that of the faculty transform a professor from an academic man into a *homo oeconomicus*. The system made him start calculating the time he would need to earn a bit of money and to compare it to the amount of money he could earn by presentations for business firms or foundations. Since the latter activity guarantees higher rewards, he would do that.

The natural scientists were aware of the plans by their faculty to introduce an incentive scheme, and commented on these plans according to the threats they perceived. All mathematicians complained that the specificity of their discipline is ignored in the discussion about performance indicators. Mathematicians produce comparatively few articles. Impact factors of journals are irrelevant because their communities are very small and highly specialised. Their research does not depend on costly equipment or personnel and in most cases does not require external funding.[12] Low enrolment and high dropout rates make things even worse. The

[12] Interestingly, even the professor who is highly successful in acquiring external funding states harsh discontent with the idea that external funding could be an indicator for research activity, saying that "...

other disciplines in the faculty want to use performance indicators for cutting back the budget of the institute of mathematics and redistributing these resources.

Like the mathematicians, the other scientists assessed the coming incentive scheme from the perspective of their own research output. The biologist who works mainly with industry and has low publication output fiercely opposes the idea of impact rankings and citation rates as reward indicators for research success. However, he is not afraid of budget cuts if indicator based evaluation systems are introduced, claiming that the majority of his funds are not university money anyway. The geologist who conducts highly successful research on an internationally attractive topic principally agrees to the introduction of impact factors and citation counts as research indicators. However, he also stressed that this represents only a small part of the *vita activa* of an academic and should not be given too much weight. His colleague is of a different opinion. Publication indicators as a measure of research performance are bad news for him. He argues that most discoveries in his geological specialty are primarily of regional interest and that it makes no sense at all to go to an American journal with the results. So he publishes mainly in regionally focused journals with low to moderate impact.

A commonly shared future expectation of natural scientists[13] is growing stress and lack of time for research if research evaluation with reporting and accountability rituals is introduced and the current structure of academic decision-making and university management remains the same.

Profile-building

While external or internal evaluations have little or no effect, the profile-building activities of the university and the faculties are a force to be reckoned with, at least for the historians and geologists. The historians disagree with the faculty's will to form a profile around a specific topic, which includes the building of 'excellence clusters'. All interviewees complained about the "hype" and saw no possibilities (and no reason) to adapt their own research to that topic. One research centre, which was founded by the faculty in order to meet political demands included in the performance agreements, is focused on the field in which the three interviewed professors work. However, all three professors reported that this centre has no impact on their research. The main effect of the centre, which has a very small budget – just sufficient for occasionally inviting guests – is to boost administrative meetings and taking time away from research through the need to attend presentations. One interviewee observed that from the 40 academics formally involved in the centre only six to ten are regularly seen at centre events.

While the historians are annoyed but so far unaffected by the profile-building of their faculty, one geologist became a victim of such activities because his specialty does not belong to the envisaged profile. The faculty decided to 're-profile' a vacant chair by dedicating it to a different specialty, which is unrelated to that of the

---

this is criminal in an intellectual sense because it only shows that you can grab money". In his opinion the same holds true for citation counts and journal rankings.

[13] Even shared by those who could expect rewards from evaluations or incentive schemes.

interviewee. The idea behind this decision is that the newly established specialty is fashionable and expected to raise a high amount of external research funding, particularly in the thematically-focused programmes of governments and the European Union and in a special DFG-fund dedicated to expensive equipment. Big, opportunistic and cost-intensive science supersedes low-cost science, this interviewee regretfully stated.

According to the performance agreements between university and ministry the social sciences should engage in cluster building activities with a local Max-Planck-Institute (MPI). One of our professors in the political sciences reported that he was asked by his dean if he could be the contact person to develop good relationship between the university institute and the MPI. Cluster building here means that the directors of the MPI become members of the faculty[14] and MPI-researchers get engaged in courses for graduates and Masters candidates. However, these latter kinds of cooperation are at the discretion of the interviewed professor who feels no pressure from above to steer this 'cluster'.

From the biologists only one professor mentioned profile-building activities, which in his case had only a minor impact on research. His institute introduced an evaluation of its own two years ago. With financial support from the rectorate they invited international peers to comment on self-reports and publication lists, listen to presentations and talk to academics. The evaluation was deemed successful, recommendations were made for some kind of restructuring and priority setting but unfortunately they never led to any direct results, mainly because implementing them would have been costly.

Adaptation Patterns[15]

Overall, no systematic adaptation of the content and conduct of research to evaluations could be observed. All interviewees claimed that their own motivation is the key determinant of engagement in research, and expressed the feeling that they can still do what they want and follow the research paths they think are useful. Apart from one exception, no one raised funds for purely opportunistic reasons or changed towards a fashionable research topic because he was forced to by external pressures. All stated that in most cases they conduct research the good old-fashioned way: getting inspired by experiments, by literature study or by colleagues they meet at conferences, then defining a research problem and only then starting to think about funding sources and programme funding schemes. Industry funding plays a certain role but seems to be the smallest funding pillar even for biologists, geologists and applied mathematicians. The only opportunistic behaviour the interviewees reported in some cases is to adapt to the "proposal lyrics" and "catch words" the reviewers of the funding agencies want to hear and read.[16]

[14] Which was a matter of fierce resistance and conflict in the faculty in the past.

[15] The following discussion applies the categorization used by Gläser and Laudel in this volume. Not surprisingly, the occurrences of adaptive mechanisms are almost completely different from those under the strong Australian RES.

[16] The historians did not even do that.

As could be expected from this general observation, there was also little *adaptation to indicators*. A major aspect of this (non-)adaptation is publication behaviour. All interviewees declared that they select journals where they can be sure to address their peers. While nearly all interviewees stressed that they follow discipline or specialty specific assumptions about the quality of journals, impact factors play a role only in biology and in one internationally oriented specialty of geology. But even where impact factors do affect a researcher's reputation, scientists still question whether it is worth the enormous amount of time and work to push an article through the peer review process of a top journal, as one professor in biology stated.

One professor in history prefers either to publish in conference proceedings or to write monographs because the length of his footnotes does not fit the usual journal format, while the political theorist argued that the journal format is inadequate to present a valuable philosophical thought, and that journals would reach a much smaller audience than a well written book in the catalogue of a renowned publisher. For those political scientists with empirical research topics the incentive scheme in place at the Faculty of Social Sciences sustains existing publishing behaviour. In the case of the one professor who agreed that the incentive scheme would influence his publication behaviour, this impact was restricted to reducing the diversity of journals to which he would submit papers. Bearing his own behaviour in mind, he expected such incentive schemes to reinforce a conservative bias by making the top journals even bigger and forcing the others to sink into mediocrity or to disappear from the scene. Such a concentration process could make it difficult to find a place for articles that lay beyond the mainstream in the future.

Opportunistic fundraising and *accommodating patrons* were very rare. Accommodating patrons was only reported by the head of an institute of applied mathematics. The institute's projects in the thematically focused framework funding programmes are attached to political fashionable themes such as multimedia and bioinformatics. The institute's specialty is data analysis and this is very adaptable. If the wind of politics changes "... we change too and in the end we hang our flag towards the new direction because otherwise the funding stops". So this mathematician follows a clearly visible adaptation pattern of opportunistic fundraising that has negative impacts on research of its own. Thematic programmes of governments are discontinued as soon as the next fashionable topic enters the agenda of politicians. When these projects are discontinued no one asks for the results and the research lines are simply dropped. Project staff leave and the scientific knowledge collected over several years gets lost. The professor complains that there is no sustainability in this kind of funding.

The biologist who collaborates with industry partners and sells biomaterials to them is – according to his self-assessment – not an example of opportunistic fundraising or accommodating patrons. He claims that he sells only such products that he produces anyway during his basic research. He would only engage in such industry cooperation that closely fits the basic research he is already conducting: "I only engage in industry related research if it covers our [scientific] self-interest."

However, a common pattern of adaptation reported by nearly all interviewees was the *'stretching'* of research due to the overall scarcity of research funding and

research time. All interviewees except the political scientists (who mentioned it implicitly) stated explicitly that they need to stretch their research. The latter means that academics expanded the time horizons of their research projects significantly. A project with an estimated duration of two or three years could last much longer under the above mentioned unfavourable conditions. In competitive research fields the need to stretch ones research due to a lack of individual research time, good infrastructure or timely too restrictive external funding of additional research staff is a clear obstacle for keeping up with the cutting edge of research.

Of all interviewees, only one geologist reported that he has *downsized* his research. He adapts to the scarcity of funds by selecting research objects (landscapes and formations) he can afford to travel to. Implicit downsizing of research happens in many natural science specialties due to ageing equipment and a lack of money to invest in new apparatus.

An adaptation pattern that was not observed in Australia but is expressed by several German professors in the natural sciences and by one historian, who all are in the decade before retirement, is '*inner emigration'*. Some of these professors terminated their positions on the boards of academic self-governance and gave up fighting unfavourable developments they felt they could not influence. They expressed the feeling that the actual changes towards the Bachelor/Masters study system and the rewarding or sanctioning results of publication counts and other tools of a coming research evaluation regime would hit primarily the younger colleagues. At the end of their careers they feel no necessity to adapt to mechanisms they do not agree with.

'Inner emigration' does not mean giving up research but withdrawal from academic self-governance and avoiding administrative duties. For example, one of our geologists founded an 'associated institute' and shifted all his research activities to that institute[17] . This professor had the impression that his specialty has no future due to declining student enrolment, which triggered severe budget cuts, the removal of supporting staff like secretaries and the declared will to distribute vacant chairs in his specialty to other disciplines when chair holders retire. In his associated institute he can create his own research environment and produce results in applied research for industry with the staff of his choice. He is now moving more and more project activity from the official university system to the associated institute, which he runs together with a retired colleague.

This example shows that good access to external funding sources can make individual professors relatively independent from further budget cuts in the university. For many natural scientists, recurrent funding is already so small that, as one biologist stressed, they cannot perform their teaching without using the personnel paid by external research grants. If a research evaluation regime is introduced, the professors who conform to the indicators might be rewarded with

[17] An associated institute is a research institute that is attached to the university but works under private rather than public law.

'peanuts' while those who do not conform but hold external grants do not care because they have turned their attention to other sources.[18]

CONCLUDING REMARKS

All the interviewed professors claimed that they can still conduct intrinsically motivated and curiosity driven research. While several interviewees adapt to limiting conditions for their research by stretching their projects into the future, none of the interviewees felt that these conditions had an impact on the contents of the research or changed his research track in unwanted directions. The majority of professors who claimed to need external funding for conducting research were successful in the competition for grants. The choices of research topics, objects and methods are hardly restricted by the conditions for research, and the diversity not reduced by external influences.

A new major impact on research conditions that has emerged recently and can be linked to the emerging weak RES is 'profile-building'. This occurs in two contexts, namely at the national level as the trend towards topical funding of big collaborative projects and within universities as a concentration of resources on certain topics and collaborations. This profile building, which appears in other RES as well (see Gläser and Laudel for Australia and Schiene and Schimank for Lower Saxony in Germany, this volume) has the potential to limit the choices of research topics available to academics.

But even if a weak RES like the kind introduced in the Faculty of Social Sciences in our case study is applied, its steering impact on professors is expected to be low. Those who conform with the set of indicators already show the behaviour that the system wants to achieve. Their intrinsic behaviour was merely rewarded and amplified. The others who do not conform and can be supposed to be the real targets of the system will probably not change for several reasons. Firstly they have tenure and cannot be dismissed. Nor would it be legally possible to redefine their tasks, remove their basic funding completely, or raise their teaching or administration loads, etc. The most likely scenario is that they get sidelined and turn to 'inner emigration', which would be the opposite of what the evaluation regime is intended to achieve.

Secondly some of these professors have no chance at all to conform to the usual indicator based standards because of the cognitive structures of their disciplines or of their specialties. These professors do not need external funding, address only small scientific communities (which means low-impact publications), or use publication forms other than journal articles. These academics cannot change even if they want to. For this group the steering effect of an incentive system would be close to zero. Thirdly, a maximum loss of 10% of basic funds for weak performing professors (due to the indicator set) is a sanction most of them can be expected to

[18] The same pattern holds true for the political theorist who claimed to weigh the incentives his faculty could give him against the money he could earn from lectures and presentations outside the university.

live with. As Minssen et al. (2003: 89) stress, a performance based share of 10% from total funds is too low to have significant intellectual effects.

Given the limitations for performance-based funding of chairs and the opportunities for German university professors simply to ignore the various RES and profile-building activities, it is difficult to imagine an effective RES for German universities without further, more radical, changes. The major instrument through which change in German universities can be achieved appears to be the appointment of new professors who (a) fit the planned new profiles of universities and (b) can be seriously punished for bad performance by being stripped of their recurrent funding. It is still unclear what effects the limited introduction of the currently weak RES will have for the generation of recently employed professors who have tenure but are subject to performance assessment, with parts of their basic funding appointments and salaries depending on a positive evaluation of their activities. At the moment, it has no effects at all because standards and procedures are not defined and additional money for significant salary increases is not available. In general, in Germany the efforts towards research evaluation across the sixteen states, between the states and 'their' universities and even – as the case study shows – between the faculties under the umbrella of one and the same university are still too diverse and fragmented to come to a serious prognosis about the future impact of research evaluation on the contents of university research.

REFERENCES

DFG [Deutsche Forschungsgemeinschaft] (2006), *Jahresbericht 2005. Aufgaben und Ergebnisse*, Bonn: DFG.

Fangmann, Helmut (2006), 'Hochschulsteuerung in Nordrhein-Westfalen. Strukturen und Instrumente, Sachstand und Perspektiven', *Beiträge zur Hochschulforschung*, 1/2006, 54-65.

Hartwig, Lydia (2006), 'Neue Finanzierungs- und Steuerungsinstrumente und ihre Auswirkungen auf die Universitäten. Zur Situation in vier Bundesländern', *Beiträge zur Hochschulforschung*, 1/2006, 6-25.

HRK [Hochschulrektorenkonferenz] (1993), *Zur Forschung in den Hochschulen. Stellungnahme des 170. Plenums vom 12. Juli 1993*, http://www.hrk.de/de/beschluesse/109_485.php?datum=170.+Plenum+am+12.+Juli+1993 (accessed 22 May 2007).

Jaeger, Michael (2005), 'Leistungsbezogene Mittelvergabe und Qualitätssicherung als Elemente der hochschulinternen Steuerung', *HIS-Kurzinformationen*, A/12/2005, Hannover: HIS.

Jaeger, Michael, Michael Leszczensky, Dominic Orr, Astrid Schwarzenberger (2005), 'Formelgebundene Mittelvergabe und Zielvereinbarungen als Instrumente der Budgetierung an deutschen Universitäten: Ergebnisse einer bundesweiten Befragung', *HIS-Kurzinformationen*, A/13/2005, Hannover: HIS.

Kehm, Barbara and Ute Lanzendorf (2006), 'Germany – 16 Länder Approaches to Reform', in Barbara Kehm and Ute Lanzendorf (eds.), *Reforming University Governance. Changing Conditions for Research in Four European Countries*, Bonn: Lemmens, pp. 135-186.

Krull, Wilhelm and Simon Sommer (2006), 'Die deutsche Vereinigung und die Systemevaluation der deutschen Wissenschaftsorganisationen', in Peter Weingart and Niels C. Taubert (eds.), Das Wissensministerium – Ein halbes Jahrhundert Forschungs- und Bildungspolitik in Deutschland, Weilerswist: Velbrück, pp. 200-235.

Kuhlmann, Stefan and Thomas Heinze (2004), 'Evaluation von Forschungsleistungen in Deutschland: Erzeuger und Bedarf. Teil I: Konzeptionelle Grundlagen', *Wissenschaftsrecht*, 37, 53-69.

Lange, Stefan and Uwe Schimank (2007), 'Zwischen Konvergenz und Pfadabhängigkeit: New Public Management in den Hochschulsystemen fünf ausgewählter OECD-Länder', in Katharina Holzinger, Helge Joergens and Christoph Knill, (eds.), *Transfer, Diffusion und Konvergenz von Politiken*, PVS-Sonderheft 38, Wiesbaden: VS, pp. 522-548.

Leszczensky, Michael and Dominic Orr (2004), 'Staatliche Hochschulfinanzierung durch indikatorgestützte Mittelverteilung. Dokumentation und Analyse der Verfahren in 11 Bundesländern', *HIS-Kurzinformationen*, A/2/2004, Hannover: HIS.

Mayntz, Renate (1994), *Deutsche Forschung im Einigungsprozess: Die Transformation der Akademie der Wissenschaften der DDR 1989 bis 1992*, Frankfurt a.M. and New York: Campus.

Minssen, Heiner, Beate Molsich, Uwe Willkesmann and Uwe Andersen (2003), *Kontextsteuerung von Hochschulen? Folgen der indikatorisierten Mittelzuweisung*, Berlin: Duncker & Humblot.

Röbbecke, Martina and Dagmar Simon (2001), *Reflexive Evaluation. Ziele, Verfahren und Instrumente der Bewertung von Forschungsinstituten*, Berlin: Edition Sigma.

Schimank, Uwe and Stefan Lange (2006), 'Hochschulpolitik in der Bund/Länder-Konkurrenz', in Peter Weingart and Niels C. Taubert (eds.), *Das Wissensministerium – Ein halbes Jahrhundert Forschungs- und Bildungspolitik in Deutschland*, Weilerswist: Velbrück, pp. 522-548.

Wolf, Hans-Georg (1995), 'An Academy in Transition: Organizational Success and Failure in the Process of German Unification', *Social Studies of Science*, 25, 829-852.

CHAPTER 8

CHRISTOF SCHIENE AND UWE SCHIMANK

RESEARCH EVALUATION AS ORGANISATIONAL DEVELOPMENT

The Work of the Academic Advisory Council in Lower Saxony (FRG)

INTRODUCTION

In a recent letter to the *Wissenschaftliche Kommission Niedersachsen* (WKN, in English 'Academic Advisory Council' - AAC) of Lower Saxony a professor of engineering expressed in harsh words his personal frustration about his recent experiences of research evaluations. In his opinion, evaluation of research at German universities is "useless, counterproductive, and uneconomical". The goals which are pursued with evaluation exercises are "popular", but cannot be reached in this way. On the contrary, evaluation amounts to an "excess of bureaucracy", and it "discourages" able and willing researchers. Those professors who act as evaluators of others are, in his view, collaborating with the enemy. They show "a lack of character, for instance a lack of regard for cooperativeness among professional colleagues". Instead of all this "parasitic formalism", he has a quite simple answer to the question of how good research is produced:

> The personal joy of researching a specific subject and – if everything goes well – about the results motivates high performance. Success is not determined by assessments and examinations. The real feedback loop demands more staying power from a researcher than a formalised administrative act. For example, one has to wait until research results are adopted by other researchers or utilised in an industrial production process. I don't make an extra effort in order to be positively evaluated or to increase my pay.

Summing up his arguments, the author insists that the financial costs of evaluations would be much better invested in more research assistants at the universities.

Another letter, written by a former university president, came to quite different conclusions about the effects of regular evaluations. From this perspective there are several very positive, indeed desperately needed effects, each of which is beneficial on its own as well as in connection with the others. To begin with, "the preparation and analysis of the evaluation exercises has considerably changed the communication

R. Whitley and J. Gläser (eds.), The Changing Governance of the Sciences, 171–190.

within universities, mainly between the university leadership and the academic disciplines, but also between academics from the evaluated disciplines". Secondly,

> the necessity to prepare periodic evaluation exercises (...) and to analyse their results (…) led to the creation of a very forceful administrative unit for this task. At the level of faculties and disciplines the role of deans and heads of institutes changed insofar as they became involved (...) in questions of development planning, change management, and quality assurance.

Thirdly, the former university president emphasises that

> if universities were required to prepare informative data about research on a regular basis, problematic cases might be identified by analysing this data and could be subjected to closer investigations.

His experience is that the results of the AAC's evaluations give to the university leadership

> … valuable argumentative support for introducing measures of strategic profile-building in the university, deciding about the reallocation of resources, establishing an internal system of selective support of research activities, and presenting the performance of the different disciplines to the public and the Ministry of Science and Cultural Affairs.

Both letters[1] use by now quite familiar arguments: the latter totally pro, the former totally contra evaluation. What is interesting about this clash of opinions – which, by the way, is quite representative of current debates within the German university scene – is that both protagonists do not totally disagree about the basic facts, with respect to what evaluation is and how it proceeds; they disagree about the actual immediate effects of evaluation; but they disagree most in their assessment of the long-term effects of evaluations on the kind of research undertaken in German universities: What will happen over the years, and is it desirable?

We will not be able to give here decisive answers to the questions of which (if either) observer of the German university system is right, and to what extent. Instead, we think this fundamental disagreement is in itself a rather interesting indication of the changes which are beginning to spread in Lower Saxony's universities as a result of the work of the AAC.

Germany is a late-comer with respect to evaluations. By now all federal states (*Länder*) - which are responsible for their universities - have begun evaluations of teaching and, to a lesser degree, research. In a few states, among them Lower Saxony, evaluation agencies have been established.[2] Evaluation methods and criteria differ considerably. In most cases, some kind of peer review is an important part of the procedure, but there are also examples of indicator-based formulas, mechanically used to distribute parts of public funding to universities. The AAC undertakes a type of research evaluation based on informed peer review where no immediate and mechanical financial consequences are tied to the assessments of research units and

[1] Both unpublished. All quotations we make from documents and reports are translated by us.

[2] For a descriptive overview of evaluation activities in Germany see Kuhlmann and Heinze (2004a; 2004b); a case study is presented by Lange in this volume. In the beginning of 2006 the German Science Council started an ambitious research rating of all German universities with two disciplines, chemistry and sociology, as test beds for procedure and criteria.

individual researchers. For this kind of evaluation the AAC is the most ambitious representative in the German scene.[3]

What we would like to focus on in this contribution is one specific element of the research evaluation undertaken by the AAC. We will not deal with questions about the quality, relevance or efficiency of research which play a prominent role in the AAC's evaluations. Instead, we will pick out another important element of contemporary rhetoric of research evaluations: the insistence on profile-building and critical masses of researchers collaborating in a specific research area. We will show, firstly, that this element is a built-in bias of the AAC's evaluation procedure; secondly, it is an explicitly stated idea of 'good' research for the future; and thirdly, it is a concrete result of this kind of evaluation.

We will take our empirical illustrations mainly from two very different scientific disciplines: history and chemistry, as the results of evaluations in these areas are well documented.[4] Whilst they are not extreme contrasts they belong at opposite ends of the spectrum in many respects.[5] Chemistry is a natural science strongly tied to extra-scientific criteria of utility and users of its knowledge, whilst history belongs to the humanities and conducts primarily curiosity-driven research. Chemistry is strongly internationalised, history is not; chemistry is strongly dependent upon external funding, history is not; chemistry often requires substantial resources and extensive co-operations, history is usually modest in both respects.

PROCEDURE

Evaluation is an integral element of the new mode of university governance often epitomised as 'new public management' (NPM).[6] The basic goals pursued in this profound transformation of university governance are efficiency and relevance. The former refers to the scarcity of public funds available for university research, which demands to allocate money in a less wasteful way. In addition, university research is pressed to have a higher extra-scientific return on investment – not only, but especially with respect to the ongoing economic crisis of Western countries. An insistence on efficiency as well as on relevance have been essential elements of the new framework of legitimation of research policy since the late 1970's.

To summarise the program of NPM: An increased competitive pressure within and among universities, produced by various institutional mechanisms as well as by an increased scarcity of available financial resources, should result in a new emphasis on efficiency and quality of research. In addition, intensified external guidance of universities requires extra-scientific relevance of research; deregulation

---

[3] For a comparison of the AAC with the Dutch, the British, and the Irish evaluation exercises see Orr (2003).

[4] As data sources we use several partly published, partly still unpublished reports of the AAC. In addition, we use data from a recent meta-evaluation of the AAC (WKN 2006).

[5] See also the phenomenology by Becher and Trowler (2001) of the different 'academic tribes' to which both disciplines belong.

[6] See De Boer et al. (2006) for an overview of these governance changes in four European countries, among which Germany is the latecomer, and more extensively the German country report by Kehm and Lanzendorf (2006).

with respect to a removal of externally imposed bureaucratic strait-jackets is one of the pre-conditions to master these changed external circumstances; another one is the strengthening of a hierarchical self-management of universities by strong rectors and deans and a weakening of academic self-governance.[7] This amounts to the building-up of a university as a corporate actor capable not only of surviving but of prospering in a highly competitive environment (Krücken and Meier 2006; see also Weingart and Maassen in this volume).

NPM is frequently associated with the slogan 'more market!' This is true, indeed, if tuition fees are introduced as a major source of university finances, or if contracts and other kinds of third party funding of university research become more and more important, as is the case in Germany as everywhere else. Under these circumstances students, business firms, ministries and funding agencies are buyers who decide which of the competing universities, faculties, or professors they give their scarce money to. Scientific journals, too, are, in a broad sense of the word, buyers of articles. As is well known, university scientists are competing strongly for the scarce publication opportunities offered especially by top journals.

However, even if all of these 'markets' are extended a considerable part of university finances are delivered as yearly block grants from government in most European university systems. These block grants provide for the basic infrastructure of universities: salaries of personnel, costs of buildings, libraries, laboratories etc. If the government wants to allocate these block grants according to the relative performance of universities, and if the university leadership wants to do the same with respect to faculties, institutes, or individual professors, both need reliable information about the relative standing of these performers. In addition, such performance information is also very helpful to students who have to decide where to study or to business firms or other extra-scientific actors who must decide to whom a research contract is given.

This is the point where evaluation comes into play. Evaluation exercises provide decision-makers such as ministries, university presidents, deans, students and others with information about the relative standing of evaluated units – universities, faculties, institutes, or professors – in terms of their teaching and/or research performance. Those who decide upon the allocation of block grants basically have two possibilities of how to make use of this performance information which Richard Whitley (in this volume) distinguishes as 'strong' or 'weak research evaluation systems':[8]

- 'Strong' research evaluation ties immediate and schematic financial consequences to relative performance, as in the formula-based government funding of Australian universities (See Gläser and Laudel in this volume). Thus, 'strong' evaluation works as a mechanical extension of market-forces: The allocation of block grants follows without any discretion where the 'invisible hand' of the various 'markets' mentioned shows the way. Strong performers in terms of attraction of students, research grants and contracts are rewarded, weak performers are punished, with the highly

[7] See Brinckmann (1998) as the most explicit proponent of NPM in the German debate.
[8] See also Orr (2003: 24) about 'high stakes' and 'low stakes' evaluations.

irresistible consequence pointed out as the "Matthew effect" (Merton 1968) that strong performers become stronger and weak performers become weaker over time.

- The other possible use of performance information is made in 'weak' evaluations. Here, decision-makers exert a 'visible hand' (Chandler 1977) and explicitly decide case by case whether they simply follow what the 'markets' point to or whether they deviate. Deviation may occur in two directions. To amplify market dynamics, good performers may be 'over-rewarded' and even average performers, not just bad ones, may be strongly punished. However, more interesting is the case where bad or average performers are given a chance to improve. In this respect, David Campbell (2005: 15) speaks about the initiation of organisational "learning processes" by evaluation.

We are interested here in those evaluations which aim at learning in the broad sense of organisational development. This has to take place within the evaluated unit, but can be accompanied by more or less support from outside – including, besides money, decisions which install certain preconditions for performance improvement. Learning how to improve performance can be very much supported by relevant information about possible causes of deficits. Thus, a kind of evaluation which produces only indicator-based information about the level of performance reached by an evaluated unit does not give any hints about where to look for possible changes which could bring about an improvement. However, a qualitative peer review – supported by appropriate quantitative indicators - which not only judges research performance but includes an assessment of research conditions, will often point out probable points of intervention quite precisely. Moreover, such an assessment should help to establish whether poor performance is caused by factors that can be changed or by circumstances beyond the university's control.

The evaluation procedure of the AAC belongs to this type. Its explicit intention is to direct and stimulate the action of the 'visible hand' of organisational leadership and, where necessary, government. The respective ministry, university presidents, deans, heads of institutes and other actors responsible for the university or relevant sub-units, shall do something to improve or maintain the level of research performance; moreover, they are told what is to be done (Orr 2003: 61). This presupposes that the information gathered and condensed gives an overall qualitative picture of an evaluated unit with respect to the relations between input, throughput, and output.

The procedure outline of the AAC makes sure that this goal is pursued (WKN 1999; Orr 2003: 28-35). The evaluations are discipline-oriented. They evaluate all university departments or institutes of the selected discipline in Lower Saxony. The usual order of events is as follows: a short framework paper is provided by the AAC to the evaluated disciplines within the universities, to help them prepare a report on the last five years of research activity and future planning. Universities are then visited by the group of evaluators; approximately six professors from the evaluated discipline in other German federal state or foreign universities. These evaluators speak to the university president, the respective dean, each professor of the discipline, some members of scientific staff, and some doctoral students, and discuss

their findings. Based on these discussions, a draft report about the discipline and its relative performance at all Lower Saxony's universities is written by the evaluators and edited by the official in charge from the main office of the AAC.

The evaluated units and individuals are then asked for their comments, which reach the AAC via the president of the university. On this basis, the final report ('assessors' report'), which contains evaluations and recommendations, is written and submitted by the evaluators to the AAC. Typical recommendations include the re-dedication of professorships, the participation of external colleagues in the recruitment procedure for vacant professorships, the establishment of special study programs for doctoral students, or the improvement of library or laboratory conditions. The AAC discusses the report and its recommendations and this is published, with the exception of the evaluations of individuals which are given to the individual and the university president, with a complete copy sent to the ministry. A follow-up proceeds, which may consist of several measures according to recommendations. One standard instrument of re-evaluation is an intermediate report after three years.

Departments or institutes can be evaluated on two levels: as a whole, and as smaller 'research units', self-defined by the researchers according to local and disciplinary conditions. A 'research unit' can range from a team of scientists (e.g. in the natural sciences) to an individual chair (e.g. in the humanities). The research performance of each unit, measured by international scientific standards and conventions within the respective discipline, is always assessed, first on an absolute basis and secondly, on a relative basis in comparison to the performance of national and regional competitors. Thus, the evaluation report gives a detailed overall and comparative picture of the respective discipline at universities in Lower Saxony. It assesses not only the activities of each 'research unit', but discusses as well the discipline's situation at each university visited and the overall situation of the discipline in Lower Saxony.

By the end of 2005, the AAC – established in 1997 as an independent body of experts that advises Lower Saxony's government in all science policy-related questions – had completed 25 exercises of research assessment within universities, starting with chemistry and history in 1999/2000 (WKN 2000a; 2000b). A key principle underlying these evaluations is the endeavour to take into account the specificities of scientific fields like engineering, the social sciences, the humanities or the natural and life sciences, and even of particular disciplines or, if necessary, interdisciplinary fields (Schiene 2004). While the procedure outline mentions mainly four general fixed criteria (quality, relevance, effectiveness and efficiency) with some more specific underlying indicators, the group of evaluators has to operationalise its discipline-specific definition of 'good' research. This adaptability of the framework of performance criteria manifests itself also in the evaluators' weighting of the selected indicators.

This procedure of evaluation, on the one hand, gives a strong position to the respective disciplinary community, represented by the peers. Obviously, this fact contributes to the legitimacy of the AAC's evaluations, in contrast with sometimes quite superficial quantitative evidence collected by ministries (see Weingart and Maassen in this volume).

On the other hand, with respect to our special focus it must be emphasised that the evaluation procedure of the AAC, in line with other elements of NPM, presupposes and, in this way, helps to create the university as an organisation with a strong capacity of collective action, at university or sub-unit level. Traditional evaluation activities of the science system rely strongly on the individual and the scientific community. This is the case in formalised procedures in the peer review of candidates for professorships, project proposals or manuscripts for publication, and more importantly, in the totally decentralised and casual evaluation of publications and their authors by quotations and comments (or non-quotations) in someone else's publications or lectures.[9]

Here, the individual researchers are responsible for the quality and relevance of their work, and they submit their finished work to the scientific community who will incorporate it into either the most accepted or least contested body of knowledge or, most of the time, immediately forget it.

Now, however, the organisation becomes an important actor in evaluations. The university is supposed to draw conclusions from the evaluation of its professors, and its leadership is often determined to do so. This still appears new and strange to German professors, as shown by their anger regarding this imposition (see the scientist quoted in the introduction). Some of them even see their constitutionally guaranteed 'freedom of science' violated by these new evaluation exercises.[10]

Until recently no German university president had to care much about the research performance of his university's professors, at least not with respect to block grants from government. Nowadays, and even more so in the future, the research performance of individual professors is of vital interest to a university. Thus, a university president cannot but react to performance information he gets from evaluations. If the evaluation procedure is a 'weak' one so that organisational development is possible, he must demand improvements from weak performers if financial consequences for the university and the performer are to be avoided. Furthermore, if the evaluation procedure reveals information about causes of deficits, and recommendations about what should be done about them, he can confront the weak performers with this interpretation of their situation.

The evaluations undertaken by the AAC deliberately serve this purpose. Against this background, it is important to study more closely the idea of good or even excellent university research which is pursued in these evaluations.

RESEARCH IDEALS

A descriptive overview of typical recommendations of the AAC in various disciplines shows what evaluators perceive as important requirements of improved research performance. The major recommendations can be briefly listed as follows:

[9] See Gläser's (2006) systematic account of scientific disciplines as 'production communities'.

[10] However, the German constitutional court recently declared evaluations – with certain provisions – as legal measures of science policy (Bundesverfassungsgericht 2004).

(1) establishment of new professorships, re-dedication of vacant professorships within the discipline, elimination of vacant professorships or their transfer to a different discipline;[11]
(2) participation of external peers in the recruitment commissions for vacant or new professorships;
(3) additional scientific staff for professorships;
(4) reduction of permanent scientific staff in favour of temporary employment contracts with younger scientists;
(5) study programs for postgraduates;
(6) additional financial means from the government;
(7) a more performance-oriented allocation of block grants within the university or faculty;
(8) increased acquisition of project grants or research contracts;
(9) infrastructural improvements of buildings, libraries, laboratories; and
(10) intensification of internal and external coordination and cooperation.

There are no big differences between all the disciplines the AAC already dealt with, and chemistry and history in particular, with respect to what is recommended.[12] This finding is not at all trivial. One might have suspected, for instance, that more stress is laid on external funds in chemistry than in history; but this is not the case although, of course, the actual reality of both disciplines strongly differs along these lines. Despite factual heterogeneity of research practices, roughly the same recipes for good research are issued by evaluators. We will reflect upon this surprising fact later, after we have taken a closer look at the idea of good research underlying this list of recommendations.

A careful reading of the evaluation reports displays a cognitive and evaluative construction of very few pillars which carry easily and firmly all the mentioned specific recommendations. Three basic prerequisites of good university research are explicitly stated by the evaluators from which the recommendations follow as logical consequences. Two of these prerequisites can be mentioned here very briefly: first of all, a primacy of inner-scientific criteria of research quality as against criteria of extra-scientific relevance,[13] and secondly, a reduction of teaching

[11] The latter measure strengthens not the discipline which is evaluated but another discipline. Of course, this – from the point of view of the evaluators as disciplinary peers – 'altruistic' recommendation is only given in 'hopeless cases'. The first-mentioned measure can strengthen an evaluated unit if the disposable financial resources are used to equip its other professorships better, for instance with scientific staff.

[12] Only a few, and mostly trivial, recommendations are discipline-specific. For instance, it is no surprise that in chemistry more stress is laid on technical infrastructure than on libraries whereas the latter are of special importance for historians.

[13] Despite all the rhetoric and demands, last but not least in the proclamation of 'mode 2', that curiosity-driven research is 'out' and the use-value of research for society is of overriding importance evaluators of the AAC share the traditional concept of the clear primacy of inner-scientific as against extra-scientific criteria of research quality. Of course, this does not exclude that relevance for all kinds of extra-scientific needs – from demands of industry to medical or even military demands – is seen as a very important, but not as the most important criterion, after all. To make this understanding of research as concise as possible: Whenever extra-scientific relevance is of high importance, inner-scientific quality must be even more important because it is seen as a precondition for the realisation of relevance.

pressures in favour of time available for research work.[14] These two pillars of the idea of good university research need no further comment here because they have been conventional wisdom for a long time. But there is a clearly visible third pillar now which is quite new and still debated at German universities. Good research, according to this view, expresses itself in a distinct profile of an institute or a faculty, and this in turn requires a critical mass of cooperating researchers equipped with adequate resources within an evaluated unit.

The traditional view of good university research in Germany is related to a single person – each individual professor – as the unit which has to build up its profile;[15] if it happens that the institute or even the faculty displays a profile, too, this is either a welcome by-product of individual profile-building or a voluntary 'joint venture' of several professors but nothing which the professors are normatively expected to pursue in a collective effort. In other words, individual profiles suffice according to the traditional view. Against this position the AAC and its evaluation groups posit the necessity of an up-scaling to collective profiles. This starts with the procedural outline of the AAC: Two of the five goals to be reached by the evaluations refer explicitly to collective research profiles (Orr 2003: 29). Again and again, evaluation reports communicate that it is not enough even if all individual professors in a particular institute or faculty are excellent researchers:

> Today, even in philosophy a successful research enterprise requires a minimum of willingness to cooperate with academics researching other subjects within the same institute, in an interdisciplinary context with other academics in the university, or with research institutions and projects nationwide and international (…) Of course the assessors do not rule out the possibility of important individual benefits (…) However, experience shows that even in philosophy quality and methodological innovation of research usually correlate with the promotion of promising academics and the engagement in research networks. (WKN 2005: 11 – Assessors' report - Philosophy)

It is not an exaggeration to state that in this view the first mentioned pillar is only solid in connection with the third one: No real research quality without collective profile! The evaluators claim not only that the whole shall be more than the sum of its parts but that, under any circumstances, the sum of the parts is not enough. Moreover, this collective profile is supposed to differ from the research priorities elsewhere, at least between the inspected universities:

> This evaluation shall serve to realize the possible advantage [of research variety] as it recommends a profile building that is even more determined, distinct, and marked, and a corresponding setting of research priorities at the various institutions. (WKN 2000b: 8 – Assessors' report - History)

At the very least there has to be a common research agenda within an institute or faculty so that the topics of individual researchers are related to each other in a meaningful way:

> Those widely acknowledged research activities in the period under review originated mainly in the individual interests of academics, but do not refer to each other, and they were only in a few cases organized in research projects. [T]he development of a

[14] See Schimank (1994; 1995) for an analysis of the inherent danger of a marginalisation of research by a growing teaching load in the institutional fabric of the German university system.

[15] Actually, this individual represents and directs the small team of a professorship.

> research profile was not appreciated as an important task of the discipline. (WKN 2004c: 36 – Assessors' report – German Language and Literature)

But even better are research co-operations:

> Establishing new and interdisciplinary research fields crucially depends on the intramural cooperation between the institutes of chemistry, which traditionally work rather independent and isolated from each other. It also depends on an increase in cooperation with other disciplines in the university. (WKN 2000a: 8 – Assessors' report - Chemistry)

Thus, the 'loose coupling', sometimes amounting to total disconnectedness, of research activities which has been and still is a predominant feature of university research in many institutes and faculties shall be transformed into a more tightly coupled network:[16]

> It is not inconsistent if the assessors recommend to the institute to look more closely at its profile and research priority setting (which probably requires more exchange and cooperation among its units and individual members, who often work in strong isolation from each other. (WKN 2000b: 12 – Assessors' report – History)

Or:

> The point is also the extent to which an institute succeeded in developing a specific profile through internal and external co-operations, a profile that is recognizable from the outside. (WKN 2005: 9 – Assessors' report – Philosophy)

Evaluators are aware of the fact that their insistence on collective profile-building strikes quite a new chord in German universities. This demand is not yet self-justified. At several places in the reports two reasons are given for the demanded transformation from traditional 'small is beautiful' to 'large is beautiful'. First, evaluators see an irresistible inner-scientific dynamic towards larger-scaled research problems and activities in principally all disciplines, not just in particle physics or parts of the life-sciences:

> Given today's diversity of research areas and scientific fields, it is necessary (…) to study complex research questions in a network with colleagues from one's own university, other national institutions, and international institutions. Cooperation (…) should therefore be among the central goals of a research unit. (WKN 2002: 11 – Assessors' report – Law)

Secondly, evaluators perceive an equally irresistible extra-scientific dynamic towards 'bigger questions' directed at all disciplines. This relates not just to inter-disciplinary problem fields such as climate change, but also to many disciplinary topics:

> Many urgent questions in modern society are so complex that they can only be answered by the collaboration of different disciplines. This is why project-oriented and organized research, and thus interdisciplinarity and cooperation in all social sciences, have become more important. (WKN 2004b: 13 – Assessors' report - Political Science and Sociology)

[16] For similar tendencies in the Dutch research policy such as the establishment of 'research schools' see van der Meulen (in this volume). See again Lange (in this volume) for the status quo in a traditional German university.

Or:

> [The complex problems which many diseases raise] require intense research collaboration among divisions and research institutions, which contribute different methodological expertise to the process of problem solving. Those collaborations are most successful when they occur in overlapping institutionalized research networks that are jointly funded. (WKN 2004a: 19 – Assessors' report – Medicine)

Both dynamics reinforce each other. Answering 'bigger questions' demands larger-scaled research approaches, and the latter stimulate the former.

We cannot discuss here how valid this assessment of science-society dynamics really is. For our purposes it suffices to say that it is shared by the large majority of evaluators of the AAC; and we have no reason to think that the AAC selects its evaluators by their adherence to this view. Rather, we suspect that the elites of all disciplines think that this view gives a correct picture of what happens in and with contemporary science. Even if this view were just a self-fulfilling prophecy it would work as forcefully as it actually does.

From a neo-institutionalist perspective, these two interrelated reasons are good examples of fictions of rationality (Schimank 2005b: 372-393). Taking the two dynamics for granted, it is only reasonable to insist on critical masses and profile-building. But these kinds of assertions oscillate between truth and prejudice. The latter aspect is stressed if one speaks of 'myths' (Meyer and Rowan 1977). Indeed, rationality fictions always pretend to be proven general truths although they have, at best, an inconclusive scientific backing for a limited sub-class of relevant phenomena. Is it really true for all disciplines, and for all research fields, that inner- as well as extra-scientific dynamics present ever 'bigger questions'? There are certainly examples which can be presented – but are they exceptions or the rule? Are such tendencies perhaps only a temporary phase which will fade away soon? Nobody knows for sure – but many assume to know the truth, and mutually reinforce each other in this belief, and moreover, in the belief that the respective rationality fictions are derived from empirically confirmed theories.[17] Some rationality fictions even lack any scientific evidence at all. Nevertheless, the standing of a scientific truth is often proclaimed for rationality fictions because in modern society this is the best safeguard against any scepticism, and against being criticised in case of failure. As something believed to be a scientific truth, a rationality fiction is effectively reified. It is treated as if it were firmly based on 'the way things are'. So it becomes, in effect, an intersubjectively shared routine of thinking and acting.[18] A decision based on what everybody holds as true may not reach its aim; but this failure then can be attributed to unfortunate external circumstances, and not be considered a wrong decision. In this way, rationality

[17] This was shown very clearly for many recent management fads (Kieser 1997). NPM, by the way, belongs to them.

[18] By definition, a routine is never rational in the sense of "procedural rationality" (Simon 1982) because it 'jumps' to its conclusion instead of a painstaking and time-consuming reflection of the respective problem and the alternatives to deal with it. Thus, even in cases where a careful consideration would have resulted in the right recommendation of building a 'critical mass' usual evaluation procedures which almost never make this effort rely on a rationality fiction.

fictions are a good device for 'blame avoidance' (Weaver 1986) which is an overriding concern of all kinds of decision-makers.

Many of the recommendations given by the evaluators are to be understood as measures of profile-building. This is as obvious for additional professorships with targeted dedications or for changes of dedications of existing professorships as it is for the targeting of additional scientific staff, finances, or infrastructure. And of course the frequent emphasis on co-operations goes explicitly in the same direction. Evaluators hope that a reciprocal causality is initiated: Co-operation leads to profile-building which leads to more co-operation etc.

The question remains: Why do evaluators in history, in sociology or even in philosophy emphasise this third pillar of the idea of good university research as much as evaluators in chemistry or in medicine do, although actual research practices of the latter disciplines fit to this idea much closer than those of the former? The explanation of this surprising fact could lie in the selection of evaluators. What all of them have in common is their high scientific reputation. It seems to be the case that the homogeneity of the idea of good research is rather high among the elites of heterogeneous disciplines. This fact, in turn, can be traced back to five complementary explanations:

(1) Firstly, in elite networks there occurs much 'mimetic' and 'normative isomorphism' (DiMaggio and Powell 1983). Mutual observation among disciplines leads to an imitation of successful others, disregarding the question of whether success recipes are transferable from one discipline to another; and success means nothing but a higher standing in the inner-scientific rank-order of prestige. That a natural science such as chemistry fares well with collective profile-building does not mean that the same is true for history – still, this conclusion is often drawn. If mutual observation is underlined by expert opinions – which actually go back to mutual observation – the homogenisation of rationality fictions is stabilised.

(2) Secondly, even if members of disciplinary elites personally have different opinions they may believe that it is wise to 'go with the pack'. Thus, expressed ideas of good research may be just 'talk' perhaps meant to buffer real 'action' against interferences (Brunsson 1989). But this makes no difference in its long-term consequences. A required and monitored compliance of the evaluated units with these recommendations results in a dynamic of homogenisation of research; and in maybe ten years research in history will be more similar to research in chemistry than it is today.[19]

(3) Thirdly, academics with highly rated inner-scientific reputation are usually intensely requested policy-advisors with a good instinct for recommendations that may have the chance to be politically supported and realised in the future. Profile-building is an idea of good research that is easy to 'sell' in a science policy context of decreasing fiscal budgets and 'unpopular' political decisions to be made.

[19] However, there will most probably remain differences with respect to the understanding of and the specific measures for critical masses and profile building in both disciplines.

(4) Fourthly, there may be a normative bias in favour of collective profiles built into the evaluation procedure. Individual professors are only the secondary level of evaluation; the primary level are collectives: institutes or faculties. In addition, only the findings about these primary evaluated units are published. Does not this procedural regulation suggest to the evaluators that a unit which is artificially constructed has to have some features of a 'real' whole?[20]

(5) Fifthly, the evaluation procedure implies a certain pressure on evaluators to formulate recommendations – except, perhaps, in cases of overall excellent faculties or institutes; and the repertoire of possible recommendations is limited. Thus, to some extent there is a sheer statistical chance for the recommendation of critical masses and profile-building. Moreover, chances are higher because this appears to be a comparatively harmless recommendation in the sense that it seems to be easier to realise than some of the others.

In addition, evaluators may expect from profile-building an intensification of mutual monitoring and sanctioning among colleagues within an institute or a faculty – which amounts to a partial delegation of quality assurance to a self-organisation of those whose quality is at stake. This is surely an important component of the organisation-building effort which the AAC wants to initiate. Again, the university comes into view as an organisation with collective action capacity:

> The assessors have recognised (…) the potential to create a specific research profile in economics, which would clearly improve the chances to reach a top position. The following considerations list measures that seem (…) suitable for profile-building and the improvement of performance. (WKN 2001: 34 – Assessors' report – Economics)

For all these reasons the disciplinary elites might play the role of a trans-disciplinary research policy avant-garde which has begun to mobilise its disciplinary followers to move towards the same universal idea of good research. This effect is even reinforced by the fact that ministries also prefer to deal with a limited number of 'critical masses' than with a 'flea circus' of countless individual professors.

The important thing about this third pillar of the guiding idea of good research of the AAC's evaluations is that it fits the procedural logic of these evaluations analysed in the first section. Both idea and procedure point in the same direction: to the university as an organisation with a capacity of collective action. At the same time, both idea and procedure downgrade the individual professor as an actor in her own right. The procedural logic frames the individual professor as a member of an organisation which – represented by its leadership – is made responsible for her performance; and the idea of collective profile-building extends this frame from the monitoring and sanctioning of her research work right into this work itself. The research topics, the time-schedule, perhaps even the theories or methods used, or the

[20] In the German university system the fictitious approach of treating an institute or faculty as a research cooperative is especially pronounced because, different from the American system, there are usually no chairs with identical or similar denominations at the same place (Ben-David 1971). This way of concentration on certain sub-fields which gives a collective profile to an institute or faculty can only rarely be found in German universities.

loci of publication shall no longer be a purely individual choice but shall become embedded in a collective enterprise.

EFFECTS

The first cycle of evaluation exercises in Lower Saxony is nearly finished. All evaluated disciplines will be requested to hand in an intermediate report after three to four years that describes lessons which were drawn from the evaluations. Quite a number of the concrete recommendations of the evaluations of many disciplines have been implemented and start to show the desired effects. The university leaderships have very often picked up recommendations and tried to realise them even against the will of the affected faculties or institutes:

> The recommendations are precise and informative for the university leadership, but are completely at odds with the self-perception of the evaluated scientists. (letter from a university president; unpublished)

Thus, from the leadership's point of view most recommendations looked reasonable. This can also be seen from the fact that university managers included such recommendations in the performance contracts that they negotiated with the Ministry since 2002, and into the subsequent internal performance contracts with faculties: "[The results were] very useful as an argumentative backing for strategic goals in internal performance contracts." (letter from a university president; unpublished).

In this way, recommendations become binding commitments on the side of the university as well as on the political side. The ministry demands from the universities that they implement these recommendations, which implies that the university leadership demands this from the faculties and institutes; and if a serious effort is made to realise what is recommended the university can ask for the agreed basic funds from the ministry:

> [The university leadership] regarded [the assessors' reports] as particularly important because decisions deviating from these recommendations now require a special justification. (letter from a university president; unpublished).

In a number of cases evaluators and the AAC become even more involved in 'change management':

> The described high acceptance [of procedures] is a central argument for (…) the group of assessors' key involvement in the follow-up of complex recommendations. In this way the processes, which are often accompanied by considerable structural changes, are much more readily accepted by staff than they would be in the usual negotiations between staff and university leadership. (letter from a university president; unpublished)

Recently the AAC established two ad-hoc committees (humanities; engineering and natural sciences) in collaboration with former evaluators and universities. These committees are bound to support the step from 'diagnosis' to 'therapy' by conceptual efforts.

However, at least two factors which significantly hinder a quick and sustainable development along these lines must be mentioned. That quite a number of intended

effects get stuck is first of all due to financial limitations. Budget cuts[21] not only made it impossible for universities to follow a number of recommendations but, even more dramatically, have led to a considerable lack of trust on the university side in the ministry's intentions. Not only are these cuts of basic funding too big to be compensated by an increased acquisition of external funds; the latter becomes more difficult and in some cases impossible because it requires adequate basic funding. The establishment of a *Sonderforschungsbereich* (collaborative research area) promoted by the DFG, for example, presupposes a certain infrastructure and scientific staff to be already existing. In a couple of cases evaluators should perhaps have recommended the closing-down of a discipline at one university to transfer these resources to another instead of having two places where the discipline can neither live nor die. But the courage to speak up openly against a location of one's own discipline can realistically not be expected from evaluators who are, after all, bound by their disciplinary loyalty, even to those colleagues who are bad performers.

Secondly, these budget cuts go along with a still increasing teaching load which also impedes or even makes impossible the realisation of many recommendations. In addition to growing student numbers the implementation of the Bologna process in the EU (design of new study programs, accreditation, and quality assurance in teaching) consumes a lot of time and energy at German universities in a zero-sum relationship with research.

Bearing these difficulties in mind, we want to draw attention here not to the many specific effects of the recommendations within universities but to one often reported general effect. Undergoing the evaluation exercise from preparation via visitation to the implementation of recommendations intensifies the communication within the respective faculty or institute as well as between the evaluated unit and university leadership especially with respect to joint future objectives and strategy (Orr 2003: 33, 64). This starts with the fact that being confronted collectively by an evaluation is in some places the very first time that professors speak to each other about how their individual research activities are or could be related to each other (Schiene 2004: 86). The concrete recommendations, and the insistence on collective profile-building, provoke an even more explicit articulation, justification, planning and co-ordination of until then highly disconnected research activities:

> So, [after the evaluation] the practice of cooperation among staff has been considerably enriched, for example through the shared organization of research colloquia (…). Newly established meetings, even if they primarily concern issues of academic self-government or teaching, lead to more communication about research questions. (History - unpublished intermediate report of a university).

This externally enforced communication manifests itself in at least three aspects. Firstly, the recommendations are not directed in carefully separated pieces to each professorship but refer in substantial parts to the discipline as a whole at the respective university and are usually discussed as such in the faculty or institute by

[21] Lower Saxony cut about 40 million in 2004, another €10 million in 2005 of the block grants given to universities. This had a noticeable disadvantageous impact on the function of the AAC as mediator between government and universities.

all those who were assessed. Each 'research unit' as well as each university can compare itself with others according to the applied indicators. Secondly, a considerable number of recommendations can only be realised collectively; to this end, fitting organisational structures have to be conceived and implemented. Thirdly, some of the recommendations point directly to the university as an organisation. In sum, individuals move via interaction to organisation-building.

A decisive aspect of this dynamic is the shift of perspective many professors have to learn. Traditionally, their university or even their faculty or institute were nothing but resources to be exploited for one's own professorship's needs. Everything outside the professorship but inside the university was potential loot, some of it having the status of the commons where everybody is free to take whatever and how much he can, some being property of others which can perhaps be taken from them under favourable circumstances. Now, the professor has to conceive of himself and his professorship as part of a larger whole which is not just a common hunting-ground but shall have the quality of a corporate actor on its own who, moreover, has certain claims on its individual members. In future, research activities have to fit into an institute's or faculty's profile and contribute to a collective good, the research performance of the respective organisational unit.

In this way, the implementation of concrete recommendations and the general dynamic of organisation-building go hand in hand, the latter being the often latent accompanying effect of the former. Nevertheless there is still a widespread resistance of professors against the implementation of university reforms in general, and the recommendations of the AAC in particular, especially with respect to collective research profiles and organisation-building.[22] The habitual extraordinarily high individual autonomy of each professor is defended energetically – by a considerable number of professors not the least because it implies the freedom to remain a bad researcher or to do no research at all. Within the still existing traditional governance regime of German universities individual professors are rather powerful in the defence of their own domain.

Thus, to make a real difference the evaluation exercises of the AAC cannot remain isolated activities but must be embedded in an overall reshaping of university governance which gives proper hierarchical authority and competencies to deans and the university president. Although Lower Saxony by now has gone more in this direction than most other German Länder, the traditional extreme respect for an individual professor's autonomy which results in a culture of standstill or a 'do nothing university' is still strong.[23]

In this difficult interim phase, it is strategically important for the ministry to resist the temptation, stemming from its distrust in the universities' willingness to change, to intervene too strongly into university decision-making and, in this way,

[22] This resistance sometimes uses financial arguments as a disguise. For example, at one university the faculty of chemistry excuses its inactivity with regard to the recommendations of the evaluation by financial scarcity and insecurity. But "... evaluators cannot share this view ..." and trace the immobility back to the fact "... that in X one is primarily concerned with a defense of everyone's own 'hunting ground'". (Chemistry - unpublished intermediate report of a university)

[23] To adopt an expression from former British minister of education David Blunkett, quoted in Orr (2003: 68).

weaken the university leadership which it wants to strengthen.[24] With respect to this necessary self-restraint of the ministry as well as to the above-mentioned obstacles it is clear: Only if evaluations and implemented recommendations do indeed lead to improved research performance some years later this new instrument of university governance is legitimated; but this critical dependence upon success requires, in turn, a favourable context for the implementation of recommendations which in most places does not exist but has to be established by political action with respect to finances, teaching load, and governance.

If, by and by, everything really goes in the direction of the AAC's idea of good university research in Lower Saxony, this means – to reiterate our main argument – that disconnected individual research is gradually transformed into collective research profiles of organisational units such as institutes or faculties. This does not amount, as is often superficially declared, to a state of affairs where the academic profession is substituted by the university as a formally organised framework.[25] It is certainly true, as we just mentioned, that the new governance regime with which this idea of good university research is associated emphasises strong organisational leadership and de-emphasises the academic profession's collegial self-governance, which is dominated by implicit non-aggression pacts among professors so that an individual professor's autonomy is maximized (Schimank 1995: 222-258).

However, professors continue to play an important part in the governance of the university system. The individual professor's influence and power to defend her own status and autonomy will weaken, as well as the formal collective power of professors in intra-university collegial bodies. But especially through inter-university mechanisms of peer review, disciplinary elites of professors will have a clear collective impact on policies and decisions of resource allocation; and this impact will even grow if policy-makers come to the conclusion that competitive pressure on the quasi-market for block grants is best exerted on the basis of performance information from peer review. The existence and work of the AAC is a clear example of this development. Thus, it is more accurate to say that the academic profession becomes organised and more stratified so that it attains a collective action capacity – if necessary, against its individual members. And one of the things which become possible as a result of this is the establishment of collective research profiles based on critical masses of researchers within faculties and institutes of universities.

In closing, we would like to draw attention to the other side of the coin. Granted that this general organisational development of German universities will happen in Lower Saxony, significantly initiated and accompanied by the AAC, and, for the sake of the argument, granted further that many of the hoped-for positive effects on research performance will indeed be realised: One nevertheless should not overlook possible negative side-effects. We see especially one such unwanted effect with a presumably serious long-term impact on the research system: a weakening of

[24] With respect to the autonomy of universities the German federal university system displays considerable variance. Obviously the degree of autonomy the *Länder* are willing to concede to their universities becomes a crucial determinant of their capability of organisational development.

[25] For a more extensive discussion of the following see Schimank (2005a).

university research as the most important evolutionary variety pool of research in general (Nowotny 1990).

Collective research profiles presuppose a reduction of an individual professor's autonomy of his selection of research topics, and this extends not always, but at least sometimes also to decisions about research approaches (methods, theories). This may be a good thing for 'normal science' (Kuhn 1962) and its systematic incrementalistic variation of the existing knowledge base.[26] Such a 'small steps approach' (Sunstein and Ullman-Margalit, 1999) can be executed in a much more coordinated and efficient way if researchers involved in the respective disciplinary sub-field commit themselves to a joint venture under a paradigm agreed upon by all; and local clusters of such a joint venture within a faculty or an institute are adequate organisational forms for this purpose. However, if such joint ventures include each and every professor, the potential for radically new research topics and approaches suffers dramatically, and a totalised mainstream drives out unorthodox approaches. But paradigmatic revolutions more often result from unorthodox approaches than from a mainstream approach which finally exhausts itself if put to its limits. The critical issue about this is that the universities are almost the only places within the research system where there is – or has been until now – an institutionalised secure space for unorthodox approaches. The individual professor's autonomy expresses this institutionalisation.[27]

To return to the beginning: Although it should be clear by now that in our view the statements of the university president about the work of the AAC we quoted at the outset do make much sense, our second protagonist, the professor of engineering, has one strong point:[28] His insistence on an individual professor's ability to pursue her own research ideas, however idiosyncratic and bizarre they may look to others, including evaluators, should not be done away with lightly. A delicate balance has to be found between collective research profiles on the one hand, and individual autonomy on the other; and this balance can be found only by the academic profession. Among others, institutions such as the AAC should be used as an opportunity structure for an identification and implementation of this balance.

REFERENCES

Becher, Tony and Paul R. Trowler (2001), *Academic Tribes and Territories. Intellectual Enquiry and The Culture of Disciplines*, Buckingham: Open University Press, Second edition.

Ben-David, Joseph (1971), *The Scientist's Role in Society. A Comparative Study*, Englewood Cliffs, N.J.: Prentice-Hall.

Brinckmann, Hans (1998), *Die neue Freiheit der Universität. Operative Autonomie für Lehre und Forschung an Hochschulen*, Berlin: Sigma.

[26] Although we are aware that Kuhn's distinction of 'normal' and 'revolutionary' science is too simple, we can make use of it for our purposes here.

[27] However, the argument should not be overstressed. That only individuals, not groups can be radically creative is partly a myth. Still, the number of creative units is strongly reduced if all individuals are strictly integrated into collective research efforts.

[28] There may be also some truth in what he expresses as his main message: There is undoubtedly considerable 'red tape, in the evaluation business which has to be corrected.

Brunsson, Nils (1989), *The Organization of Hypocrisy – Talk, Decisions, and Actions in Organizations*, Chichester et al: Wiley.

Bundesverfassungsgericht (2004), *1 BvR 911/00 vom 26.10.2004*, Absatz-Nr. 1-191, http://www.bverfg.de/entscheidungen/rs20041026_1bvr091100.html (accessed 19 October 2005).

Campbell, David (2005), *Evaluation of University Research in Europe – Conceptual Framework and Comparative Typology*, University of Klagenfurt, IFF: unpublished manuscript.

Chandler, Alfred (1977) *The Visible Hand. The Managerial Revolution in American Business*, Cambridge, MA: Belknap Press.

De Boer, Harry, Jürgen Enders and Uwe Schimank (2006), 'On the Way Towards New Public Management? The Governance of University Systems in England, the Netherlands, Austria, and Germany', in Dorothea Jansen (ed.), *New Forms of Governance in Research Organizations. From Disciplinary Theories Towards Interfaces and Integration*, Dordrecht: Springer, pp. 137-152.

DiMaggio, Paul J. and Walter W. Powell (1983), 'The Iron Cage Revisited', *American Sociological Review*, 48, 147-160.

Gläser, Jochen (2006), *Wissenschaftliche Produktionsgemeinschaften. Die soziale Ordnung der Forschung*, Frankfurt am Main: Campus.

Kehm, Barbara and Ute Lanzendorf (2006), 'Changes in the Governance of Public Research in Germany', in Barbara Kehm and Ute Kanzendorf (eds.), *Reforming University Governance. Changing Conditions for Research in Four European Countries*, Bonn: Lemmens, pp. 135-186.

Kieser, Alfred (1997), 'Myth and Rhetoric in Management Fashion', *Organization*, 4, 49-74.

Krücken, Georg and Frank Meier (2006), 'Turning the University into an Organizational Actor', in Gili Drori, John Meyer and Hokyu Hwang (eds.), *Globalization and Organization: World Society and Organizational Change*, Oxford: Oxford University Press, pp. 241-257.

Kuhlmann, Stefan and Thomas Heinze (2004a), 'Evaluation von Forschungsleistungen in Deutschland: Erzeuger und Bedarf. Teil I: Konzeptionelle Grundlagen', *Wissenschaftsrecht*, 37, 53-69.

Kuhlmann, Stefan and Thomas Heinze (2004b), 'Evaluation von Forschungsleistungen in Deutschland: Erzeuger und Bedarf. Teil II: Produktion und Verwendung evaluativer Information sowie Möglichkeiten ihrer künftigen Organisation', *Wissenschaftsrecht*, 37, 125-149.

Kuhn, Thomas S. (1962), *The Structure of Scientific Revolutions*, Chicago: The University of Chicago Press.

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

Meyer, John W. and Brian Rowan (1977), 'Institutionalized Organizations: Formal Structures as Myth and Ceremony', *American Journal of Sociology*, 83, 340-363.

Nowotny, Helga (1990), 'Individual Autonomy and Autonomy of Science: The Place of the Individual in the Research System', in Susan E. Cozzens, Peter Healey, Arie Rip and John Ziman (eds.), *The Research System in Transition*, Dordrecht: Kluwer, pp. 331-343.

Orr, Dominic (2003), 'Verfahren der Forschungsbewertung im Kontext neuer Steuerungsverfahren im Hochschulwesen: Analyse von vier Verfahren aus Niedersachsen, Großbritannien, den Niederlanden und Irland', *HIS Kurzinformation*, A1, 16-74.

Schiene, Christof (2004), 'Forschungsevaluation als Element der Qualitätssicherung an Hochschulen', *Zeitschrift für Evaluation*, 1, 81-94.

Schimank, Uwe (1994), 'How German Professors Handled Increasing Scarcity of Resources for Their Research: A Three-Level Actor Constellation', in Uwe Schimank and Andreas Stucke (eds.), *Coping with Trouble: How Science Reacts to Political Disturbances of Research Conditions*, New York: St. Martin's Press. pp. 35-59.

Schimank, Uwe (1995), *Hochschulforschung im Schatten der Lehre*, Frankfurt am Main: Campus.

Schimank, Uwe (2005a), ''New Public Management' and the Academic Profession – Reflections on the German Situation', *Minerva*, 43, 361-376.

Schimank, Uwe (2005b), *Die Entscheidungsgesellschaft. Komplexität und Rationalität der Moderne*, Wiesbaden: VS.

Simon, Herbert (1982 [1976]), 'From Substantive to Procedural Rationality', in Herbert Simon (ed.), *Models of Bounded Rationality*, vol. 2, Cambridge MA: MIT Press, pp. 424-443.

Sunstein, Cass R. and Edna Ullmann-Margalit (1999), 'Second-Order Decisions', *Ethics*, 110, 5-31.

Weaver, R. Kent (1986), 'The Politics of Blame Avoidance', *Journal of Public Policy*, 6, 371-398.

WKN [Wissenschaftliche Kommission Niedersachsen] (1999), *Research Evaluation at Universities and Research Institutions in Lower Saxony. Procedure Outline*, Hannover: WKN.

WKN [Wissenschaftliche Kommission Niedersachsen] (2000a), *Forschungsevaluation an niedersächsischen Hochschulen. 'Chemie', Bericht der Gutachter*, Hannover: WKN.
WKN [Wissenschaftliche Kommission Niedersachsen] (2000b), *Forschungsevaluation an niedersächsischen Hochschulen. 'Geschichte', Bericht der Gutachter*, Hannover: WKN.
WKN [Wissenschaftliche Kommission Niedersachsen] (2001), *Forschungsevaluation an niedersächsischen Hochschulen. 'Wirtschaftswissenschaften', Bericht und Empfehlungen*, Hannover: WKN.
WKN [Wissenschaftliche Kommission Niedersachsen] (2002), *Forschungsevaluation an niedersächsischen Hochschulen. 'Rechtswissenschaften', Bericht und Empfehlungen*, Hannover: WKN.
WKN [Wissenschaftliche Kommission Niedersachsen] (2004a), *Forschungsevaluation an niedersächsischen Hochschulen. 'Medizin', Ergebnisse und Empfehlungen*, Hannover: WKN.
WKN [Wissenschaftliche Kommission Niedersachsen] (2004b), *Forschungsevaluation an niedersächsischen Hochschulen. 'Politikwissenschaft und Soziologie', Ergebnisse und Empfehlungen*, Hannover: WKN.
WKN [Wissenschaftliche Kommission Niedersachsen] (2004c), *Forschungsevaluation an niedersächsischen Hochschulen. 'Germanistik mit Niederdeutsch, Niederlandistik und Skandinavistik; Kulturanthropologie / Europäische Ethnologie; Medienwissenschaften', Ergebnisse und Empfehlungen*, Hannover: WKN.
WKN [Wissenschaftliche Kommission Niedersachsen] (2005), *Forschungsevaluation an niedersächsischen Hochschulen. 'Philosophie', Ergebnisse und Empfehlungen*, Hannover: WKN.
WKN [Wissenschaftliche Kommission Niedersachsen] (2006) *Forschungsevaluation an niedersächsischen Hochschulen und Forschungseinrichtungen. Bewertung des Evaluationsverfahrens*, Hannover: WKN.

CHAPTER 9

BAREND VAN DER MEULEN

INTERFERING GOVERNANCE AND EMERGING CENTRES OF CONTROL

University Research Evaluation in the Netherlands

INTRODUCTION

In analysing the emergence of new evaluation practices for academic research in the 1980s, one cannot but link these to changes in the post-war relationships between universities and governments. After WW II academics became accustomed to an autonomy regime, in which governments were willing to provide funds for academic research exercising control neither on the academic performance nor on the returns of the investments. Quality control was left to the academic sector or, more precisely, quality control was seen as implicit to the dynamics of science and not something that had to be organised separately.

In the Netherlands, this policy approach was rejected in the 1980s and new evaluation practices were implemented. In the context of science policy, committees were established to advise the government on discipline-oriented science policies. As part of their work these committees started to evaluate university research. In 1982, funding of university research was made conditional on the positive assessment of research programmes. The aim was to organise 80% of all the institutionally funded research in such programmes. Since then, evaluation evolved through several stages from loosely co-ordinated exercises organised by the government, through a well-established standardised practice to the current situation in which research units can organise their own evaluation.

Usually, the development of state evaluations of university research is seen in terms of the establishment of new instruments by which the government exercises control over the universities and university research (OECD 1997; Geuna and Martin 2003). We even find this approach in more anthropological studies in which the evaluation of university research is taken as yet another example of the rise of an audit culture within Western societies (Shore and Wright 2000). Such analyses suggest that the government can define and control the evaluation frameworks, and

R. Whitley and J. Gläser (eds.), The Changing Governance of the Sciences, 191–203.

neglect the role of university management, researchers and evaluators. Moreover, a focus on the policy dimensions of evaluations ignores the evaluative practices of science itself. Though evaluations may be implemented from 'outside', it is likely that when they stay and develop into a practice, they become part of the world of science, and add to the reputational organisation of the disciplines (Whitley 1984).

The development of university research evaluation in the Netherlands can be used to study how evaluations become both government instruments and part of the world of science at the same time. One can distinguish two different but related governance arrangements. First, there is the institutional relations between governments, universities, departments/research groups, and researchers. These relations are nationally defined and institutionally filled out by funding arrangements, work contracts and distribution of research policy responsibilities over the different levels. Second, evaluations of research are embedded in disciplinary contexts that frame the dynamics of research programmes and researchers' strategies both cognitively as well as professionally. The disciplinary context is an implicit governance arrangement in which notions of research quality are defined and research strategies are socio-cognitively coordinated.

Each of the governance arrangements has its own evaluation practices in which different notions of research quality are defined. The emergence of a new evaluation practice for university research in the Netherlands ensured that the two arrangements interacted, or, in a sense, 'interfered'. Evaluations produce performance indicators, scoring lists, ranking orders and, as such, insight into something intangible like research quality. Disciplinary bodies are included in the governance interactions to set up legitimate, peer-based evaluation procedures and create a link between disciplinary notions of research quality and the new governance instruments. Disciplinary interactions are modified by these evaluations that aim to shape the accountability relations between government and universities but also affect the reputational order.

The notion of interference should be taken quite literally. Interference of waves creates a new wave pattern because crests from both waves reinforce each other at some points, and at other points crests are destroyed by troughs. Similarly, the interference of governance arrangements may reinforce some actor positions, enforce new interaction rules, create strong notions of research quality and thus induces the emergence of 'centres of control' or 'nodal points of power' (Clegg 1989). The interference between two governance arrangements may also weaken positions and destroy existing coordination and control mechanisms.

There is a difference though. The interference of waves usually creates stable patterns. In case of interference of governance arrangements, however, actors are likely to respond to the new pattern. The weakening of specific positions and interactions may, for instance, create attempts by actors to restore their positions, e.g. by introducing evaluation processes. On the other hand, when new centres of control emerge from the new evaluation practice, their role and function may be used for other functions as well and actors may link up with these centres to improve their own positions. As a result, even when an evaluation practice has become stable in terms of its procedures, definition of research quality and methods, its meaning in governance arrangements can change.

I analyse university research evaluation in the Netherlands here to understand the evolution of an evaluation practice. I present three episodes and use each of them to highlight aspects of how this evaluation practice became embedded and how it changed governance arrangements. The first attempt to evaluate university research started without clear ideas about who should do the evaluation, how such an evaluation should be done and what the function of the results could be. Only during the process of preparing the evaluation, doing it and discussing the report, did the actors develop these ideas and, as it were, construct an evaluation. In the second episode the uncertainty about methods continued but the evaluation became linked to university research funding and to the organisation of university research. A new centre of control emerged in university research management with the development of research programmes. The third episode discusses the evaluation practice that was implemented in the nineties by the Association of Dutch Universities.

Standardisation of the procedures and reporting enabled university managers to use the results in other contexts. While this practice seems stable, the end of the third episode shows, almost dramatically, how the success of the practice induced dynamics that may result in its end.

EPISODE 1: THE CONSTRUCTION OF AN EVALUATION

In 1969 the then Advisory Council for Science Policy recommended the establishment of "ad hoc committees of experts… that should assess the nature and suitability of the research financed by the government in restricted fields of research" (RAWB 1969: 32, translation by the author). In the first half of the 1970s the Dutch government established three such committees to assess research linked to government policy sectors: education, social policy and environmental planning. In 1976 the focus shifted towards the sciences and in the years afterwards disciplinary committees were set up for chemistry, biochemistry, biology and physics. In the mid-eighties, these were followed by committees for economic science, law, philosophy, educational research and pedagogics, sociology and policy science, theology and for building research, which all operated in the context of the budget cuts between 1985 and 1990. Without a well-developed evaluation practice, each of these committees had to define what has to be evaluated, who can evaluate it, and how it can be evaluated, for what actual purpose and for whom it needs to be evaluated (Van der Meulen et al. 1991).

The chemistry evaluation showed how the Minister, disciplinary bodies and the committee together built up something as an evaluation. In 1976, in the Science Budget the Minister announced that he would set up a committee for chemistry in the Netherlands. The goal of that committee was to review all research activities chemistry including those in the private sector. In 1977, in the next Science Budget the Minister decided to distinguish biochemistry from chemistry and to set up a separate committee for the former area. The ambition to include industrial research in the exercise was dropped. The Dutch chemical companies refused to take part in the exercise. Still, no committee was formed. The Minister delegated the task of selecting members of the committee to the Academy Committee for Chemistry,

which took the opportunity to discuss the tasks of the committee as well. In April 1978 the committee was established. As a result of the negotiations between the Ministry and disciplinary bodies, it was set the ambitious task of describing and assessing the development, quality, effectiveness, coherence and relevance of the sub disciplines of Dutch chemistry in international perspective, and to advise about a policy for chemistry in the Netherlands.

Two years later the report was published.[1] It was a result of a committee that developed its methodology incrementally. For more than a year, committee members discussed aims and objectives of the exercise and general principles of science policy at their meetings. Meanwhile, the committee also started to interview key persons in the discipline and to collect data about publicly financed research activities. The report gave a detailed description of publicly financed research. It included a table with the research efforts per research group, research area and research subject. Also the number of publications and patents were counted per subfield. Policy recommendations, though, reflected the discussions at the meetings, rather than the results of the data collection.

About the quality of the research, the committee spoke only in general terms ('the quality of chemistry in the Netherlands') referring to more than fifty interviews with representatives from faculties and research institutes, disciplinary bodies, industry, and key scientists from abroad. The committee did not assign quality rankings to research groups, individuals or local research programmes. In an appendix the committee reported a 'quantitative literature study' in which data were gathered on the number of publications in chemistry from the Netherlands. The data were organised and compared by subfields. Interestingly, some of the journals in English were considered as 'national journals' for the UK, US and Canada. The committee doubted whether it should take publications in these journals into consideration. Likewise the committee tried to balance the results for those fields that publish predominantly in Dutch.

The report was sent to a range of science policy bodies, national and disciplinary ones – and they all sent their responses to the Ministry. Most bodies supported the recommendations to improve researchers' mobility and to strengthen interaction with industry, but mentioned that there is little relation between the assessment of Dutch chemistry and these recommendations. In December 1981, almost two years after the report was published, the Ministry concluded that the report does not provide much of a base for a chemistry-oriented research policy. It especially regretted that the report did not set priorities and *posteriorities*[2] for chemistry.

While the government did not adopt the results and recommendations of the chemistry report, disciplinary bodies elaborated on the work of the committee. In the years afterwards, an increasing number of reports, produced by the Royal Netherlands Chemical Society and by the chemistry bodies of the Netherlands Academy of Arts and Science, of the then Research Council and of the *Academische*

[1] At that time, the committee for biochemistry still had to begin. It would publish its report only in October 1982.

[2] *Posteriorities* are negative priorities: a typical Dutch science policy term from the eighties, related to the budget constraints at that time and to the logic: 'if the government wants us (universities, researchers) to put our (sic!) money in priorities, it also has to say where to take it from.'

Raad[3] discussed and coordinated chemistry research in the Netherlands. Through these reports the chemistry community did what the ministry could not do: develop strategies to improve research conditions and distribute research priorities over the faculties. The chemistry evaluation moved from the science policy governance arrangement to the national subfield and stayed there; separate from its original context.

Evaluation then was a process which extends before and beyond the visit of a peer committee and the publication of the results. It started with the first mention of a university research evaluation (by a government report, a disciplinary body or in a planning scheme of an audit body) and continued when the evaluation results are taken up in the two governance arrangements. The evaluation became part of interactions about funding, mergers of departments, research programmes etc. 'University research' as an object of governance then developed from something that is funded, through something that needs to be evaluated, towards something that has certain properties and problems represented in a report. In between all sort of translations need to be made to construct the objects of evaluation, the evaluation framework, the principles for the evaluation, the expertise needed to organise the evaluation and the experts that can do the assessments (Van der Meulen 1995).[4] These translations are important, as they determine to what extent the evaluation links up with existing governance arrangements or remains a *Fremdkörper*.

EPISODE 2: CONSTRUCTION OF AN EVALUATION OBJECT AND CENTRES OF CONTROL

In 1979 the Ministry for Education and Sciences announced a new funding scheme for university research. Universities were forced to organise their research in research programmes to be assessed ex ante by peer committees. Universities that could not provide a considerable part of their research budget with approved programmes could lose funding. Positive assessment implied a budget protection of the programmes for five years (see also Ball and Verkleij 1999).

For all actors it was unclear how assessments had to be made. Peer committees were asked to evaluate scientific quality and societal relevance, but without any indications how these concepts had to be interpreted. The committees used different indications for judging quality. Some of the committees assessed only the programme descriptions. Others focused on the scientists within the programmes or, being peers, based their judgements on their general knowledge of the group's research performance. None of the committees was able to evaluate societal relevance systematically. Moreover, with an eye on the policy context with stringent budget conditions, most evaluation committees were not very harsh in their

[3] Before the Association of Dutch Universities was established, the *Academische Raad* (Academic Council) was the national council in which the universities met. It had a fine-grained structure of (sub)disciplinary and thematic sub-councils.

[4] Translations in the sense of Actor Network Theory, by which an object is constructed through associations with others and gets a meaning in relation to its context (see e.g. Callon 1986).

judgements. A few, though, were very strict and assessed a considerable number of programmes as being of low quality (Blume and Spaapen 1988).

In 1985, the universities negotiated successfully a change in the evaluation procedures and in 1987 a second round of evaluations of conditionally financed programmes was initiated. The emphasis shifted to accountability and instead of ex-ante assessments research programmes were evaluated ex-post. Within the Ministry, the idea remained that evaluation outcomes should have consequences for budget allocations to universities, but no rules or procedures were developed to implement such reallocations. Although a strict evaluation procedure was still lacking, making ex-post evaluations turned out to be less problematic and, contrary to the first round, most evaluation outcomes were seen as valid. Within the university system the outcomes became important, and university managers recognised the opportunities to use the evaluation system for university research management. At the end of the second round, in 1992, universities and researchers accepted the idea of accountability. But they also felt that the feedback they receive from peers was rather limited considering the evaluation effort. Universities wanted to have full responsibility for the evaluation, which they indeed did get. In 1993, the responsibility for the evaluation of university research was transferred to the Association of Dutch Universities (VSNU).

Evaluation practices changed relations and redefined balances of trust, accountability and control in relationships. The ex ante evaluations had little direct impact upon the funding and the Minister remained empty handed. The results of the evaluation did not really move outside the walls of the universities and did not shape disciplinary interactions. But a new entity emerged within the institutional relations of government-universities-researchers: the research programme as an object for evaluation.

Research programmes have aims and objectives, a programme leader and research staff, a work plan and a budget – all that is needed to make up well-organised research. In the natural and biological sciences, research programmes were not uncommon; though even in these disciplines the evaluative grid forced the faculties to demarcate clearly the programmes and thus they started to programme at faculty level. In other disciplines, researchers hade to join efforts in a common five-year programme for the first time. Most of these research programmes started simply as an administrative entity, but soon after became not just the object of an evaluation, but also the place where research management by the university, by the faculty and by the research group came together. The paper construct created a new reality and 'programme leader' became a responsibility, a function within the organisation accountable for the performance of the programme.

Through the new funding system and its related evaluative practice, the research programme emerged as a new centre of control, which was both part of the institutional governance and the socio-cognitive governance of university research.[5] These research programmes became the places where the research activities in the laboratory and behind the researchers' desks, the field work and literature reviews

[5] Clegg (1989) uses 'nodal points of power' after Laclau and Mouffe's (1985) 'nodal point of discourse' and 'obligatory passage point' from Callon (1986).

can be described in management terms. The programmes have research lines, research aims and research questions and produce empirical and theoretical findings and scientific publications. Peer committees can read the programmes and judge them. But they also provide a frame for counting, calculating and distributing full time equivalents, institutional block grants, lab space, PhDs and relate these input factors to the number of scientific publications and approval by peers.

The management of the programmes can be easily described as a form of self-organisation by the scientists. They have developed the programme themselves and through the programme the researchers can continue research activities as usual (see also Morris 2002). It is a limited description. The strength of research programmes as management device derives from the simultaneous imposition of external control and internalisation of the new norm that research needs to be organised in a research programme that produces quantifiable outcomes (Shore and Wright 2000). The interference between the disciplinary and institutional governance processes leads to a 'peak' in the governance landscape from which activities are controlled. The irony is that the more the research programmes organised research activities, the more useful they became for evaluations and the more researchers became susceptible to accountability relations with research managers at other levels of the university.

EPISODE 3: STANDARDISATION, COMPARABILITY AND SECOND ORDER DYNAMICS

In 1993 the VSNU acquired responsibility for university research evaluation. The association developed an evaluation system that informed in detail the university board and departments about the performance and progress of research programmes. The Ministry of Education and Science became informed about their quality in general terms only. This evaluation approach developed quickly into a well-established practice in which all universities took part. All university research became subjected to evaluation in four-year cycles. At the beginning of a cycle, the universities enact a protocol for the whole evaluation cycle, which defines the disciplines and their year of evaluation, the responsibilities of every actor involved in the evaluation, the criteria for evaluation, the minimal information on which the evaluation has to be based and the procedure of the evaluation (see Appendix).

The actual evaluations were preceded by self-assessments of the departments, reports written in formats according to the guidelines of the protocol. The self-assessment reports consisted of a description of the research programme, an overview of the performance of the last five years, future plans as well as a list of five key publications. Full publication lists were usually added as an appendix. If programmes had been evaluated before, the self-assessments had to make clear how previous recommendations had been taken up. The evaluations were done by peer committees, who evaluated on the base of the self-assessment reports complemented by interviews with programme leaders. In most disciplines, especially the laboratory-based ones, site visits were made. In a few disciplines, especially those with a large number of research programmes, programme leaders were asked to visit the committee, instead of the site visits.

Research programmes were evaluated on four aspects - quality, productivity, scientific relevance, viability – on a five point scale (excellent, good, satisfactory, unsatisfactory, poor). A comparison of the research evaluations of chemistry and of law from the first cycle makes clear that the protocol left space for disciplinary interpretations of these four aspects (Table 1). In chemistry, VSNU's standing disciplinary committee translated 'quality' and 'productivity' into output indicators, based upon peer reviewed international publications. The meaning of 'relevance' and 'viability' is outlined and it is left to the evaluation committee to decide what data are needed for assessing these aspects. The evaluation committee for law research started with less specific interpretations of the four aspects. For quality, the committee looked at the key publications and trusted its own peer competence to evaluate the quality, or asked external peers to assess the publications. For productivity, it developed a formula by which programmes can be ranked into different categories. The other two aspects were not evaluated systematically. The report mentioned these only if programmes did very well on them.

Despite the differences, the final scoring lists in the evaluation reports created an important opportunity for university managers: Chemical research can be compared to law research. A research programme at a law faculty with a '5' for quality and a '4' for productivity has higher status than a chemistry programme with, for instance, scores of '4,3,4,3'. Local reputations of researchers and research groups, developed over along time, suddenly have an 'independent' basis. Evaluations may reveal

Table 1. Criteria used in chemistry and law research evaluations

Evaluation aspect	*Chemistry*	*Law*
Scientific Quality	Quality of output International visibility	Quality of key publications
Scientific productivity	Number of PhD theses Number and kind of international publications Number of patents Number of invited lectures	Calculation of output according to a formula
Scientific relevance	Research topics and methods Expected impact on progress of chemistry and other sciences Expected impact on progress of technology	Relevance and viability are assessed for those programmes that deal with particular relevant subjects and that are clearly progressing.
Long term viability	Future research plans Human resources Research facilities	

under-performances of politically strong groups and excellent performance of weaker groups.

The standardisation of the evaluation results enabled the university managers to use them in decision processes at the university level. Reports were used incrementally in strategic decisions, e.g. on investments, new professorships and the organisation of the faculty (Westerheijden 1997). The outcomes of the research evaluations and required reputations were also used within other contexts and facilitated good results in other evaluations, provide access to research collaborations and key positions in the disciplinary field. Moreover they were used by researchers in the acquisition of competitive funding. In new funding schemes in the 1990s to stimulate the creation of centres and networks of excellence, high scores in the university research evaluations were necessary to join consortia. Something of a Matthew effect (Merton 1968) occurred, not at the individual level but at the level of research groups and programmes.

The evaluations organised by the Association of Dutch universities added new features to the evaluative relationships in the Dutch university research landscape. The research programme as a centre of control was complemented by the evaluation practice of the Association of Dutch Universities as a new centre of control, from which the evaluations are organised. Here, self-assessments, peers, judgements, bibliometric indicators come together to create tables of performance for all disciplines. The two centres are interdependent: The evaluation practice needs the research programme in order to have an evaluation object that can be presented as an entity in terms of the protocol. The force of the research programmes rests on the evaluation as long as it produces reliable results that can be used to manage research internally and externally.

The association between the two is crucial for understanding how the evaluation practice affected the institutional relationships within universities and between university and government. Between the university and the government, the old days before the 1980s seemed to have returned: the government funds research but has no opportunities to exercise control. The difference though is that while before the 1980s quality control was implicit in the dynamics of science, it is now explicitly organised. The evaluation practice buffers the influence of the government on university research activities, while at the same time strengthening the position of the university management at university and faculty level. This university management becomes the link between the research programmes and the Association of Universities.

The standardisation of evaluation practices created new opportunities and further dynamics in the evaluative practice. It implied that evaluation itself becomes transferable, and universities do not necessarily need their association to manage the process. In 1998, researchers successfully asserted to the then new Minister of Science and Technology that research evaluation had run out of control and that a reduction of evaluations was needed. In the 1990s, policy schemes for graduate schools and for different kinds of centres of excellence introduced evaluations in addition to the VSNU evaluations.

The Minister asked the Royal Netherlands Academy for Arts and Sciences (KNAW), the VSNU and the national research council to develop solutions to

reduce the evaluation load. They came to the conclusion that the university research evaluations can be delegated even to a lower level, the university. A joint committee of the three bodies developed a protocol for the evaluations, which resembles the one of the VSNU, but in which each university can define the disciplinary scope of an evaluation. The minister agreed to the plan, and from 2002 onwards each university itself has been responsible for organising the evaluations. Some evaluations are still organised jointly, especially when there is a strong disciplinary body, and some universities take the opportunity to break through the disciplinary grid, and organise evaluation for multi-disciplinary areas and institutes.

There is another irony here: as soon the new evaluation structure was implemented in which the university, not the association of universities, becomes responsible for the evaluations, the Minister started a discussion about his lack of insight and control in the allocation of university funding. Old policy aims to reallocate funding between the universities and reward performances are reformulated in the context of improving the innovation role of the university. In a recent advice to the Minister, a committee of high level experts concluded that the new evaluation structure has raised the number of evaluation panels, in that respect it is inefficient, and that internal resource allocation and performances of the universities are not transparent. The movement of the evaluation of university research from the government, to disciplinary bodies, through the VSNU to the universities seems to have moved into a dead-end.

CONCLUSIONS

In the Netherlands there is no direct relation between the evaluation outcomes and government allocation of funding. Basic funding for university is provided as a lump sum which is calculated for every university on the base of a formula that includes some teaching performance indicators and some historically determined elements. If not affecting the allocation of basic grants, do evaluations have any impact at all? In our analyses, we have looked differently at evaluation of university research. Not as a government *instrument* per se, but as an emerging phenomenon within institutional and disciplinary governance arrangements. We will not recapitulate the evolution of this phenomenon, but look at the impacts of the phenomena of university research evaluation and how it has shaped the Dutch research system. These impacts go far beyond the allocation of funding.

One impact is through the way by which in evaluations university research is presented. Representations are crucial for governance. Administrators and managers perceive the object of governance through maps, numbers and formulae and only through such representations they can control complex and heterogeneous activities like research (Cooper, 1992; see also Law, 1986). ‘University research’ that is evaluated is not the activities within the laboratory, the tinkering with instruments, making questionnaires, hermeneutic interpretation of medieval texts. The early evaluations tried to capture these activities through qualitative peer reviews and interviews about the quality of the research. In the VSNU evaluations, the site visits of the committees may reveal some of these aspects as well. But in the evaluation

reports, 'university research' becomes university research groups, with publications and a bibliometric profile, a certain number of staff and PhD's, and together these groups make up a set of which some are 'excellent', others 'good' and some are 'poor'. The managerial intangibility of research activities has been tackled and its properties can be translated easily by managers within decision procedures.

A second impact is how the evaluations create definitions of research quality and implicitly also of proper performance. The difference between how the first chemistry committee in the eighties looked at international research publications and how quality and productivity were defined in the VSNU protocol for chemistry is striking. The first committee doubted whether all international publications should be taken into account and publications in Dutch were seen as a legitimate reflection of the national role of chemistry research at universities. In the latter evaluation such international publications were the norm and national publications did not add to the quality of the research. Similar tendencies towards a restriction of research quality in terms of international journal publications can be found in other fields, despite recurrent pleas, including from the Netherlands Royal Academy of Arts and Sciences, for more differentiated interpretations and acknowledgement of other publication strategies.

The main impact, though, concerns how the evaluations have reshaped institutional and disciplinary governance arrangements. Along the interdisciplinary dimension, the new evaluation practices has inserted standardised notions of research quality and research performance and objectified disciplinary reputations. Despite a strong science policy tradition to focus on the social and economic benefits of university research, scientific publications and performances dominate the evaluation of research, as seems to be the case in the British RAE. Even within the Royal Academy of Arts and Sciences there is concern that this emphasis on international scientific publications is too strong and that other outputs than publications in peer-reviewed journals need to be rewarded as well.

Along the institutional dimension, the evaluations have rearranged the relationships between government, university, departments and researchers. Evaluations were started to empower the government in setting priorities and reallocate funding. This has not occurred. Instead, university management has become a key actor in the institutional governance arrangement. For the government it functions as a gatekeeper, being able to buffer interventions from the government. For research groups, the evaluations provide them with the means to make programme leaders and faculty deans accountable. This intermediate position rested on the organisation of research into assessable research programmes and the organisation of evaluations in a standardised practise. At the beginning of the 21st century, it seemed a stable configuration had settled down. However the recent delegation of responsibility for the evaluations towards the university level has put the configuration under tension again and it remains to be seen how evaluations and governance arrangements evolve in the near future.

REFERENCES

Ball, D. F. and A. Verkleij (1999), *University research as a business: a comparison of research assessment exercises in the United Kingdom and the Netherlands*, unpublished manuscript.

Blume, S.S. and J.B. Spaapen (1988), 'External Assessment and "Conditional Financing" of Research in Dutch Universities', *Minerva*, 26, 1-30.

Callon, Michel (1986), 'Some elements of a sociology of translation: domestication of the scallops and the fishermen of St. Brieuc Bay', in John Law (ed.), *Power, Action and Belief: A New Sociology of Knowledge?*, London: Routledge & Kegan Paul, pp. 196-233.

Clegg, Stewart R (1989), *Frameworks of Power*, London: Sage.

Cooper, Robert (1992), 'Formal Organization as Representation: Remote Control, Displacement and Abbreviation', in Michael Reed and Michael Hughes (eds.), *Rethinking Organization: New directions in organizational theory*, London: Sage, pp. 254-272.

Geuna, Aldo and Ben R. Martin (2003), 'University research evaluation and funding: an international comparison', *Minerva* 41, 277-304.

Laclau, E. and C. Mouffe (1985), *Hegemony and Socialist Strategy*, London: Verso.

Law, John (1986), 'On the methods of long-distance control: Vessels, navigation and the Portuguese route to India', in John Law (ed.), *Power, Action and Belief: A New Sociology of Knowledge?*, London: Routledge & Kegan Paul, pp. 234-263.

Merton R. K. (1968), 'The Matthew effect in science', *Science*, 159 (3810), 56-63.

Morris, Norma (2002), 'The developing role of departments', *Research Policy*, 31, 817-833.

OECD [Organisation for Economic Cooperation and Development] (1997), *The Evaluation of Scientific Research: Selected Experiences*, Paris: OECD.

RAWB [Raad voor Advies van het Wetenschapsbeleid] (1969), *Vervolg-interimadvies inzake de overheidsuitgaven voor onderzoek en ontwikkelingswerk*, Tweede Kamer, zitting 1968-1969, 9800 hoofdstuk VIII, nr. 50.

Shore, Cris and Susan Wright (2000), 'Coercive accountability: the rise of audit culture in higher education' in Strathern, Marilyn (ed.), *Audit Cultures: Anthropological studies in accountability, ethics and the academy*, London: Routledge, pp.57-89.

Van der Meulen, B.J.R., D.F. Westerheijden, A. Rip and F.A. van Vught (1991), *Verkenningscommissies tussen veld en overheid; Evaluatie onderzoekverkenningscommissies*, Achtergrondstudies Hoger Onderwijs en Wetenschappelijk Onderzoek 8, Zoetermeer: Ministerie van Onderwijs en Wetenschappen.

Van der Meulen, B. J. R. (1995) 'Understanding evaluation processes in research systems in transition', *Science Studies*, 8, 24-35.

VSNU [Vereniging van Universiteiten] (1998), *Standard Evaluation Protocol Research*, unpublished document, Utrecht.

Westerheijden, D. F. (1997), 'A solid base for decisions: use of the VSNU Research Evaluations in Dutch Universities', *Higher Education*, 33, 397-414.

Whitley, Richard (1984), *The Intellectual and Social Organization of the Sciences*, Oxford: Oxford University Press.

APPENDIX: PROTOCOL OF VSNU RESEARCH EVALUATIONS[6]

Protocol of VSNU Research Evaluations

The protocol includes a classification of the disciplines as well as a rolling scheme for evaluation of the disciplines.

After consultation of the involved departments, the VSNU determines a time schedule for every evaluation

The directly related standing disciplinary committee of the VSNU nominates two or more candidates as chairman of the evaluation committee as well as decides upon a profile of the expertise of the committee members.

The VSNU appoints a chairman, after consultation of the Royal Netherlands Academy of Arts and Sciences KNAW. The chairman, in consultation with the KNAW, puts together a committee of which the majority of the members are from abroad. The working language of the committee is English.

The directly related standing disciplinary committee of the VSNU specifies within the discipline-specific protocol the terms of reference for the committee.

Based upon the general protocol and the discipline-specific protocol the involved university departments make self-assessments of their performances of the last five years and describe their future plans. The unit of evaluation is a research programme. Of each programme, five key publications are put up as part of the self-assessment. In addition, a profile or mission statement of the department is requested.

The evaluation committee is requested to judge, for each programme, its quality, productivity, relevance and viability on a five-point scale. For each programme, a brief explanation of the scores is given, which might nuance the general judgement. In addition, the committee gives an assessment of the state of the art of the discipline and of each department.

The committee's judgements are based on documents, complemented by interviews with programme leaders and the department's management. Especially in the engineering and natural sciences the committee will make site visits.

The report will be finalised and stipulated and presented to the VSNU chairman, after the departments have got the possibility to react on the draft report.

The costs of the evaluation are covered by the universities involved in evaluation.

The costs depend on the discipline and the discipline-specific protocol. The base costs are Dfl.27,000 [approximately 12,000 €] per university.

[6] Source: VSNU (1998), translation by the author.

CHAPTER 10

LAURA CRUZ-CASTRO AND LUIS SANZ-MENÉNDEZ

RESEARCH EVALUATION IN TRANSITION

Individual versus Organisational Assessment in Spain

INTRODUCTION

Research evaluation has been an essential practice of the regular functioning of the research system (Zuckerman and Merton 1971; Cole and Cole 1973). Reputational competition (Merton 1957; Ben-David 1971, 1972; Dasgupta and David 1994; Whitley 2000) has been shaped by mechanisms of evaluation of research mostly identified with the practice of peer review for journals' publications (Campanario 1998a, b; Cole 1998). Some of these practices for publishing papers or awarding prizes have been extended to the allocation of the funding for research from governments or intermediary organisations (Chubin and Hackett 1991; Cole et al. 1978).

More recently, state research evaluation systems (RES)[1] have been developed in a number of countries in the context of new public management practices, scarce public funds and increasing accountability requests (Georghiou 1995), and the allocation of resources for organizations and programs has become more and more connected to the evaluation of research (Geuna and Martin 2003; Liefner 2003). Additionally, the dominant *ex ante* or project appraisal approaches have been complemented by the institutionalisation of retrospective *ex post* evaluations of research performance[2], as the papers by Kneller, Cozzens ands others in this volume discuss.

In terms of the overall governance of the public research system, evaluation carries out two main functions. On the one hand, it can be considered as a 'steering or management tool', that is, as an instrument for organising and managing research activities (Callon et al 1995). As such, evaluation may be directed to research related organisations (universities, research centres, and management institutions) in order

[1] We understand research evaluation system as the ensemble of practices and institutional arrangements in a country mediating between scientific quality controls and research policies.

[2] For a recent review of different evaluation practices associated to R&D see the work by Luke Georghiou and Philippe Laredo for the OECD (2006).

R. Whitley and J. Gläser (eds.), The Changing Governance of the Sciences, 205–223.

to improve their functioning; or it can be oriented to policy making, through the evaluation of science and technology (S&T) programmes, and to improve research policies.

On the other hand, along with the steering function, research evaluation can be used in a distributive way whereby it is used to allocate different kinds of rewards and resources among different types of actors - individuals, groups or organisations - to improve their research performance. These incentives may be economic (grants, salary bonuses) and/or symbolic (reputation and prestige). Research evaluation as a distributive instrument can, then, affect the funding of research organisations, research projects, or the allocation of rewards to individual researchers. In some countries, we find the combination and integration of both functions in single funding instruments, while in others they are separated.

The aim of this paper is to analyse the institutionalisation of the Spanish RES in the context of the transformation of the research organizational field[3] (Cruz-Castro and Sanz-Menéndez, 2007) and the public research system (PRS). The co-evolution of the changing PRS and RES is particularly interesting in Spain because of the effects of increasing political decentralisation and the importance of individual evaluations rather than organisational ones. The limited financial resources for research in Spanish universities and their dependence on the success of individual researchers in obtaining research grants through competitive bidding has meant that they have weak strategic capabilities[4], especially in comparison with those of Australian, British and US research organisations.

More specifically, we deal with the following questions:

- What explains the emergence in Spain of a RES focused on individuals rather than on organisations?
- Why is research evaluation at the level of organisation marginally connected with research funding?
- How has the decentralisation of science and technology policy affected the RES?
- How has the changing RES affected the organisation and functioning of the PRS?

During the last decade, research evaluation has become seen as essential for the steering of the public research system and there has been considerable effort directed to the different actors across the system to encourage the adoption of evaluative habits and structures. Nevertheless, we suggest that the evaluation focus continues to be at the individual level of researchers and research groups with minor developments regarding organisational and program evaluation.

The chapter is organised as follows: the next section describes the basic institutional features of the Spanish academic system and its research funding regime. This is taken as a point of reference in order to analyse, in section 3, the institutionalisation process and characteristics of RES, including some of the

[3] We borrow the concept of 'organisational field' from the institutional approaches of organisational theory (see for example DiMaggio and Powell 1983).

[4] We recall the issue because of the negative consequences for performance of highly autonomous scientists in loosely coordinated organisational settings (Pelz and Andrews 1966).

explanatory factors that account for these developments. Section 4 presents a summary of the arguments and deals with the feedback effects that new trends and institutional arrangements of the RES have had on the public research system.

THE ACADEMIC SYSTEM: GOVERNANCE AND FUNDING

For the purpose of this chapter we understand the Spanish academic system to be composed of two different subsystems: the universities (represented in the *Frascati Manual for R&D statistics* as 'Higher Education') and the *Consejo Superior de Investigaciones Cientificas* (CSIC), which is the largest the Public Research Centre (included in the R&D Statistics under the label of 'Government'). In 2004, higher education represented 29.5% of the total R&D expenditures, meaning 0.32% of the Spanish GDP and 51% of the researchers (FTE) and 39% of the R&D personnel. The government sector as a whole represented 16% of the total R&D expenditures, meaning 0.17% of the Spanish GDP, 17% of the researchers (FTE) and 17% of the R&D personnel; and CSIC represented approximately one third of the Government sector.

Two single words characterise the Spanish academic system over the last 25 years: Growth and change. On the one hand, numbers have more than doubled in the main variables of the system, such as the number of universities, professors, lecturers, students enrolled, graduate and doctoral degrees granted, etc. On the other hand, universities, and to a lesser extent public research centres, have significantly changed their way of governance and functioning.

Size and Governance

In 2004 there were 69 universities in Spain, 21 of them private. However in terms of the number of students enrolled and the number of professors and lecturers, the public universities represent approximately 92%. The 48 public universities had almost 88,000 professors and lecturers, among them 50,500 with the condition of civil servants – permanent staff – engaged in research and teaching of 1.36 million enrolled students; the public universities produced around 182,000 graduates and 7,100 PhDs in 2003/2004.

Some historical figures provide a clear picture of the growth of the system. In 1983, the number of universities was 33 (3 of them private and owned by the church[5]), with almost 31,000 professors and lecturers, engaged in teaching of 700,000 enrolled students; the public universities produced almost 80,000 graduates and 1,900 PhDs per year.

However, transformation has not been just the result of the growing demand of the Spanish society for higher education and the increase of public budgets for education; changes were also the result of the transformation of the university in two important dimensions: Firstly, the universities have moved from 'bureaucratic centralism' dominant at the Dictatorship times to a 'self-regulation' mode of

[5] Until the early 1990s the only private universities existing in Spain were those related with the Church, a privilege that they got in the times of Francoist dictatorship.

governance (Sánchez-Ferrer 1997) and, secondly, from being just teaching universities, they have evolved into a model of the university that has, simultaneously, set up research and third mission activities (García and Sanz-Menéndez 2003).

The Spanish 1978 Constitution recognised the autonomy of universities (Spanish Constitution, article 27.10), while the 1983 University Reform Act (LRU) defined the constitutional arrangements for university governance, management and functioning. Although universities were defined as self-governing bodies, they have been highly dependent on public funds; therefore, despite the high level of 'autonomy' the universities were rather poor in financial terms and consequently quite dependent on the political authorities.[6] Additionally, the legal reforms enabled the universities to recruit and select their own academic staff and to appoint, after an 'examination procedure' managed by the university, new professors with civil servant status (Mora 2001).

Apart from the academic staff recruitment procedures, the governance of universities has also followed quite autonomous mechanisms, developed under principles established by the 1983 Universities Act but implemented by the specific procedures in each university (through their own statutes). University authorities (rectors, vice-rectors and deans) are elected by their own constituencies, which include permanent professors, temporary lecturers, administrative personnel and students, and their 'responsiveness' to society depends basically on their will and the 'financial pressure' that governments could exert.

Between 1985 and 1996, as part of the decentralisation or federalisation of Spain, supervision and control of the universities were transferred to the different regional governments.[7] Additionally, some regional governments have created new public universities, either transforming former 'colleges' in some provinces into universities, or creating from scratch, in order to reduce students' enrolment pressure on the old universities. In any case, the growth of universities in provincial capitals has become strongly related to local and regional politics.

While most aspects of the governance arrangements of universities have been stable for almost two decades, some significant changes have taken place in their regulatory environment with the approval of a new Universities Act (LOU) in 2001. This Act represented a significant increase in the regulatory powers of regional governments and many regional governments have approved Regional Universities Acts since then. In fact, the 2001 Universities Act gave legal recognition to an emerging process of differentiation of relationships between the regional authorities and their universities. Given the diversity in the capabilities of regional governments to implement steering mechanisms, a broad variety of outcomes is likely to emerge.[8]

---

[6] Estimates of aggregated income of universities coming directly through regular transfers from public budgets (either regional or national) amount to 78% of the total (Hernández Armenteros 2004).

[7] The national government only has direct control and supervision capabilities over the UNED (a distance learning university like the Open University in UK) and UIMP (a 'summer courses university' that does not yet provide degrees).

[8] There are 8 Regions (Basque Country, Balearic Islands, Asturias, Cantabria, Navarre, Castile-La Mancha, La Rioja, Extremadura) each of them supervising only one university. In those regions, on many occasions, the Chancellor or Rector of the University is a more relevant person in the region than

Under the new regulations, the Chancellors of universities are elected following universal democratic rules; and consequently the university management structure often seeks re-election to take care of the interest and demands of their constituencies. Whilst this change represented a move towards increased university autonomy and increased internal accountability, the 2001 Universities Act also introduced a significant change in the mechanisms of access to the civil servant status for university professors or to get a university contract, namely, national *habilitation* and accreditation respectively. The reasons for this re-centralised quality control - with respect to the recruitment of academic staff - lies in one of the side effects of the decentralisation of selection established in the 1983 Universities Act, that is, a high degree of inbreeding and the consolidation of internal labour markets dynamics within university departments (Cruz-Castro and Sanz-Menéndez 2006).

In addition to the universities, research is conducted by CSIC, which is the largest public research organisation in Spain. This is an umbrella organization – similar to the Max Planck Society in Germany or the CNRS in France - with more than a 100 institutes all over Spain, over 10,000 employees and 2,500 tenured scientists (Sanz-Menéndez and Cruz-Castro, 2003). In contrast with the Universities, CSIC has not been transferred to the regional authorities. Despite the fact that CSIC authorities, appointed by the Minister of Education and Science, have a significant discretionary power of allocation of resources, they also try to get support and consensus from their researchers and institutes. CSIC has been – until the seventies - the 'reservoir' of public research in Spain (OECD, 1964), always in strong interaction with universities. It is different from other public research centres for several reasons. Firstly, it is not mission-oriented; secondly, only researchers with a PhD can obtain tenure as academic staff; and thirdly, CSIC researchers are involved in the competition for the same research funds, and they are also the only ones subjected to the same evaluation systems as university professors.

Funding

In the context of the formal autonomy of universities and researchers to pursue their own research objectives, the way in which funding is organized is a critical element in evaluating the ability of governments to steer research activities (Braun 1993; Whitley 2003). The standard Anglo-Saxon literature on the relation between science and politics usually interprets the increasing relevance of competitive project-based funding for research as a signal of the demand of the authorities to the research community for more responsiveness to the programmatic research goals defined by the government. However, the Spanish situation does not fit this model because neither universities nor even the public research centres obtain significant amounts of stable block grant funding for research.

At the time of the transfer of the universities' control from national Government to the regional ones, the funding system was relatively homogeneous. It followed an incremental line item budgeting, in which each single item of expenditure of the

the President of the Regional Government. Bigger regions with more than one public university are Andalusia (10 universities), Catalonia (8 universities), Madrid (6 universities) and some others.

budget was increased with respect to the budget of the previous year, but a system based on a formula model was emerging, where the main criteria were teaching loads (number of registered students) and the numbers of teaching staff; with almost no block grant funding for research.

Today, university funding has been decentralised to the regional authorities, so the situation in each university varies depending on the strategies and priorities of regional government. The annual transfer for university funding is included in the Regional Government Annual Budget that the regional parliament approves at the end of each year. Due to the very different political priority assigned by the different regional authorities to the higher education institutions and research policy, the mechanisms used by governments to finance universities are quite diverse. Incremental line item budgeting has been replaced by two types of models, which in many cases go together: formula models and contractual arrangements (González López 2006). The first one is usually based on different combinations of students' enrolment, size of the staff and other numerical data. In contractual agreements (*Contratos Programa*) the funds are usually linked to the accomplishment of goals or requisites previously agreed (see Lange in this volume on a similar practice in Germany).

CSIC funding from the government has traditionally followed an incremental line item budgeting. In the budgets of 2005 and 2006 a political decision to increase R&D budgets has meant a very significant increase of more than 20%. This increase has opened the possibility, along with new legal changes, of developing a model of relationship based on contractual agreements with performance indicators. In late 2005 and early 2006 the first steeps in that direction have been taken (Fernández de Labastida 2005), and a legal change is on the way.

In the absence of any significant block grant funding for research activities, university professors and CSIC researchers need to obtain research funds through competition from funding institutions (national or regional) or through contracts with companies. In the mid 1980s a national external funding system for research in universities and CSIC was established[9] for the first time, in a model labelled as a 'quasi research council' (García and Sanz-Menéndez 2005). This model has distinctive characteristics as regards Research Councils as we know them. First, the majority of the funding bodies depend directly on Ministries, and the Heads of these bodies are normally political appointees. Secondly, these organisations lack major administrative capacities since they are weak bureaucracies with little permanent staff and broadly populated by the agents of the system: the researchers. Finally, and most importantly, the quasi-research councils operate in a context of unclear boundaries between principals and agents, but they can, nevertheless, enjoy a high degree of institutional stability along the years.

The primary mechanism for government implementation of Spanish R&D policy has been through the funding of research projects, and this has accounted for most of the non-specific objective funds earmarked for the public research system. These budget funds are awarded through an annual public call for proposals, usually for

[9] For a deep analysis of the institutional construction of the Spanish science and technology policy see Sanz-Menéndez (1997).

3-year research projects. It is in this project-based context that the Spanish type of research evaluation has been strongly associated with the allocation of funds.

Additionally, the regionalisation of the country has produced a distinctive feature of the research funding system: multilevel dynamics (Sanz-Menéndez and Cruz-Castro 2005). The increased involvement of regional authorities in providing competitive funding for research has increased the pluralism in the system with respect to the definition of research objectives, without reducing, however, the problems that Spanish research institutions and organisations have in defining strategic behaviour and solving the collective action problems of their researchers.

Some tension and conflict between the scientific and political rationales of research resources distribution among research organisations can occur. Although the formal organisational autonomy of the universities with respect to the government and the political system has been reinforced over time in terms of self-governance, there is significant financial dependence of the universities on their regional authorities. When the research organisations and political arenas are too close, there is always a possibility of a kind of practice of allocation of resources based on interests groups' politics, rather than on the scientific logic of the best performance or on a more explicit managerial approach. The political distributive rationale at the regional level tends to be egalitarian rather than discriminating between research organisations, however the aggregate effect of the 17 different regional policies could produce differentiation.

INSTITUTIONALISATION PROCESSES AND CHARACTERISTICS OF THE SPANISH RES

In the mid-1980s, together with the dramatic growth in the budget earmarked for competitive R&D funding, there was a considerable increase in the use of peer review in funding allocation decisions. Research evaluation arrangements were institutionalised in 1986 by means of the Act for the Promotion and General Coordination of Scientific and Technical Research (Science Act). Under this Act and its developments, the funding of research activities was organised around a National R&D Plan including: (1) targeted research, articulated around priority programmes, and basic research articulated by the programme for the General Promotion of Knowledge,[10] and (2) reliance on peer review as the legitimate selection mechanism prompted by the creation of the National Agency for Evaluation and Foresight (ANEP), managed by the research community.

Overall the existing funding regime is basically a project-based one, with some elements of a programmatic regime; however the influence of the academic research community in the selection of national priorities was, and is, quite determinant. In addition to a high degree of competition, there is a significant degree of autonomy in pursuing research objectives, even in the context of national S&T priorities. The

[10] The structure of the National R&D Plan, organised in targeted and non-targeted programmes, was actually rhetorical. It did not have – *ex ante* – any resources assigned by areas, but waited for the demand for funding and the quality of the proposals.

basic features of the Spanish research-funding regime and some interaction with the RES can be summarised as follows:

- High university political autonomy but budgets largely determined externally and mainly dependent of teaching loads or students' enrolment;
- Essentially, a peer review project-based funding regime, with elements of a programmatic regime (Whitley, in this volume);
- University strategic management of research has limited impact on research groups, as compared to scientific peer's pressures and the steering of research towards priorities set up by funding bodies;
- Increasing diversity of funding sources for research as result of the regional authorities involvement in science policy;
- Public research centres have had historically less autonomy, and greater presence of block grant funding than universities, but this has changed over time; and
- High variance of evaluation standards across scientific fields, but with some trends towards convergente.

Institutional Arrangements of the Research Evaluation System (RES)

The 1983 Universities Act represented an almost complete transfer of the responsibilities of selection and access to the civil service of university professors to the universities themselves. In contrast, the arrangements of research evaluation set up with the 1986 Science Act were associated with a more centralised model of managing science and technology policy. In fact, analysing the underlying logic of both Acts one could say that University Reform Act was shaped by a liberal and self-organised model of responsibilities, while the Science Act had a much more planning oriented and interventionist model, even shaped by Bernalist models of science policy.

In the early times of the Spanish S&T policy in the 1980s, the emerging academic elite associated to the new socialist government was very much concerned about the procedures of the allocation of competitive funding (Sanz-Menéndez 1997), which in the past had been mainly based on the hierarchical approach of the senior professors in each field. The creation of an independent evaluation space, managed by scientists, as a first step into the project funding process, became a clear objective. The reforms that took place in the mid 1980s considered the set-up of evaluation structures as a priority in order to build a coherent S&T policy in Spain. The main focus was on developing a peer-review system guaranteed by the State (Sanz-Menéndez 1995) as a mechanism for research evaluation directed to the allocation of public research funds.

The National Agency for Evaluation and Foresight

A unique institutional arrangement was adopted in Spain with the establishment of the National Agency for Evaluation and Foresight (ANEP). The ANEP was created in the Inter-ministerial Commission for Science and Technology (CICYT) - the

inter-ministerial political-planning body in charge of the R&D policy - but with autonomous scientific management. Its mission was defined as

> … the scientific and technical evaluation of entities and research groups that participate in the implementation of programmes and projects of the National R&D Plan, the proposals of both R&D actors and operators, and the monitoring of results (outputs and outcomes) that could be produced in the development of those programmes and projects. (Decree RD 415/1987, 6th March, "Organic structure of the Permanent Commission of the CICYT", our translation)

The ANEP had also the mission of developing foresight activities in scientific research and technological development, but in practice the overload of work and its weak organisational capabilities precluded a serious development of any activity other than *ex ante* project appraisal. Already in 1995 the bulk of ANEP activity was scientific and technical evaluation for project selection and funding, general scientific and technical assessment and technical advice to political bodies, with single cases of evaluation of research organisations (Sanz-Menéndez 1995).

The evaluation and selection of project applications is implemented as a two-stage process in which two or three individual peers, using a mail procedure, make a first assessment of the submissions; then, a panel of experts make the final funding decisions. Thus, ANEP does not control the final approval of the projects reviewed. Its reviews are just one of the inputs, albeit a relevant one, in the process of selection. The overall project funding process was identified as 'dual' because after the individual peer reviews by the ANEP, the 'priority' adequacy to the research objectives and priorities of the National Plan were established by a panel from the funding body. The effects of ANEP's activity can be shown in the fact that approximately 50% of R&D projects evaluated for the Directorate General for Research were rejected (García and Sanz-Menéndez 2005).

The Spanish peer review model used for funding research has been based upon two critical 'roles', usually filled by academics in part-time jobs: 1) the coordinator of each scientific area, appointed by the ANEP; and 2) the manager of each scientific programme at the R&D funding units. Coordinators select the reviewers from a pool of academics. This selection is based on a mix of criteria such as scientific specialisation, research expertise, etc. Programme managers are responsible for appointing the panel of between 8 and 20 experts that will complement the assessment of each project, to which they assign scores. These new scores, together with those from ANEP, support the final decision on whether or not the proposal is funded. The evaluation criteria are the usual ones: the contribution of the proposal, research design, quality of methodology and past performance of the principal investigator and research team.

Fifteen years after its creation, the ANEP continues to be an administrative unit of the Ministry of Education and Science under the authority of the State Secretary of Universities and Research, and organises its research evaluation activity autonomously from the management of the R&D funding programmes and bodies. Funding bodies, both internal to the Ministry of Education and Science and external as parts of other Ministerial funding or Regional governments, often request the support of the ANEP in the evaluation of the project proposals or the approval of individual fellowships.

The National Commission for the Evaluation of Research Activity

The second institutional element of the Spanish RES emerged in the late 1980s. It was the establishment of *ex post* research performance evaluation procedures of individual researchers and the creation of the National Commission for the Evaluation of Research Activity (CNEAI). This move must be seen as part of the institutionalisation of individual incentives and rewards for research activities. The 1983 Universities Act (article 11) approved the possibility for university professors (and latter on the Science Act did the same for the CSIC researchers) to get additional personal income, despite their civil service status, from contract research with private entities.

A response to the movement of many university professors into 'contract research' was the attempt to increase the wages of those working, mainly with publicly funded projects, and to reward their publication profile. In 1989 the government approved a voluntary-based system of periodical (every six years) evaluation of individual research outcomes,[11] and the construction of an institution dedicated to that task: the CNEAI.

The CNEAI was institutionalised as a mechanism for evaluating academic careers and the research performance of tenured researchers. It was organised as a way of providing incentives for research activities. The rewards were small salary increases for university professors and CSIC researchers in exchange for the recognition of good research performance (reflected in a positive evaluation). Once a year, tenured academics and researchers may submit five contributions to a panel of peers that examine their career on a six-year period base. The recognition of these six-year periods of research activity leads to an automatic increase in salary, and increasingly constitutes a reputational legitimising element for the researcher (Sanz-Menéndez 1995).

The original idea was simply to reward research performance, mainly based on the concept of 'contributions', and mostly reflected in the publication of papers in international journals. The evaluation procedure is organised in 11 large research domains and the examination of the contributions submitted is made by a small set of publicly appointed experts.

The institutional arrangement of the CNEAI is quite soft, as it involves experts appointed by the Minister of Education and Science and representatives of the Regional Governments. This evaluation mechanism is arranged by the national authorities, but the small increases of wages of university professors is paid by the regional governments, because of the institutional dependence of the universities at this level. The individual application for evaluation is voluntary and if the outcome is positive, the effect is a small permanent increase in the researcher salary (approx. 110 euros/month gross, meaning approximately 3% of the total annual income).

Other forms of institutional evaluation directed to the improvement of the management of R&D programs and research organisations had a marginal presence in mid 1990s. The CICYT did not include any formal or explicit decision to

[11] Amazingly a similar system of complementing the wages of the academics, the National System of Researchers (*Sistema Nacional de Investigadores*), was established in Mexico in the same years (see Schoijet and Worthington 1993, despite their confusing interpretation of the system).

systematically perform program evaluation, and the same occurred with other R&D management bodies (Sanz-Menéndez 1995). A final type of evaluation developed during the 1990s might be termed as 'evaluative studies of R&D policies'. These were studies and research projects not formally commissioned, and not integrated or associated with the policy-making process, although some of them achieved informal acceptance and collaboration of the policy-making bodies, but with no commitment to implement the results. They were more academic exercises than proper research evaluations; their design did not provide clear definitions of the purpose or clear evaluative criteria, nor were the methodologies well suited to generating policy feedback.

In sum, formal research evaluation up to the mid-nineties was still developing. Peer review was the main method employed, which reflected the main role conferred to the 'clients' of R&D policies. Additionally, evaluation was oriented mostly to *ex ante* evaluation for project funding purposes and individual research performance recognition linked to the provision of small salary increases for tenured researchers with significant reputational effects. On the other hand, organisational evaluation directed to strategic purposes constituted an exception in this context, where it appeared only under very exploratory actions with little practical consequence.

New Developments in the Spanish RES

Two institutional developments have taken place over the last decade in the Spanish RES. Firstly, there has been the emergence of systematic evaluations of universities which focused on teaching quality rather than research, and did not have funding effects. The main instrument here has been the National Plan for Quality Assessment of Universities (PNECU). The PNECU was headed by the Council of Universities, an organisation composed of representatives of regional and national governments and the rectors of all universities. The process was managed by a Technical Committee, composed by Council's officials and experts.

The Plan started in 1996, with three following rounds (1998, 1999 and 2000) before the 2001 Universities Act, when this type of evaluation was institutionalised. The underlying rationale was to detach the process of organisational evaluation from funding and accreditation, letting each university develop their own quality policies to improve their products and services (Bricall Report 2000). The main objectives of the PNECU were: a) To promote the institutional assessment of university quality; b) to draw up homogeneous methods of assessing university quality in line with the practices currently used in the European Union; and c) to provide objective information which may be used by the various organisations to aid decision making in their particular area of expertise.[12] The Plan evaluated three main activities: a) teaching in degree programs; b) research in the departments to which the programs were assigned; and c) management in the services attached to the programs.

The methodology deployed by the PNECU consisted of a mixed system of self-assessment and external assessment, and, as a final step, the writing and publication

[12] Article 1 of the Decree, RD 1947/1995 of 1st December, establishing the PNECU.

of a final report. A major achievement of the Plan was to introduce and popularise the culture of evaluation among higher education agents. Almost all universities have participated in the Plan. The exercise also succeeded in establishing quality control units or departments in almost all universities, and significantly improved the information and statistical systems within the institutions. Although the Plan was more oriented to teaching activities than to research ones, research activities in universities were partially reviewed, paying attention to different institutional dimensions: scientific production, external relationships, human resources, support staff, economic resources, material resources and infrastructures, doctoral programs, research groups, university's internal support for research, promotion activities, internal communication and collaboration, and departments' internal promotion of research.

The second important development within the 2001 Universities Act has been the extension or stretching of individual performance evaluation, which put into motion new recruitment procedures and quality requirements (accreditation and *habilitation*). With regard to the early stages of the academic career in public universities, the Act introduced some significant changes. It created new forms of pre-tenure teaching and research positions, and in order to access to some of these contracts, candidates must previously have been accredited by the appropriate agency. In this accreditation process, a body of experts assesses the research and teaching merits of the candidates.

In addition, a quite complex system of national *habilitation* for achieving civil servant status was established. Once a year, national *habilitation* exams were organised. The number of *habilitations* is determined by the demand of the universities. The universities set up access exams for their tenure positions and select among the candidates entitled. However, many universities do not issue the call until they have got their own temporary professors habilitated, maintaining some inbreeding practices.

Both developments, those related to 'quality assurance and improvement' and 'individual national accreditation' have been associated with the creation, in the 2001 Universities Act, of a National Agency for Quality Assessment and Accreditation (ANECA). This agency, in addition to pursuing the institutional quality assessment initiated with the PNECU, started the accreditations and evaluations for university teaching staff. Most interestingly it has performed a significant task in developing methodologies and publishing evaluation procedure manuals.

Following this institutional innovation, several similar regional agencies have been set up. Any Spanish university can contract individuals accredited by the national agency, whereas individuals accredited by a regional agency can only be contracted by the universities located in the respective region. The consequences of this fragmentation for the mobility and transparency of the academic labour market is still unclear. Nevertheless, *habilitation,* the official competitive evaluation

required for university academic tenure, remains centralised as a reaction to the traditional high degree of inbreeding and low mobility of research personnel.[13]

Many regional authorities, empowered by the 2001 Universities Act, have created new accreditation and quality assurance agencies, leading to the multiplication of evaluation structures. In 2005, 9 out of 17 Regional authorities had set up university evaluation institutions.[14] Two of them (Catalan and Andalucian) started to work in the context of the implementation of the PNECU, and the others in the context of the 2001 Universities Act. These 'agencies' have diverse legal statuses, most frequently that of a 'consortium' between a university and its regional government.

Most of the agencies have been granted broad competencies related to evaluation and quality assessment, and the most important are: 1) Institutional evaluation (which includes evaluation of academic programs, management and services, and academic and research functioning); 2) Accreditation and quality certification activities; 3) Individual evaluation (which includes teaching staff accreditation and research and academic teaching staff evaluation in order to get regional salary bonuses); 4) Assistance and advice related activities (including policy planning). However, in practice, most of the activities of the regional evaluation agencies have focused on just two missions: individual evaluation[15] and some institutional assessment related with the quality of the teaching activities and services. An emerging field is the preparation of the changes associated to the implementation of the European Higher Education Space.

In accounting for these developments, two main factors are important. First, the expansion of undergraduate education has been the driver of the growth of universities in the last twenty years, and this explains partly the focus of the evaluation exercises on teaching quality rather than on research. Secondly, in explaining the multiplicity of evaluation structures, the most important factor has been the growth of the research system. This growth, in terms of research centres, laboratories and researchers, contributed to the overload of the national system for evaluation and to the conditions in which some regional authorities created regional

[13] At the time of revising the paper in July 2006, the Spanish government has sent to the Parliament a new Act on Universities that introduces changes in the *habilitation* procedure, mainly its transformation into a simple national accreditation.

[14] These are: Andalusia (*Agencia Andaluza de Evaluación* -AGAE), Balearic Islands (*Agència de Qualitat Universitària de les Isles Balears* -AQUIB), Canary Islands (*Agencia Canaria de Evaluación de la Calidad y Acreditación Universitaria* -ACECAU), Castile-La Mancha (*Agencia de Calidad Universitaria de Castilla-La-Mancha* -ACUCLM), Castile-León (*Agencia para la Calidad del Sistema Universitario de Castilla y León* -ACSUCYL), Catalonia (*L'Agència per a la Qualitat del Sistema Universitari a Catalunya* -AQU), Galicia (*Axencia para a Calidade do Sistema Universitario de Galicia* -ACSUG), Madrid (*Agencia de Calidad, Acreditación y Prospectiva de las Universidades de Madrid* -ACAP) and Valencia (*Comissió Valenciana d'Acreditació I Avaluació de la Qualitat* - CVAEC).

[15] The main activity of these agencies has been the individual evaluation of university teaching staff. This kind of evaluation has been oriented towards two main tasks: the recruitment of new teaching staff according to the new categories included in the 2001 Universities Act; and the allocation of regional salary bonuses. According to the available data, the evaluation and accreditation of teaching staff is the only activity performed by all the agencies. Moreover, the lack of common standards in the evaluation criteria threatens to jeopardise the meaning of accreditation itself.

evaluation structures in support of their local scientific policies. Regional governments have consolidated as active new actors in the system, contributing, very significantly, to R&D expenditures *vis* a *vis* the central state.

Compared with 1995, there are not only more researchers, but also more funds coming from national, regional, and supranational R&D institutions. Thus, the growth in demand has been sustained by the multiplication of financing sources. Funding models and evaluation practices have been diffused or imitated across levels and among regions. Accepted national practices have diffused quite widely, despite the fact that some experimentation and new developments have emerged in some regions.[16] In a few cases, this has even led to the duplication of evaluation institutions, since, for example, some regions have set up their own research performance based salary bonuses, applicable only to the academics in the universities located in their regions.

FEEDBACK EFFECTS BETWEEN EVALUATION AND RESEARCH SYSTEMS

Evaluation occupies a substantial part of the new discourse adopted by Spanish research actors, but assessment practices have not been fully consolidated, especially at the organisational level. Overall, the research evaluation activities performed by the ANEP and the CNEAI, mostly related to project and individual research performance evaluation, continue to be the key elements of the Spanish system. The balance of our analysis can be summarised as follows.

First, in the context of the increasing intervention of regional authorities in S&T policy, research funding schemes based on competitive project funding have proliferated in Spain. In this context, pluralism has increased. As a result of the support of regional governments for their own centres and universities, we are probably moving from a public research system largely based on "competitive pluralism" to one more composed of "competitive hierarchies" (Whitley 2003).

Due to the regionalisation of S&T policies, a major response from regional governments as actors in the system has been the setting up of parallel funding and evaluation structures. The result has been the diversification of external sources for project funding and *ex ante* evaluation structures. The outcome of this process has been the multiplication of evaluation standards and criteria for project funding.

The regional authorities' involvement in the S&T policy domain and the creation of regional research evaluation structures have resulted in an increased fragmentation of the RES as a whole. Regional governments have developed their own R&D policies autonomously from national authorities, and this has prompted the proliferation of several regional public research systems, each with the capacity to establish their own criteria for allocating funds or designing programs. The existence of regional evaluation structures parallel to national ones, using different

[16] For example, the Catalan Government has also promoted a non for profit institution ICREA that is playing a critical role in the improvement of the quality of the Catalan system with prizes and grants, but again the focus has been the selection of the best individuals, not the change of organisational practices.

standards and criteria, might create problems of duality of markets and legitimacy deficits. Due to the decentralisation processes and the multiplication of funding and evaluation structures, our prediction is that some of the regional subsystems will evolve into stronger RES and some of them will not, depending largely on the relative weight of academic elites into the policy decision-making processes.

Second, the consolidation and deepening of individual researchers' evaluations has become established as a core practice in the Spanish RES. The model developed in the late 1980s and early 1990s that focused on rewarding the research performance and granting reputation to the researchers has become a 'generalised practice' at national and regional levels. Most of the Regional Governments have created their own reward systems to provide some additional income to their researchers. Moreover, the requirements for accreditation for the entry phases of an academic career, and the *habilitation* for the consolidation of tenure, have become generalised across the system. Accordingly, the implementation of the accreditation has been duplicated by nine regions that have also created their own accreditation and quality assurance entities.

Third, although the activities of ANECA and the regional entities in the domain of institutional evaluation and certification etc. have grown in the past years, in almost no region has a 'research performance evaluation' been linked to the funding regime of universities and public research organizations. The models used for university core funding are slowly evolving into 'contract agreement' types, but based on output indicators and not on research assessment.

This brief summary leads us to the exploration of the way in which those practices of research evaluation are affecting various dimensions of the organisational field of universities and similar organisations, and the public research system as a whole. Although the specific arrangements for RES in Spain are in transition, their processes and procedures have started to produce feedback effects in the public research system. For the sake of clarity we can distinguish between direct and indirect effects.

Direct effects are consequences that result from legal regulations, organisational procedures, norms and routines, mainly related to funding mechanisms and the functioning of the labour market for researchers. There are also indirect effects that result from the legitimisation and reputation dynamics that increasing information and transparency across the system are producing. These types of mechanisms produce long term effects in the overall system.

Regarding direct effects, the most important one relates to funding distribution**.** Together with the financial weakness of Spanish universities, the research evaluation procedures embedded in the competitive project funding model have both created a strong dependence of researchers and research groups on external funding sources and differentiated research groups in terms of resources and prestige. Additionally, the limited funds of universities and other research organisations reduces significantly the authority of the managers and academic authorities over researchers and research groups, who are highly autonomous in their decisions and in pursuing their research objectives. Universities and traditional public research organisations are more 'confederations' based on distributive coalitions of individuals, departments, institutes or schools, rather than as unitary strategic actors.

The second relevant direct effect relates to staff recruitment and research career management. Before the 2001 reforms universities retained almost complete control and autonomy in the selection and appointment of teaching and research staff. In parallel, a nationwide individual performance evaluation system (CNEAI) was established, providing some additional personal income and establishing a formal reputation system. A long debate in the 1990s about the negative consequences of inbreeding for the quality control of the new university professors and lecturers created the conditions for the 2001 legal reforms. External evaluation of individual curricula has now become embedded in the new hiring and promotion procedures. The legal and normative changes in the requirements for contracting university pre-tenured and tenured staff have had an impact as regards quality control at the entry point of the academic labour market.

Some of these effects are contradictory to some extent because, on the one hand, the system has advanced into a more extensive and deeper development of the individual researchers evaluation, by means of accreditation and *habilitation* as legal requirements to be contracted or tenured by any university; but on the other hand, due to the proliferation of principals of research (regional authorities) and the available funding sources, scientists find less pressure for resource competition.

Finally another direct effect of the individual evaluation principles was built in the 2001 Universities Act which introduced a provision whereby only professors with positively evaluated research periods by CNEAI (1 or 2, depending on the categories) could be members of the selection committees for the *habilitation* processes. This regulation has meant more than 40% of tenured professors with variations across scientific fields are left out of the selection system.

There are also some indirect effects at the aggregated level that result from increasing information within the research system. In particular, the use of data from individual research performance evaluation (CNEAI) in aggregated form has begun to affect collective reputations. Since academics with rejections in their 'CNEAI *sexenios*' are not well considered by their colleagues, individual information has become almost secret, private information, affecting the individual rights. However, the information has increasingly been aggregated by institutions and scientific fields, and some data about the quality or excellence of universities and CSIC, resulting from the evaluation of their tenured professors and researchers has become available in some publications similar to rankings (MEC 2004).

At the aggregated level the CNEAI system has had some further effects. The use of standard academic criteria, mainly based on international scientific publications, has produced a slow, but clearly shaped change in the publication pattern of Spanish academics. The impact of the CNEAI system has been a significant and continuous growth, in the recent years, of the Spanish share of the ISI-Thomson databases (Jiménez-Contreras et al. 2003). These processes are similar to those described in other countries that use formula funding associated with publications included in the ISI-Thomson database (Butler 2003).

It is also interesting to note that bibliometric indicators, even those not developed in Spain in the context of direct evaluation exercises but more often in the context of bibliometric research, have had a strong indirect impact on the informal 'reputational market' of academic science. In this direction some public agencies

have recently promoted the systematic analysis of the Spanish bibliometric domain, by fields and institutions, to publish rankings (FECYT 2005).

These dynamics are also related to the most significant process in the system: differentiation, which is increasing due to the project-based nature of the research-funding regime. These differentiation effects operate, on the one hand, among research groups by their relative levels of competitive external funding acquisition, and on the other hand, among the individual researchers by the relative recognition of positively evaluated research performance periods.

In contrast to the research group and individual researcher level, we find only a marginal degree of differentiation among the research organisational level, that is, among universities and research centres. The fact that organisational evaluation as it has been designed and performed is not linked to funding decisions is only part of the explanation. Research evaluation, as a tool for strategic planning and management improvement, is advancing very slowly. Although there is a new emphasis on institutional assessment in universities, the national organisational evaluations set into motion during the period studied have not emphasised research activities and, at the sub-national level, regional institutions have followed a similar pattern.

Increasing the differentiation and specialisation of universities and research organisations will require, in addition to the indirect reputational effects, a change in the funding system of research, creating mechanisms to assess the collective research performance and linking it to differentiated funding. Additionally, the emergence of new types of research institutions in the organisational field, as reported in Cruz-Castro and Sanz-Menéndez (2007), could increase the level of competition between organisations favouring the differentiation and specialisation process.

If new funding instruments are developed to shape organisational behaviour, for instance by hiring people with research profiles, we could witness the emergence of more institutionally-oriented research performance evaluation. Alternatively, the development or reinforcement of new models of funding based on contract agreements could help with the development of an institutionally focused, soft type of research evaluation, as usually implied in the strategic plans models. Overall, though, the recent trends have reduced the ability of national R&D authorities to steer the system as a whole, because of the increasing pluralism in terms of funding sources.

REFERENCES

Ben-David, Joseph (1971), *The Scientist's Role in Society,* Englewood Cliffs, N.J.: Prentice-Hall.

Ben-David, Joseph (1972), *American Higher Education*, New York: McGraw-Hill.

Braun, Dietmar (1993), 'Who Governs Intermediary Agencies? Principal-Agent Relations in Research Policy-Making', *Journal of Public Policy*, 13, 135-162.

Butler, Linda (2003), 'Explaining Australia's increased share of ISI publications – the effects of a funding formula based on publications counts', *Research Policy*, 32, 143-155.

Bricall Report (2000), *Informe Universidad 2000,* Madrid: CRUE

Callon, Michel, Philippe Laredo and Philippe Mustar (eds.) (1995), *La Gestion Stratégique de la Recherche et de la Technologie*, Paris: Economica.

Campanario, Jose Manuel (1998a) 'Peer Review for Journals as It Stands Today'. *Science Communication,* 19, 181-211.

Campanario, Jose Manuel (1998b) 'Peer Review for Journals as It Stands Today'. *Science Communication,* 19, 277-306.

Chubin, Daryl E. and Edward J. Hackett (1991), *Peerless Science: Peer review and U.S. Science Policy,* Albany, N.Y.: State University of New York Press.

Cole, Stephen (1998), 'How Does Peer Review Work and Can It be Improved?' *Minerva*, 36, 179-189.

Cole, Stephen and Jonathan R. Cole (1973), *Social stratification in Science*, Chicago: University of Chicago Press.

Cole, Stephen, Leonard Rubin and Jonathan R. Cole (1978), *Peer review in the National Science Foundation: Phase one of a study,* Washington DC: The National Academy of Sciences.

Cruz-Castro, Laura, and Luis Sanz-Menéndez (2006) 'Careers at universities and public research centres: evidence from individual trajectories' data', Paper presented at the Conference on *Science and Technology Policy 2006: US-EU Policies for Research and Innovation*, Atlanta, 18-20 May 2006.

Cruz-Castro, Laura, and Luis Sanz-Menéndez (2007), 'New Legitimation Models and the Transformation of the Public Research Organizational Field', *International Studies of Management and Organization*, 37, 27-52.

Dasgupta, Partha and Paul A. David (1994), 'Towards a new economics of science', *Research Policy*, 23, 487-521.

DiMaggio, Paul J. and Walter W. Powell (1983), 'The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields', *American Sociological Review*, 48, 147-160.

FECYT [Fundación Española para la Ciencia y la Tecnología] (2005), *Indicadores Bibliométricos de la actividad científica española – 2004*, Madrid: FECYT.

Fernández de Labastida, José Manuel (2005), 'The role of institutional evaluation in the CSIC strategic planning and its consequences', Presentation at the *OECD/BMBF International Workshop on the Evaluation of Publicly Funded Research*, Berlin, 26. and 27. September 2005.

García, Clara Eugenia and Luis Sanz-Menéndez (2003), 'The Evolution of Knowledge Management Strategies in PROs: The Role of S&T Policy in Spain', in OECD [Organisation for Economic Cooperation and Development] (ed.) (2003), *Turning Science into Business: Patenting and Licensing at Public Research Organisations,* Paris: OECD, pp. 203-222.

García, Clara Eugenia and Luis Sanz-Menéndez (2005), 'Competition for funding as an indicator of research competitiveness: The Spanish R&D government funding', *Scientometrics*, 64, 271-300.

Georghiou, Luke (1995), 'Research evaluation in European and National science and technology systems', *Research Evaluation*, 5, 3-10.

Geuna, Aldo and Ben R. Martin (2003), 'University Research Evaluation and Funding: An International Comparison', *Minerva*, 41, 277-304.

González Lopez, María José (2006), 'Towards decentralized and goal-oriented models of institutional resource allocation: The Spanish case', *Higher Education*, 52, 589-617.

Hernández Armenteros, Juan (2004), *La Universidad española en cifras 2004,* Madrid: CRUE.

Jiménez-Contreras, Evaristo, Felix de Moya-Anegón and Emilio Delgado López-Cozar (2003), 'The evolution of research activity in Spain: the impact of the National Commission for the Evaluation of Research Activity (CNEAI)', *Research Policy*, 32, 123-142.

Liefner, Ingo (2003), 'Funding, resource allocation, and performance in higher education systems'. *Higher Education*, 46, 469-489.

MEC [Ministerio de Educación y Ciencia] (2004), *Informe del Profesorado funcionario de las universidades públicas españolas y de la actividad investigadora evaluada,* Madrid: Secretaria del Consejo de Coordinación Universitaria, MEC, Junio 2004.

Merton, Robert K. (1957), 'Priorities in scientific discovery: A Chapter in the Sociology of Science', in *American Sociological Review*, 22, 635-659.

Mora, José-Gines (2001), 'The academic profession in Spain: Between the civil service and the market', *Higher Education*, 41, 131-155.

OECD [Organisation for Economic Cooperation and Development] (1964), *Country Report on the Organisation of Scientific Research: Spain*, Paris: OECD.

OECD [Organisation for Economic Cooperation and Development] (2006), *Evaluation of Publicly Funded Research. Recent Trends and Perspectives*, Report prepared by Luke Georghiou and Philippe Laredo [DSTI/STP(2006)7], Paris: OECD.

Pelz, Donald C. and Frank M. Andrews (1966), *Scientists in Organizations. Productive Climates for Research And Development,* New York: John Wiley and Sons.

Sánchez-Ferrer, Leonardo (1997), 'From Bureaucratic Centralisation to Self-Regulation: The Reform of Higher Education in Spain', *West European Politics*, 20, 164-184.

Sanz-Menéndez, Luis (1995), 'Research actors and the State: research evaluation and evaluation of science and technology policies in Spain', *Research Evaluation*, 5, 79-88.

Sanz-Menéndez, Luis (1997), *Estado, ciencia y tecnología en España (1939-1997),* Madrid: Alianza Editorial.

Sanz-Menéndez, Luis and Laura Cruz Castro (2003), 'Coping with environmental pressures: Public Research Organizations responses to funding crisis', *Research Policy*, 32, 1293-1308.

Sanz-Menéndez, Luis and Laura Cruz-Castro (2005), 'Explaining the science and technology policies of regional governments', *Regional Studies*, 39, 939-954.

Schoijet, Maurico and Richard Worthington (1993), 'Globalisation of science and Repression of scientists in Mexico', *Science, Technology and Human Values*, 18, 209-230.

Whitley, Richard (2000), *The intellectual and social organization of the sciences*, Oxford: Oxford University Press, second edition.

Whitley, Richard (2003), 'Competition and pluralism in the public sciences: the impact of institutional frameworks on the organisation of academic science', *Research Policy*, 32, 1015-1029.

Zuckerman, Harriet and Robert K. Merton (1971), 'Patterns of Evaluation in Science. Institutionalisation, Structure and Functions of Referee System'. *Minerva*, 9, 66-100.

CHAPTER 11

SUSAN E. COZZENS

DEATH BY PEER REVIEW?

The Impact of Results-Oriented Management in U.S. Research

The introduction of research evaluation and performance measurement systems for research in most OECD countries over the last two decades has been accompanied by protest and angst on the part of research communities and firm resolve on the part of implementing agencies. Freedom is a wonderful thing to have, and researchers of course prefer good resources and self-direction over external steering and monitoring. At the same time, performance is a wonderful thing to fund, and government agencies of course see it as only rational and responsible to expect the highest quality in what they fund, to set clear goals, and to make sure that they are being achieved. The resulting dialogue has occupied many conference rooms and journal issues.

Has all the effort to design research evaluation systems paid off? When public health officials, in the face of controversy, made the effort to have fluoride put in the water supply, they expected to see a decrease over time in tooth decay - and they did. Have government research authorities seen improvements in the research portfolios of the countries that have adopted the system? On the other hand, have critics seen the decline in creativity that they predicted? Is it too soon to tell? Or is it impossible to tell?

This paper considers these questions in light of the recent history of results-oriented research management at the federal level in the United States. Results-oriented management was introduced in the U.S. in the early 1990s, through the vehicle of the Government Performance and Results Act (GPRA, see U.S. Congress 1993). The principle has since been incorporated widely in legislation (McMurty 2002). The Office of Management and Budget (OMB) has moved the emphasis on performance another step forward with its Program Assessment and Rating Tool (PART)[1], which appears to be a strong evaluation system, in Whitley's terms (see Whitley, this volume): formal, public, and linked to resource allocation. Indeed, many research agencies have strengthened their evaluation systems over the last decade, with the most striking examples coming from federal laboratories in mission agencies.

[1] http://www.whitehouse.gov/omb/part/ (accessed 6 November 2006).

R. Whitley and J. Gläser (eds.), The Changing Governance of the Sciences, 225–242.

While responding to legislative requirements, however, U.S. research agencies have escaped the dangers many feared a decade ago. Quantitative performance goals are rare, and confined largely to facilities and management processes. Qualitative retrospective peer review has been officially accepted as a measure of performance across a broad range of programs. And the basic research agencies have persuaded OMB that well-run project selection processes should be taken as their primary performance indicator, rather than pre-planned discoveries (OMB 2005).

Consider the following quotations from the Office of Management and Budget's R&D Investment Criteria.

> While the criteria are intended to apply to all types of R&D, the Administration is aware that predicting and assessing the outcomes of *basic* research in particular is never easy. Serendipitous results are often the most interesting and ultimately may have the most value. Taking risks and working toward difficult-to-attain goals are important aspects of good research management, and innovation and breakthroughs are among the results. However, there is no inherent conflict between these facts and a call for clearer information about program goals and performance toward achieving those goals. The Administration expects agencies to focus on improving the management of their research programs and adopting effective practices, and not on predicting the unpredictable. …
>
> The intent of the investment criteria is not to drive basic research programs to pursue less risky research that has a greater chance of success. Instead, the Administration will focus on improving the management of basic research programs. (OMB 2005: 60)

Seasoned evaluators will note that 'management' is a process concept. Where, then, are the 'results' in results-oriented management of basic research, in this description? Project selection has trumped outcomes here, and results-oriented management seems to have suffered 'death by peer review'. This is a turn of affairs that many would applaud.

To say that U.S. research has avoided the worst pitfalls of strong evaluation systems, however, is to underestimate the impact of results-oriented management as a whole. In the U.S., this system has had its greatest impact in mission agencies, and in particular in government laboratories. I argue that the impacts of these tools are best seen in terms of their consequences for innovation systems, both sectoral and national (Freeman 1987; Lundvall 1988; Malerba and Orsenigo 1997; Nelson 1993). In the United States, those consequences have been largely positive, for structural reasons that are anticipated well by Whitley's organizational framework.

ANALYTIC FRAMEWORKS

We cannot understand the impact of research evaluation systems in the United States in isolation. Research evaluation systems have been strengthening in the United States in the context of overall systems of results-oriented management, and it is the overall system that creates effects, not just one part of it. The goal of results oriented management is to shift the attention of government program managers from inputs to outcomes, from spending to results. Towards this end, the first major results-oriented legislation, GPRA, requires three documents: a strategic plan, covering a five year period and updated every three years; a performance plan, setting specific

annual target levels of performance; and a performance report, indicating whether the targets were met and if not, why not (U.S. Congress 1993). The strategic plans are organized by programs (most agencies aggregated their programs at fairly high levels for GPRA purposes; see COSEPUP 1999), each of which has objectives and performance goals that contribute to agency goals. Within a research program, goals and objectives are translated into project selection criteria, passing performance expectations on to each project.

Evaluation is the process that aggregates results of various projects and compares them with agency goals and objectives. It asks: are we achieving what we hope to achieve, to fulfil our mission? The answers to this question feed back into strategic planning, and on through the cycle again.

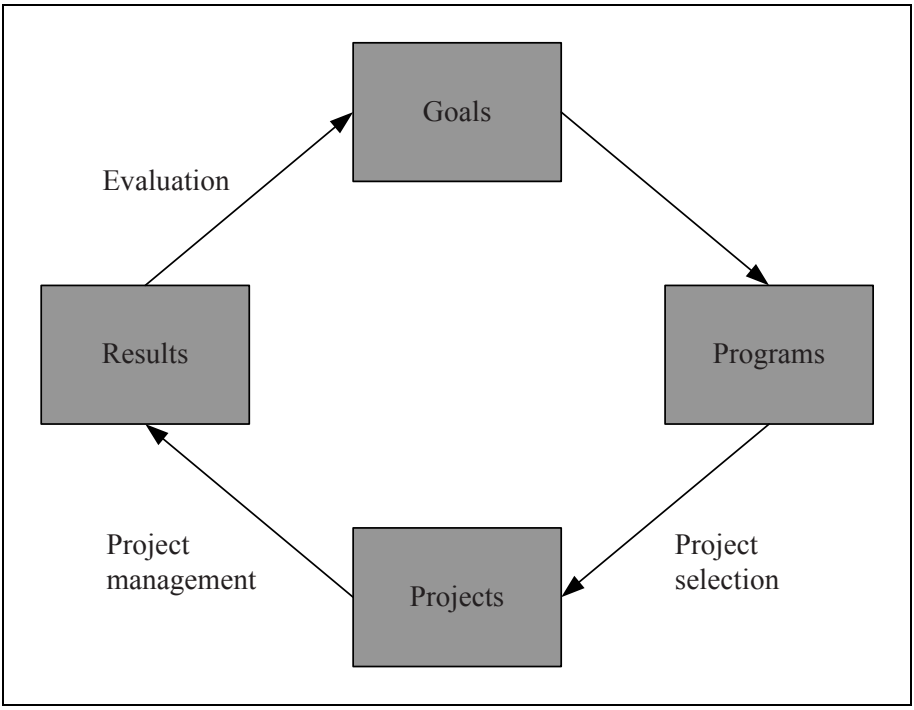

Figure 1. The Cycle of Results-Oriented Management

Whitley's analytic framework (in this volume) attempts to account for the effects of research evaluation systems under different organizational circumstances. The framework is constructed from five types of funding, two levels of strength in research evaluation systems, four characteristics of public research systems, and four characteristics of fields. Together, these variables predict effects in areas that I will summarize as competition, disciplinarity, and risk. In brief, Whitley claims that where disciplinary elites control resource allocation, the introduction of strong evaluation systems will increase competition among both individuals and

institutions, and inhibit risk-taking and interdisciplinary research. Where there is already a high level of use-oriented funding from a variety of funding sources, these negative effects will be 'limited' or 'weakened' in Whitley's scheme.

What is striking about the list of consequences is that they are all unintended, and therefore do not include the changes that government managers seek when they introduce evaluation systems. Two of those goals seem to me to be particularly important to take into account in an overall assessment of the impact of evaluation systems.

The first is trust. GPRA's first goal was increasing the confidence of the American public in government. (The Act has never been held accountable to this objective, to my knowledge.) Are research evaluation systems succeeding in building public confidence and trust? Many U.S., research agencies, particularly those that have taken up research evaluation reluctantly, have also stressed the communication value of stating strategic goals succinctly and using them to organize activities in ways that make the function of the agency more accessible. Have those agencies tested the effectiveness of this new communication pattern? Better annual reports are not the measure; their impact could be.

While an increase in public trust may be hard to measure, trust also figures in specific relationships with key external stakeholder groups for research. In the United States, research-based industry is the most influential, both at a generic national level and in industry-specific innovation systems. Likewise, particularly in health and environment, organized citizen groups are a crucial part of the evaluation context of public research. The trust of these groups in research could be built through their participation at a number of stages of results-oriented management, from strategic planning through ex post evaluation. Has that interaction increased? Is there more trust now than a decade ago?

The goal of building trust might be seen as window dressing in the legislation mandating performance evaluation systems. But it is impossible to brush off improving performance itself as an objective. On the tenth anniversary of GPRA, I interviewed a number of U.S. agencies about its effects in their organizations. Some of them were quite convinced that their effectiveness had improved, and others laughed at the question. The reactions are correlated with some of the variables in Whitley's scheme, and indeed with his characterization of consequences, as well. Where agencies saw their role primarily as stimulating new ideas (the risk-taking and inter-disciplinarity dimensions in Whitley's list of effects) – that is, in basic research - they were not likely to see GPRA as having improved their effectiveness. Where agencies saw their role as adding to a knowledge base to contribute to solving particular problems, however – that is, in mission agencies – they reported that GPRA, and the family of management approaches it represented, was a useful tool, particularly in helping with strategic alignment of activities. These agencies could not point yet to specific impacts attributable to re-aligned activities, but they had a high sense of confidence that their portfolios were now organized in a way that would increase impact. (On GPRA implementation and results, see also COSEPUP 2001 and GAO 2004.)

The agency reactions point to the dual character of performance in modern research systems. Research organizations are usually expected to be sources of creativity and risk-taking, but the research portfolio is at the same time expected to be strategically oriented towards solving particular problems generated either by industry or civil society. The tool in the new management systems that is most directly related to the latter goal is strategic planning. Impact assessment is also quite relevant to the problem-solving goals of research systems. Has methodological attention to this area grown with the growth of the performance management systems? The presence and strength of measures that trace the results of research for society at large would be a good intermediate measure of the success of performance evaluation systems with regard to this set of public goals.

In short, we need to assess the results of evaluation systems in the broader context of learning in innovation systems. An innovation system, as the concept has been articulated in evolutionary economics, consists of actors (usually research institutes, government, and industry) and their relationships (for a review, see Edquist 2006). The better the system, the better all the actors are at utilizing and accumulating knowledge to increase the vitality and growth of the system. At the centre of the innovation system concept is the learning firm, or, if we wish to extend the model to public sector innovation, a learning organization, either public or private. Learning organizations need sources of creativity and new ideas, and they need knowledge directed at solving their particular problems. The relationship between the learning organization and its sources of knowledge is the key to the effectiveness of the system. The learning organization needs relationships with research organizations that are not only creative and interdisciplinary, but also open and responsive. Results-oriented performance management systems may be particularly good at encouraging responsiveness. Perhaps.

THE U.S. PUBLIC RESEARCH SYSTEM

The United States is by far the largest national system among the OECD countries, with $284 billion in overall R&D spending in 2003 (National Science Foundation 2005). The federal government provides most of the public support for research, and performance management has been implemented across the federal agencies. It is therefore fair to evaluate GPRA and its successors in their effects on the system as a whole.

Figure 2 is an attempt – admittedly imperfect – to depict those relationships. To try to capture Whitley's concept of Public Research System, I have placed the sources of public expenditure on research and development in juxtaposition with the institutions other than industry that spend the money. Industrial R&D appears to the right as a market for knowledge from the research institutions, and organized civil

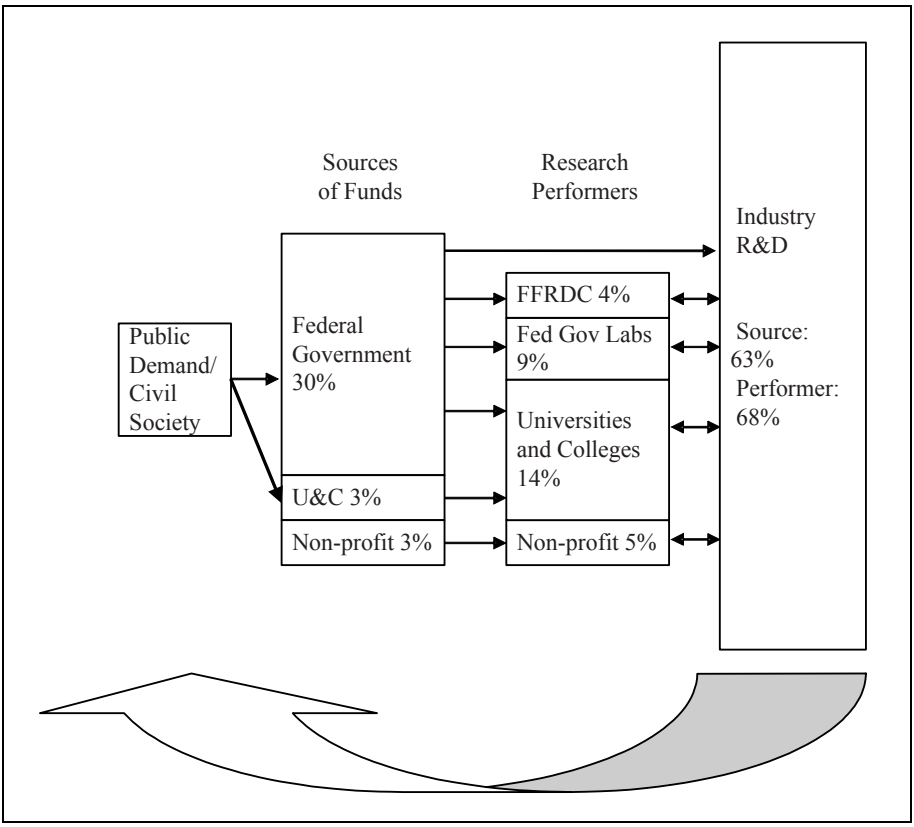

Figure 2. U.S. Public Research System (Source: NSF 2005)

society appears at the left as a shaper of federal spending priorities. The diagram has no way of depicting the strength of the market for funding created by research organizations themselves or by parts of the federal bureaucracy, e.g., the Defense Department's market for weapons. (The other learning organizations should really appear at the right along with industry.) The linking arrow is not a particularly good way to depict the relationship between consumer demand and industrial R&D. Most crucially, being a national diagram, Figure Two does not depict the international context for any of this activity, neither in the form of international competition that industry is facing nor international sources of information that learning organizations can draw on to improve quality of life.

What the imperfect figure does show, however, is that there are two major groups of non-industry research-performing organizations in the U.S., government laboratories and universities.[2] Government laboratories, which are further

[2] While the non-profit sector appears, the number is an estimate and there is little systematic information available on this sector, so I will pass over it in my discussion.

subdivided into intramural labs and federally-funded research and development centers (FFRDCs) rely almost entirely on the federal government for funding. Universities, on the other hand, are quite heterogeneous in their sources of support, with funds from federal and state agencies, along with internal funds, industrial support, and some money from private philanthropy. Like government laboratories, universities fall into two groups: about a third are private universities (which receive a larger share of their R&D funding from federal sources) and two-thirds are public universities that are state, not federal, organizations. The comparison between government laboratories and university research will thus form an important dimension of the analysis for the United States.

FORMS OF FUNDING

What forms of research funding are used in the United States, in what proportions? I find that Whitley's categorization needs added dimensions to begin to describe U.S. funding streams. His distinction between block funding and project funding is quite useful. Core funding for federal laboratories probably falls into the block funding category, although for FFRDCs, the management contract is subject to re-competition periodically (for a review of issues related to the laboratories, see GAO 1998). Federal extramural block funding is uncommon, although not completely unknown. The largest surviving program, in the Department of Agriculture, distributes $450 million annually.[3] The resources universities devote from internal resources might also fall into this category, especially in state universities, where some of that money comes from the state legislature (National Science Foundation 2005).

From an evaluation viewpoint, however, each of these funding allocations incorporates some form of strong performance evaluation system. Both universities and government laboratories have personnel evaluation systems that take research performance into account. The promotion and tenure system in U.S. universities is a strong system, and at research universities the research credentials of candidates for promotion and tenure get very careful examination. In some federal laboratories, including the National Institutes of Health (NIH), external review is used for personnel evaluation (Cozzens 2001). In addition to personnel evaluation processes, all government laboratories are subject to program review, and FFRDCs have to re-compete regularly for their support.

Most U.S. federal research funding, however, is distributed on a project basis through competitive merit review. This observation applies across the major extramural funding programs at NIH and National Science Foundation (NSF) and the smaller extramural programs at the agencies that house laboratories. But it even applies nowadays within government-funded laboratories, many of which run an internal competition for funds on a project basis with some form of merit review, either internal or external (Cozzens 2001). These processes for evaluating intramural projects are being linked in many agencies with the rest of the cycle of results-

[3] http://www.csrees.usda.gov/ (Accessed 6 November 2006).

oriented management: strategic plans to determine relevance, performance plans to set year-by-year goals, and the use of external evaluators to examine both strategic directions and project quality.

It is very hard, however, to sort out the discipline-driven from problem-driven research in the U.S. system (Whitley's 'project' versus 'programme' funding). The National Science Foundation is home to the largest number of truly discipline-based programs, as befits its role as a balance wheel among fields and responsibility for maintaining the health of university research. NSF also prides itself on stimulating emerging interdisciplinary areas, including ones that are not necessarily problem driven. And of course, it considers all the research it supports to be 'basic', in OMB's term. The Office of Basic Energy Sciences at the Department of Energy, home to much of the support for physics, would probably also identify some of its programs as discipline-based, including its support for major high-energy physics facilities. NASA Space Science would be in a similar position.

The National Institutes of Health, however, illustrate the problem in parts of the rest of the U.S. funding system. Biomedical research in the U.S. is very strongly 'use-oriented basic research', falling into Pasteur's Quadrant in the scheme introduced by Don Stokes (Stokes 1997). What are called the pre-clinical disciplines in biomedical research are both disciplinary and problem-oriented. They generate their own research agendas, but because they are so steeped in the context of science-based medicine, those agendas automatically carry a high level of relevance to human health. NIH would probably argue to OMB that most of what it funds is 'basic research' and should fall under the management principles quoted at the beginning of this chapter.

A few of the funding programs of other mission agencies might make similar arguments, but perhaps less convincingly. Most of the funding programs of other mission agencies would clearly self-identify as problem-oriented, and thus not claim the basic research exemptions in the OMB criteria. Examples would include the energy technology programs at the Department of Energy, project-based extramural funding at the Department of Agriculture, and extramural funding from the Office of Naval Research.[4]

TYPES OF RESEARCH EVALUATION SYSTEMS

Whitley groups evaluation systems into two types: weak and strong. Weak systems are informal, private, and formative. Strong systems are formal, public, and summative in the sense of having direct links to the allocation of resources. The strongest type of strong system would be funding based on a publicly-verifiable performance-based formula. The fact is that most resource allocation is done through prospective evaluation rather than retrospective in the U.S., so the chances

[4] Whitley creates a further subcategory of 'delegated' in his 'project' and 'programmatic' categories. In the U.S., centers exemplify this form of funding. They often allocate resources internally to a range of activities. Some government laboratories also might be thought to do this. I know of no interesting evaluation problems associated with the delegation process, however, so I will ignore these subdivisions for the purposes of the analysis in this paper.

of finding what Whitley describes as a strong evaluation system are rather slim. Retrospective evaluation is generally just one input into prospective allocation decisions that rest as much or more on proposed benefits. This is true even with OMB's PART system.

GPRA was widely seen to be a serious effort to introduce stronger evaluation systems across U.S. federal agencies. The ways agencies reacted to and absorbed GPRA's requirements thus provide a number of opportunities to test Whitley's hypotheses about evaluation and organizational dynamics. To do so, we need to understand the baseline situation. When GPRA was passed, I had just completed a survey of program evaluation practices for research in U.S. federal agencies (Cozzens et al 1994, Cozzens, 1997). The processes could be described as varying from weak to non-existent. I found no formula funding based on retrospective evaluation at that time. Most of the mission agencies evaluated through 'program review', a process that usually involved external evaluators, was somewhat structured in terms of the information the evaluators received, but which reported privately to agency management with the goal of program improvement. In the stronger cases, external evaluators rated individual projects on their performance. Nonetheless, these program review processes would fall rather easily into Whitley's 'weak evaluation' category.

The basic research agencies in general did even less. Especially for their core extramural project-based funding programs where peer review was used for project selection, neither NIH nor NSF did any regular program evaluation. The exceptions were the unusual and often the controversial programs, for example at NSF, the centers programs (for example, COSEPUP 1996, NSF 1993, Committee on National Needs for Biomedical and Behavioral Research Personnel 1994). Both agencies evaluated their human resource programs more regularly and professionally than their research programs. Formal program evaluations were published, in part because they were generally done to demonstrate program effectiveness to audiences outside the agency. In all cases, evaluation was just one input into the decision process around program budgets.

The current situation is that the program review processes in the mission agencies have, in general, been strengthened considerably. They are much more closely tied to strategic planning, use external advisors much more often, and are much more closely aligned to resource allocation processes. Both the Environmental Protection Agency and the Agricultural Research Service (ARS) exemplify this trend (Cozzens 2001). Both of these agencies manage government laboratories. In the pre-GPRA era, these laboratories were subject to fairly weak review processes. Under GPRA, both agencies have adopted strong, ambitious strategic goals, and developed extensive new systems for aligning laboratory activities to these goals through an internal project proposal and evaluation process.

Mission Agency Example

In the case of the Agricultural Research Service, the alignment process involves extensive external consultation, strengthening the relationships with both academe

and industry.[5] The ARS strategic plan has five goals, which translate the goals of the Department into research terms.

Through research and education, empower the agricultural system with knowledge that will improve competitiveness in domestic production, processing, and marketing.

- To ensure an adequate food supply and improve detection, surveillance, prevention, and education programs for the American public's health, safety, and well-being.
- A healthy and well-nourished population who has knowledge, desire, and means to make health-promoting choices.
- To enhance the quality of the environment through better understanding of and building on agriculture's and forestry's complex links with soil, water, air, and biotic resources.
- Empower people and communities, through research-based information and education, to address the economic and social challenges of our youth, families, and communities.

Over the last few years, ARS's projects have been grouped into 'national programs' that are aligned with the goals. Working groups of the National Program Staff have done this work, whittling an initial list of about 50 program areas down to the current 22 or so. Examples are 'crop production' and 'integrated crop systems'. These national program areas represent a major cultural change for the organization. The programs set expectations for specific projects. Under this structure, every laboratory activity is linked to the strategic plan goals. The strategic plan forms the criteria for placing projects into programs, and will form the criteria for assessing them in each year's budget submission. Thus, in the part of the process, relevance to Department goals is a dominant consideration. Another feature of the new national program structure is the appointment of 30 national program leaders. Each has an area of special expertise, and is responsible for coordination in that area.

A new Office of Scientific Quality Reviews for ARS was set up the National Program Staff as part of this new structure. Under this system, all the projects in a particular program undergo external review at the same time. This allows panels to see the overall structure of the program, rather than seeing just one project at a time. The first review under this system, of Food Safety, was completed in 2001. The external panels make decisions on individual projects (rating them as outstanding, good, or bad). They judge these units by quality and relevance, plus the capability of the investigators to do the project, and recommend changes as needed. Most of the members of these panels work outside ARS, and most work outside government. All are Ph.D. scientists with excellent research credentials. Some combine research expertise with experience in organizations, like seed firms, that use ARS research results.

ARS feels that this new system has many strengths. In the past, the national program staff organized reviews of each project when it expired, requesting three to five external reviews. Reviewers were not convened, and they were not paid. The

[5] The following description is based on an interview with David Rust, ARS National Program Staff, as reported in Cozzens (2001).

new review system allows for more of an overview of the program, and they feel that convening the panels allows them to understand program balance and generate better advice. The first review panels were very rigorous, and ARS staff were pleased with the input they received.

Between program reviews, managers in the laboratory are responsible for maintaining progress toward project goals. Each area office reviews its own activities annually, and submits a report on accomplishments and impacts. In addition, the 2000 scientists of the ARS are also subject to rigorous individual review processes. Every three to five years, an internal peer review group examines their accomplishments, to determine whether they should be promoted or remain in grade. This process has great credibility in the agency, which puts considerable time and attention into it. The program formation process already described takes projected results and impacts into account, as do the prospective program reviews. A retrospective review process is under consideration. It will use the goals and objectives of the ARS and program area strategic plans as criteria.

Basic Research Agency Changes

What remains largely untouched from the pre-GPRA baseline is evaluation at the basic research agencies, NSF and NIH. Both agencies have come a very long way in absorbing strategic planning into their management styles. Both had had unhappy experiences with strategic plans in the years before GPRA, and have now captured the essence of their missions in the form of strategic plans. Both have systematized the gathering of examples of advances and accomplishments for all their funding activities (a true step towards results-orientation). NIH at first experimented with having an external panel examine these success stories and deliver a judgment on agency performance, and NSF still has such a panel.[6] NSF incorporated performance assessment into its existing Committee of Visitors (COVE) process, and now asks its agency-level GPRA panel to aggregate the judgments of 30-odd annual COVs into a judgment of agency performance. This step could be understood as instituting program review in a similar mode to what was in operation in the mission agencies before GPRA. But neither agency has been pushed yet to produce anything more systematic or data-based by way of results information. Formal program evaluation is no more frequent or widespread than it was before GPRA. Under the current OMB understanding of performance as management rather than results, the peer review processes of these agencies are enough.

In some ways, these patterns of change are predictable under the organizational dynamics Whitley identifies. Where strong discipline-based elites were already well entrenched in the power structure of agencies, they were able to ensconce their own definition of performance in the official implementation of results-oriented management. This left their role in resource allocation processes untouched and kept control where it resided before, while co-opting communication processes like strategic planning. In mission agencies, in contrast, results-oriented management

[6] http://www.nsf.gov/about/performance/acgpa/index.jsp (accessed 7 November 2006).

strengthened problem orientation. However, at the same time, at least in some cases (see ARS), the more public character of the process led to more 'external' review by the research community, which in turn increased its role in translating public goals into research programs. This result would not have been predictable in Whitley's scheme.

CONSEQUENCES

Whitley seems to assume that weak evaluation systems do not generate the negative side effects he is interested in, and he therefore pays little attention to them. But the consequences of evaluation for trust and for strategic direction do not depend on the implementation of formal, public, evaluations linked to resource allocation. Clarification of goals and stimulating interaction with stakeholders through a more thorough review process can increase attention to performance, even if the results are not linked through a formula to resource allocation. The argument underlying Whitley's detailed analysis is that strong research evaluation systems under many circumstances intensify reputational competition and the need to coordinate one's research plans with those of colleague-competitors, solidify disciplinary boundaries, and discourage risk-taking. It seems to me that the strengthening of research evaluation systems in U.S. federal agencies has had quite different effects in many cases.

The Benefits of Competition

One of Whitley's hypotheses is that moving from block grants to any form of performance-based resource allocation will increase the sense of competition among researchers. This is a fairly obvious observation, and it is hard to picture a situation in which it would not be true. The important question is what other consequences might accompany this shift, and thus whether the benefits outweigh the costs.

I have recently finished contributing to a study of evaluation processes in a network that has enjoyed block funding for state programs for forty years (OSB 2006). This network of programs includes both research and outreach, and is the heart of federal funding for an innovation system in a key U.S. export industry. About six years ago, a program review process was introduced, including requirements for strategic planning. In the last several years, the program review system for the state programs has been formalized and linked to a small pool of 'merit funding' (less than 10% of the total funding allocated under the program; all base funding has been preserved). This small step towards performance based budgeting has been intensely controversial and painful for the network, which had previously seen itself as collaborative and now feels very competitive. Trust between the national program staff and the state programs has been undermined in part because the national program staff makes the final decision on program ratings.

Still, there is a general agreement across the network that (1) the external reputation of the overall program has been improved because it has publicly undertaken serious evaluation, and (2) performance in many of the state programs

has improved as a result of the implementation of the evaluation system. An outside observer would say that the strength of the evaluation system has indeed gotten the attention of the state programs, and that the innovation system is strengthening as a result. The increased sense of competition is at least for the moment probably a small price to pay for the improved performance of weak units. In addition, the improved reputation of the program is seen to have increased its chances for continued funding. Because strategic planning is hard to implement in such a distributed network and is still in its formative stages in the funding organization, increased intellectual collaboration, as predicted by Whitley's scheme, has not yet been achieved. But its implementation should be enforced more effectively in the competitive situation than under a block funding scheme.

Interdisciplinarity

Whitley's analysis asserts that strong evaluation systems have the potential to solidify disciplinary boundaries and inhibit interdisciplinary research. The movement towards strong systems in the U.S. has certainly not had this effect, and perhaps Whitley's variables predict this result, given the structure of the system. First, as we have seen, very little federal funding is discipline-based in the U.S. The articulation of public benefits and goals through strategic plans has if anything made it less likely that agencies will feature discipline-based programs in their budget justifications, although as we have seen, OMB has left room for disciplinary goals rather than problem solving to set the directions for funding programs.

The bastion of disciplines in U.S. research is university departments, but the growing interdisciplinarity of federal funding does not threaten their hold on hiring and promotion, and thus on careers (Committee on Facilitating Interdisciplinary Research 2004). A plurality of funding sources allows academics in many – although not all – fields to seek support in several places, and they can often publish theory while producing problem-oriented results for their sponsors. The introduction of strong evaluation systems for federal agencies has had very little effect on these possibilities, and thus on the structure of fields.

The main result of the new management styles for most academic researchers is that they need to do more serious results reporting. The agencies that support them are probably asking for more detailed final reports, with more counting of outputs and more active identification of impacts.[7] The new reporting formats may make researchers more aware of the goals their agencies are trying to achieve, and are thus likely to improve actual performance. If the researchers also participate in formulating agency strategic plans (many do through advisory boards), that effect should be increased. Larger grants will be subject to even greater scrutiny for results.

[7] Many of them may be at the same time be providing online reporting systems that make the task much easier than it was earlier, for example, NSF's system, FastLane.

Short-termism

Whitley is not alone in his prediction that strong research evaluation systems will under most circumstances (other than block funding) decrease risk taking among researchers and focus their attention on short-term results. This concern has been voiced repeatedly in the U.S. in the discussion around strengthening evaluation systems (see for example COSEPUP 1999). The quotation from OMB in the introduction to this paper is just one example, and illustrates an interesting phenomenon: in part because of the concern over possible ill-effects of evaluation systems, risk taking has been institutionalized as a value of the U.S. research system precisely during the time when research evaluation systems have been strengthening. The systems have incorporated risk-taking into the concept of performance and provided new opportunities to make risk-taking an official goal. They have thus encouraged rather than discouraged it, at least at the level of rhetoric.

An example that tries to avoid the aversion of formal procedures to uncertain outcomes and deviant approaches is NIH's new program of Pioneer Awards – mini-block grants given to a small set of highly creative researchers to explore their best ideas in an unfettered way (see the NIH Director's Pioneer Awards[8]). Described as "a high-risk research initiative of Research Teams of the Future," the program calls for only a short proposal that is evaluated on its creativity, taking the track record of the investigator in path-breaking research into account. NSF has also turned to "discovery science" as a theme and called for "breakthrough" research in its new initiatives, even in the social sciences (for example, the National Science Foundation's Human and Social Dynamics Program[9]). Just as it was hard earlier to judge whether the actual levels of risk-taking in the system were high or low, it is difficult to judge the actual effects of these steps, but at the least having such official visibility and approval should reinforce the voices of reviewers who want to recognize a spark of originality in a proposal under review, even in the regular programs of these agencies.

Strategic Alignment

Whitley's final concern about strong evaluation systems is that they encourage research management and competition among institutions for reputation. The competition for project funding gives the upper hand to disciplinary orthodoxies, in his view, and makes really creative new programs less likely. The U.S. experience bears out the first prediction, but not the second. In both universities and government laboratories, results-oriented management in funding sources has encouraged more entrepreneurial strategic management, focused on taking advantage of the opportunities provided by the various federal funding sources. As funding agencies have clarified their goals and translated them more effectively into programs, research performing agencies have re-organized to respond to the new

[8] http://nihroadmap.nih.gov/pioneer/ (accessed 6 November 2006).
[9] http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=11678&org=NSF (accessed 6 November 2006).

opportunities. Universities like Arizona State are re-organizing themselves into interdisciplinary schools that align with problems in the external world. FFRDs like Oak Ridge National Laboratory are providing incentives to their researchers to seek new partnerships with universities and private firms to extend the range of their funding possibilities. Such research organizations have risen to the challenge of articulating how their efforts will have an impact on the outcomes funding sources want.

The strategic alignment that Whitley predicts is visible in many contexts. I have already described the processes of strategic alignment taking place within the Agricultural Research Service (ARS), through program re-grouping (Cozzens 2001). In the new environment, laboratory leadership must be aware of context and manage internal resources skillfully to take advantage of competitive openings. The new structure of national programs is designed to communicate the benefits of ARS research to the public. By bringing projects together under themes that are clearly tied to the strategic plan, ARS hopes to make its usefulness evident. Its home page has been organized to give public access to this information. An annual report for each national program area appears, with links to a map that shows the projects associated with each. In addition, ARS holds occasional workshops to consult with customers and stakeholders in specific program areas, developing action plans that are also displayed on the Web. An important communication benefit of the new structure has been with the scientific staff. According to agency officials, staff used to wonder whether anyone was paying attention. Now they know that someone does. ARS researchers are thinking more about their impacts now, and are aware of the need to solve problems that customers and stakeholders want solved. In addition, according the agency's budget staff, the questions they are receiving from Congressional staff have changed, to focus more on outcomes for the public.

The Naval Research Laboratory (NRL) provides another example (Cozzens 2001). The laboratory itself has developed internal competitive management processes that allows much more flexibility in responding to client needs than a block funding system would, in no small measure because people who do not want to respond flexibly can be fired there (a situation that is unusual among government employers). Each research team must garner its own portfolio of resources, mostly external, to show its worth to laboratory clients; otherwise, it is subject to closure. Even at universities, the federal environment calls forth strategic investments. Internal resource allocations tend to be made to increase the potential of the campus to compete for external funds, preferably from a combination of state, federal, and industry sources, as for example when universities provide matching funds for major centre or training grants.

Open Windows

However, research management in this sense has not eliminated creativity from the U.S. research system. The value of new directions has been institutionalized, and there are many open windows in the U.S. system where they can be supported. For example, in the NRL context, the core grant from the Office of Naval Research

provides the opportunity to develop such directions. This core support is allocated through competing proposals that are ranked by a team of senior managers according to the proposal's ability to build core competencies that will extend the laboratory's client-funded work. Core competencies that show promise of attracting external clients get first priority, but the activities undertaken with core support focus on learning and on new skills, rather than producing products for paying clients. At NSF, the bulk of resources are still allocated competitively to investigator-initiated grants (importantly, not contracts), and even science and technology centers and large grants for interdisciplinary graduate training programs are funded across a broad range of fields, with an emphasis on original combinations and breaking new ground (for example, in the Science and Technology Centers[10] and Integrative Graduate Education and Research Training[11] grants). U.S. researchers compete on their new ideas, not just on old ones.

CONCLUSIONS

Results-oriented management thus serves the U.S. innovation system in two ways, through the combination of open windows for creativity and interactively aligned problem-oriented work. The overall picture of changes in the U.S. research community is (1) increased interaction with stakeholder groups around strategic plans, (2) increased interdisciplinarity, (3) increased encouragement of new ideas and risk-taking, and (4) better strategic alignment in all research organizations. Some of those who have lived through the changes would certainly ask whether the cost has been worth the benefit. The objection makes the most sense in the basic research agencies, which have put in no less effort than the mission agencies into implementing the relevant law but made the smallest changes in structure and procedures. In mission-oriented agencies, the changes have been deeper and are judged to be more significant.

The pluralism of the U.S. research system has long been considered its strength, and this analysis of the experience with results-oriented management bears out that judgment. U.S. research organizations have long enjoyed considerable strategic autonomy, especially universities because of their independence from the federal government. The movement towards performance management over the last decade has encouraged the spread of project-based, merit reviewed research portfolios, not only in the university sector but also in government laboratories. Retrospective review processes have been strengthened, creating better quality assurance across the board. And strategic planning has stimulated the explicit articulation of a full range of goals and objectives necessary to a healthy innovation system, from risk-taking and new directions to strategic alignment and outcome orientation. Sectoral innovation systems (such as agriculture) have been strengthened through the alignment and improvement of work in government laboratories, while basic

[10] http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=5541&org=NSF (accessed 6 November 2006).
[11] http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=12759 (accessed 6 November 2006).

research agencies have been given the flexibility they need to produce the unexpected. Added up, this is a recipe for a better innovation system.

REFERENCES

Committee on Facilitating Interdisciplinary Research and Committee on Science, Engineering and Public Policy (2004), *Facilitating Interdisciplinary Research,* Washington, D.C.: National Academies Press.

Committee on National Needs for Biomedical and Behavioral Research Personnel, Studies and Surveys Unit, Office of Scientific and Engineering Personnel, National Research Council (1994) *Meeting the Nation's Needs for Biomedical and Behavioral Scientists,* Washington, D.C.: National Academies Press.

COSEPUP [Committee on Science, Engineering, and Public Policy] (1996), *An Assessment of the National Science Foundation's Science and Technology Centers Program*, Washington, D.C.: National Academies Press.

COSEPUP [Committee on Science, Engineering, and Public Policy] (1999), *Evaluating Federal Research Programs: Research and the Government Performance and Results Act*, Washington, D.C.: National Academies Press.

COSEPUP [Committee on Science, Engineering, and Public Policy] (2001) *Implementing the Government Performance and Results Act: A Status Report*, Washington, D.C.: National Academies Press.

Cozzens, Susan E. (1997), 'The Knowledge Pool: Measurement Challenges in Evaluating Fundamental Research Programs', *Evaluation and Program Planning*, 20, 77-89.

Cozzens, Susan E., Steven Popper, James Bonomo, Kei Koizumi and Ann Flanagan (1994), *Methods for Evaluating Fundamental Science*, Washington, D. C.: RAND, Critical Technologies Institute.

Cozzens, Susan, with Barry Bozeman and Edward A. Brown (2001), *Measuring and Ensuring Excellence in Government Laboratories: Practices in the United States*, Report prepared for the Canadian Council of Science and Technology Advisors, http://www.csta-cest.ca/files/States.pdf (accessed on 6 November 2006).

Edquist, Charles (2006), 'Systems of Innovation: Perspectives and Challenges', in Jan Fagerberg, David C. Mowery and Richard R. Nelson (eds.), *The Oxford Handbook of Innovation,* Oxford: Oxford University Press, pp. 181-208.

Freeman, C. (1987), *Technology policy and economic performance: Lessons from Japan*, London: Pinter Publishers.

GAO [General Accounting Office] (2004), *Results-Oriented Government: GPRA Has Established a Solid Foundation for Achieving Greater Results*, GAO 04-594T, Washington, D.C.: GAO.

GAO [General Accounting Office] (1998), Department of Energy: Uncertain Progress in Implementing National Laboratory Reforms, RCED 98-197. Washington, D.C.: GAO.

Lundvall, B.A. (1988), 'Innovation as an interactive process: From user-producer interaction to the National Innovation Systems', in G. Dosi, C. Freeman, R.R. Nelson, G. Silverberg and L. Soete (eds.), *Technology and economic theory,* London: Pinter Publishers, pp. 349-369.

Malerba, F. and L. Orsenigo (1997), 'Technological regimes and sectoral patterns of innovative activities' in *Industrial and Corporate Change*, 2, 45-74.

McMurty, Virginia A. (2002), *Government Performance and Results Act: Overview of Associated Provisions in the 106th Congress*, Washington, D.C.: Congressional Research Service.

NSF [National Science Foundation] (1993), *Guide to Programs in the Division of Research, Evaluation and Dissemination with specific guidelines for individual programs,* NSF 93-143. Arlington, VA: NSF, http://www.nsf.gov/pubs/stis1993/nsf93143/nsf93143.txt (accessed 6 November 2006).

NSF [National Science Foundation, Division of Science Resources Statistics] (2005), *National Patterns of Research and Development Resources: 2003*, NSF 05-308, Brandon Shackelford: Arlington, VA.

Nelson, R.R. (1993), National Innovation Systems: A Comparative Analysis, Oxford: Oxford University Press.

OSB [Ocean Studies Board, National Research Council, Committee on the Evaluation of the Sea Grant Program Review Process] (2006), *Evaluation of the Sea Grant Program Review Process*, Washington, D.C.: National Academies Press.

OMB [Office of Management and Budget] (2005), *Guidance for Completing the Program Assessment Rating Tool' (PART)*, Washington, D.C.: OMB, http://www.whitehouse.gov/omb/part/fy2005/2005_guidance.doc (accessed 6 November 2006).

Stokes, Donald E. (1997), *Pasteur's quadrant: basic science and technological innovation*, Washington, D.C.: Brookings Institution Press.

U.S. Congress (1993), *Government Performance and Results Act of 1993, 103rd Congress*, S20/HR826, Washington, DC: Government Printing Office.

Concluding Reflections

CHAPTER 12

JOCHEN GLÄSER

THE SOCIAL ORDERS OF RESEARCH EVALUATION SYSTEMS[1]

A SOCIOLOGICAL APPROACH TO RESEARCH EVALUATION

Research evaluation systems (RES) are governance tools that are intended to change science by improving its quality, and which cause concern about change that will alter its content. The contributions to this volume demonstrate that RES have the potential to intrude deeply into the conditions, social relationships and practices of knowledge production. Some RES influence a crucial condition of knowledge production, namely the access to resources. Some of them also change relations within scientific communities by both opening up and demanding new ways of participation by academics in science policy making, thereby shifting the balance of power between those who evaluate and those who are evaluated. RES also contribute to the increasingly powerful discourse on 'good research', which transfers models of 'good research' across organisations and disciplines. Knowing the mechanisms that produce these effects of RES enhances our understanding of the ways in which scientific knowledge is produced at individual and collective levels. This is why the sociology of science should be interested in RES. It is unlikely that science policy and higher education studies can fully grasp the impact of the new political instruments on knowledge production without an input from, and their collaboration with, the sociology of science.

The contributions to this volume also demonstrate the variation of RES and their institutionalisation processes. RES emerge from different initial situations and become institutionalised at different levels – at the state level in Germany and Spain, at the national level in Spain (for individual academics), the Netherlands, Japan, Sweden, and Australia. The institutionalisations start at different points in time and proceed with varying speeds, which is why we have well established RES in the UK and Australia but can only speak of an advent of RES in many other countries.

[1] I would like to thank Kate Barker, Grit Laudel, Uwe Schimank and Richard Whitley for helpful comments on the first draft of this chapter.

R. Whitley and J. Gläser (eds.), The Changing Governance of the Sciences, 245–266.

Our knowledge about the development and effects of RES has developed rather unevenly. While the science policy processes and organisational dynamics have been investigated for several countries, we have just begun to explore how RES interact with scientific communities and their research. This volume sets the stage for sociological research on RES by outlining what is, and what should be, known about this major institutional innovation in the science policies of many countries. The contributions propose what is sociologically important about these RES and sketch the ways in which these phenomena could be investigated.

In this chapter, I want to focus on the links between science policy and knowledge production by interpreting RES as changes in the social orders of knowledge production. This perspective overcomes a characteristic limitation of higher education studies, whose focus on organisations and governance excludes the primary social context in which scientific knowledge is produced, namely the scientific community. Disregarding scientific communities leads to a neglect of the production of scientific knowledge and the social structures that govern this production. Research is reduced to professional work by academics in their universities, and the interactions of academics with universities are conceptually reduced to the aspect of governance. A classic example of this approach is Clark's influential book on the 'higher education system', in which he first acknowledges the primacy of the social structures of knowledge production ('disciplines', Clark 1983: 28-34) but later reduces their influence to the participation of the 'academic oligarchy' in the management of higher education. His 'triangle' of state, market, and academic oligarchy (ibid.: 137-145), the literature building on this idea (see e.g. the list provided in Enders 2002: 75), and more elaborate alternatives such as the system of five governance modes proposed by Schimank (2005) have in common that they effectively decouple the governance problem from the endogenous dynamics of the production processes that is the target of governance.

While these approaches are sufficient to analyse some aspects of governance, e.g. the interactions of universities and the state, they are blind to an important part of the work that is conducted in universities, namely the production of scientific knowledge. In order to fully understand the dynamics of the modern university, we need 'to bring work back in' (Barley and Kunda 2001), which means bringing scientific communities in. Scientific communities have their own distinct social order, which extends across all science policy institutions and organisations. Their impact on academics often is much stronger than that of governance practices within universities. A perspective on governance as changing social orders enables the investigation of interactions between institutions, organisations, and science in a coherent framework at the same level of abstraction.

THE PROBLEM OF GOVERNING SCIENCE

A focus on science policies in general, and RES in particular, supports a picture of the scientific enterprise as being conducted by well organised and clearly delineated national units that are self-governed to a certain extent but answer also to science policy-makers and organisational hierarchies. A first contribution of the sociology of

science to the analysis of the politics of RES is the insistence that this picture is itself a construction. It has been jointly constructed by science policy-makers and scientific elites for governance purposes, and RES contribute to shaping the sciences as such units by defining and delineating them for the purpose of evaluations. The contributions by Whitley, Schiene and Schimank, van der Meulen, Lange, and Kneller in this volume make clear that defining fields, research programmes or 'critical masses' and investing them with resources makes the boundaries of these entities effective and thereby 'real'. A similar point can be made with regard to the scientific elite. By inviting the elite to evaluate research, and by limiting eligibility for some of the funding to the elite, science policy-makers construct it as a visible group that is separated from its community by special treatment and is given new powers within their fields. As Whitley has argued in his introductory chapter, neither effect is straightforward or deterministic. The actual structuring effects of RES depend on the practices and social structures of knowledge production on which they are projected.

To assess the impacts of RES on science it might be useful not to begin with the politically constructed fields but to recall the ideal-typical international scientific community. Scientific communities are actor constellations whose members autonomously decide what problems to solve, how to solve them, and how and to whom to offer results. The social order of this strongly decentralised collective enterprise is achieved by the reference of all members to a shared body of knowledge, which needs to be publicly accessible. Although each producer's perception of this shared body of knowledge is unique, the fact that it is common to all of them enables the amalgamation of contributions. Membership of such a community is constituted by individual perception and actions informed by that perception rather than rules or ascription. A membership resting on self-perception can remain hidden, and can easily change. Nobody knows all members of a community, and episodic, partial, and simultaneous memberships of several communities make scientific communities informal, fluid and thus ill-delineated social collectivities (Gläser 2006).

Since the social order of the scientific communities rests on autonomous local decision-making informed by observation (of knowledge), it is a spontaneous social order in the sense of Hayek's use of the term (Hayek 1991). 'Spontaneous' means that the social order is an emergent effect of unilateral adjustment by community members rather than the result of purposive coordination. This mode of production is superior under conditions of multiple uncertainties that frequently occur in the production of scientific knowledge. In many situations of knowledge production the following points may be unknown:

- What exactly the problem is that needs solving (how it should be formulated);
- Whether there is a solution to the problem at the current stage of knowledge;
- How the problem could be solved;
- What knowledge can be regarded valid and reliable and should therefore be used for solving the problem; and
- Who can solve the problem.

In these situations decentralised autonomous decision-making is efficient because it means that many independent attempts to formulate and solve problems as possible

are undertaken simultaneously. While many (and sometimes most) of the attempts are bound to fail or to become redundant, the decentralised approach provides the highest likelihood that the problem is solved as quickly as possible. A second advantage is that tasks are assigned to producers by 'self-identification' (Benkler 2002: 414-415, Gläser 2007). The individual producers, who know their abilities best and will formulate problems in a way in which they can solve them, decide themselves what to do. This decentralised decision-making on tasks guarantee the best possible fit of tasks and producer.[2]

The spontaneous social order of scientific communities not only facilitates rapid progress under conditions of uncertainty but also makes scientific communities highly adaptable to changes in knowledge or societal conditions of their work. However, it also implies that the scientific community has no structures or processes that enable collective decision-making. Nor can it reach decisions that bind all its members. As a result, the progress of a scientific community's knowledge production resembles a 'drift' rather than a directed movement. This internal 'governance deficit' of the scientific community, i.e. its inability to coordinate large groups of members and to focus their efforts on achieving an aim in a set amount of time, constitutes a specific 'community failure'. This varies between scientific communities[3] and is not necessarily a failure in the context of knowledge production. To compensate for their internal governance deficit, scientific communities utilise other social orders. Internal decision-making, communication and collaboration are supported by a variety of formal organisations such as professional organisations, journals, conferences, and institutionalised peer review procedures. Communication is enabled by a symbiosis of scientific communities with the social order of the market (for journals and books).

The informal and fluid social morphology of scientific communities also poses a problem for their interaction with societies. Linking themselves to societies is an existential need for scientific communities because they cannot create the resources needed for knowledge production (except for the knowledge itself). The resources for research must be provided by (national) societies, which do so because of the scientific communities' contribution to intellectual and material welfare. This exchange relation – the social contract between science and society - is the basis on which modern sciences exist.

In order to be maintained by national societies and to contribute to their welfare, scientific communities need interfaces that link them to the alien social orders of their 'host societies'. Since these societies are 'organisation societies' (Perrow 2002), the interfaces are mostly formal organisations. Research organisations (including universities) host segments of scientific communities and provide them with

[2] It is also liable to individual misperceptions (definition of problems that cannot be solved or are irrelevant to the scientific community) and multiple solutions to the same problem. However, both 'inefficiencies' can be established only after a problem has been solved, i.e. it is impossible to tell *ex ante* which attempts of solving a problem will be wasted.

[3] The relationship between varying degrees of uncertainty and variations in the coordination of knowledge production has been addressed by Whitley (2000). An extreme example of large-scale coordination in scientific communities – the coordination of large experiments in high energy physics – has been described by Knorr-Cetina (1995).

the resources for their research, while intermediate organisations contribute to resource distribution (as funding agencies) and to the decision-making about research and research policy. Societal expectations concerning research, resources provided for research, and contributions of research to societal welfare are channelled through these organisations and the institutions that govern them, which together form the national academic system (Whitley 2003). Shifting demands and contributions, and the need to improve the governance of science, trigger innovations in national academic systems, one of which is RES.

THE PLACE OF RES IN THE GOVERNANCE BY SOCIAL ORDERS

Three Rationales for RES

The traditional arrangement of research funding in many countries has been that research organisations received public block funding for basic supplies for research, which they administered according to their internal needs or on the basis of regulations issued by the state. The amount of block funding depended on perceived needs, which were estimated by the state or negotiated between the state and the organisations. Researchers could apply for competitive grants to further support their research. As the contributions on Sweden, Germany, Japan, The Netherlands, and Spain demonstrate, this arrangement still dominates the funding of university research in these countries. But in many countries the traditional arrangement is now rapidly changing. Research organisations are expected to demonstrate performance rather than need. This is the basis on which they begin to compete directly or indirectly for their block funding. The contributions to this book identify three reasons for these changes. The first reason is that, since the Second World War, science has been immensely successful in making a case for its funding. In a short time, science has achieved indispensability for all advanced industrial societies, and has rapidly grown. By the same token, it has turned its funding into a significant budget position that attracts closer scrutiny and is coming under increasing pressure of justification. The auto-catalytic growth of scientific knowledge production can only be sustained by exponentially growing science budgets, which had become impossible by the 1980s. Since then, science has come to face a 'steady state' of slow or zero effective growth (Cozzens et al. 1990), and societies have become accustomed to the idea that it might be necessary and possible to economise on science (Whitley, this volume). This development has confronted policy in general and science policy in particular with a new question: How to economise on science, which is essential for societal welfare but whose internal workings are opaque to the outsider?

A development that is closely related to the strain on public budgets is the move from blind faith in science to trust based on scrutiny (Cozzens, this volume). While trust in science is still essential because of the esoteric nature of its conduct, science policy demands that science provides reasons why it should be trusted. The expectations include demonstrated performance and demonstrated responsible procedure

(including economic use of resources) *in terms that can be understood outside science*.

Finally, science has to put up with the ramifications of the dominant political paradigm of public policy, namely the idea of 'New Public Management' (Lange, this volume, Schimank 2005). 'New Public Management' is the crystallization in public policy and administration of one of the most powerful ideologies of our time, namely the assumption that market competition and market exchange are the best ways of allocating resources to tasks and thus of conducting tasks regardless of the content of tasks and of the conditions under which they are conducted. A corollary belief is that organisations should best be run like businesses – again regardless of the content of their tasks and of the conditions under which these tasks are to be performed.

These three concerns have triggered attempts by science policy to get a tighter grip on the research it funds.[4] The major obstacle to the design of new governance regimes ensuring the quality of research is the *laissez faire* approach to quality that is characteristic of scientific communities. The decentralised decision-making in scientific communities encourages a wide range of production and sorts out the high quality contributions after peer review and publication. Quality is ultimately established by use in subsequent knowledge production processes, i.e. *ex post* and in the same anarchic process in which contributions are produced. This practice corresponds to the uncertainties of research, which make it difficult to immediately distinguish between marginal, radically new, and bad contributions. This is why allowing them all to be produced and sorting out those that are useful afterwards appears to be an effective if not efficient approach. The scarcity of resources results in the *laissez faire* approach being curbed by the prospective peer review of grant proposals, which channels resources to the best projects.[5] This system is considered sufficient in the US (Cozzens, this volume) but not (anymore) in countries with largely state-financed university systems. The science policy-makers of these countries feel the need to introduce 'accountability' and 'quality assurance' in the second major stream of research funding, i.e. in the research funding of organisations, among them universities. In this process, the *laissez faire* approach to 'quality control' in scientific communities is confronted by the focused approaches of the other social orders that are utilised by science policy.

In his introductory chapter Whitley distinguishes between weak RES that are not public, not transparent, and have little or no consequences for funding, and strong RES that are transparent, public, and have significant consequences for funding. The case studies in this volume enable a further differentiation. They demonstrate that there are public and transparent RES with little or no consequences for funding (e.g. the 'third episode' of research evaluation in the Netherlands). We also observe RES that have strong consequences for universities, which are however structural rather

[4] A fourth trend is the increasing demand for contributions by science to societal welfare, which I will not discuss here because it is addressed by governance instruments other than RES, e.g. targeted funding.

[5] A certain degree of pluralism can be maintained if there is a sufficient variety of funding sources, peer review procedures, and assessors. The descriptions by Cozzens of the pluralistic grant system in the US and by Gläser and Laudel of the monotheistic system in Australia provide an interesting comparison.

than financial (e.g. the peer review system in Lower Saxony). These RES are not necessarily public or transparent. Owing to the independent variations in several dimensions it seems useful to look at RES by applying two dimensions, namely the information provided by RES and the strengths of the consequences (table 1). This leads to a further subdivision of Whitley's basic distinction into four types of RES.

Table 1. Differentiation of RES according to information about and consequences for universities

		Consequences for universities	
		Weak	*Strong*
Information about	*Necessary improvements (local, nontransparent information)*	*Intrusive RES*	
		Information about research performance	*Hierarchical intervention*
	Comparative position (public, transparent information)	Competitive RES	
		Competition for reputation and students	Competition for funding, reputation, and students

The assumption underlying this table, which is supported by the contributions to this volume, is that the kind of information produced by an RES and the way in which it is communicated create two kinds of RES, which can be either weak or strong. To begin with competitive RES, it is important to realise that any public and transparent information about research performance of universities is *comparative* information that inevitably creates a competitive situation. Universities that are publicly compared to each other compete for reputation because they depend on public opinion and on academics' choices of employers. Since comparisons of research quality also inform student choices, universities that depend on these choices compete for students and research students. If the comparison informs funding decisions, they additionally compete for that funding, which constitutes the strongest consequence for them. These arenas of competition and the dependence of universities on the various competitions create a spectrum from weak to strong competitive RES.

Some RES create an entirely different kind of information, namely in-depth information about strengths and weaknesses of one university's research conditions, performance, or potential. The character of this in-depth information means that it cannot easily be compared and only makes sense in the context of the individual university. This information is therefore often communicated in privileged communication between the government and the university. Since it contains opinions about specific strengths and weaknesses of the university's research, it is necessarily intrusive. The strength of the intrusion depends on whether it is accompanied by

recommendations to change research or research conditions, and whether the government is in a position to enforce these changes. The weakest intrusive RES would merely inform universities about their research performance, while the strongest would include changes to the university's research that can be enforced.

The strong intrusive and competitive RES resemble the social orders of hierarchies and markets. Therefore it seems worthwhile to emphasise this distinction by discussing the empirical cases presented in this volume along the distinction between (weak and strong) intrusive and (weak and strong) competitive RES.

Intrusive RES

Five country studies in this volume analyse university systems that grant their universities only limited (albeit increasing) autonomy. According to the logic described in the previous section, a state interest in research performance has begun to enter state-university relationships. Manifestations of this interest go back almost forty years in the Netherlands, are more recent in Germany and Japan, and only exist in the first traces in Sweden. Spain is a notable exception because a state interest in the research performance of *universities* is only now developing as part of the regionalisation of university governance, while there is a unique well established interest of the state in *individual* research performance.

Four of the five countries have intrusive RES that gather information about research performance by means of peer reviews or universities' self-assessments, and use this information for the structural improvement of research. The peer-review based RES in Lower Saxony counteracts the trend towards more autonomy for German universities because it constitutes an intervention into internal university matters that has not existed before.[6] Strangely enough this intervention also strengthens the position of the university vis-à-vis its professors because it entails assessments and decisions that are in the interest of the universities but could not have been made by them against the resistance of the university professoriate. The same argument can be made for the 'performance agreements' between the German federal states and each of their universities. The performance agreements are based on self-assessments and (e.g. in Lower Saxony and Northrhine-Westfalia) on prior evaluations by peer-review. They also constitute a new degree of intervention into university matters but are nevertheless welcomed by universities because they legitimise structural change that would be difficult to achieve by internal decision-making. The reason for this paradox is the double weakness of German universities, which have little autonomy in their relationships with the state and little authority in their rela-

[6] 'University' is used in this chapter to denote a corporate actor, i.e. a formally organised social collective that is able to arrive at decisions about collective goals that are binding to its members, and to utilize its joint resources for achieving these goals. National academic systems obviously vary in the extent to which their universities enjoy this 'actorhood' (Weingart and Maasen, this volume), which depends on both the autonomy granted by the state and the authority over their academics. In order to capture this variation, I describe universities according to the role their management plays in decision-making and resource utilization, i.e. according to the strength of 'hierarchical self-governance' vis-à-vis 'academic self-governance' and 'external hierarchical control' (Schimank 2005).

tionships with their professoriates (Japanese universities are in a similar situation, see Kneller, this volume).

A similar but less strong RES has been introduced in Sweden, where an evaluation of all research and teaching by international peers led to recommendations to universities which, however, are not mandatory. The weakest form of intrusive RES was in place in the second 'episode' of RES in the Netherlands. The evaluation of programmes within universities was first conducted as an *ex ante* evaluation with some consequences for funding and then turned into an ex post evaluation without consequences for funding – or any other consequences in the sense of decisions made by the state. However, the information provided to universities is used by them in their internal management of research performance. This RES can be considered a very weak intrusive form, which increased the autonomy of universities as strategic actors. A similarly weak form is currently in place in Japan, where the government mainly sees to it that data on research performance is collected and that self-evaluations are conducted by universities.

The most unusual case of a RES can be found in Spain. To date, research evaluation is taking place at the individual level, where a quite elaborate system has been developed. The system is transparent but not public. It has weak consequences for the funding of academics, namely for their employability (via *habilitation* and accreditation) and salaries. It can also be said to intrude on the conduct of university research because an evaluation at the national level is a prerequisite for accreditation and *habilitation*, which in turn affect the universities' opportunities to hire academics. The national system of individual evaluation can therefore be considered as a weak intrusive RES. Its consequences for the content and conduct of research and teaching are of great interest but have not yet been investigated except for changes in publication behaviour (an increasing inclination to publish in international journals, Jiménez-Contreras et al. 2003). The contribution by Cruz-Castro and Sanz-Menéndez also reports the emergence of an institutional structure that would support national and regional RES for university research. National and regional authorities have created many evaluation agencies that are principally designed to include research evaluations but haven't done so as yet. This is why the sustainability of the RES for individual academics cannot be assessed at this time. Its major disadvantages appear to be that it provides incentives for individuals but not for any larger aggregation conducting research, and that it does not provide grounds for the allocation of research funding. It might therefore be crowded out by the emerging RES for universities and by internal performance management schemes that will be developed by universities when they gain sufficient autonomy.

Intrusive RES have emerged and are emerging because of a responsibility of the state for the conditions of high quality research in universities. Their strength varies according to the role the state chooses to play. Only strong intrusive RES require a powerful 'visible hand' to achieve the necessary interventions. They depend on a low autonomy of universities and strong hierarchical governance. Weak intrusive RES leave the implementation of change to universities by providing information about necessary and possible improvements of research. When this information is accompanied by recommendations at all, they are not binding. Weak intrusive RES are thus perfectly compatible with a high autonomy of universities. However, it

seems likely that they are nothing but a transition stage. After all there is no reason why universities should do anything about their research when the only punishment for low quality research is to be informed about it in private.

Competitive RES

When information about university research performance is publicly available, transparent and comparative, universities are in a competitive situation because they depend on public opinion, students' choices, and government funding, all of which may be affected by this information. The use of funding formulae in Germany (where they are designed as zero-sum games), the ratings of university departments in the current Dutch RES and the application of an RES to inform funding of Japanese universities from 2010 have in common that they are competitive approaches that have no (Netherlands) or weak (Germany, Japan) consequences for the allocation of research funding. These RES apply peer review (Netherlands) or quantitative indicators (Germany, Japan) in order to evaluate past research performance. They all arrive at information that makes universities comparable and thus creates competition between them. These RES delegate the task of improving research entirely to the universities. The government reserves the rights to measure performance, to inform universities, and to provide incentives for improving research, but no longer decides on internal matters. The refusal of governments to intervene, the gradual increase in the amount of funding controlled by formulae in Germany, the announcement of consequences for funding in Japan, and the renewed discussion about using the evaluation to inform funding in the Netherlands suggest that weak competitive approaches might be a transitional stage towards strong competitive RES. They may go as far as they can under current conditions of professorial autonomy (Germany), provide the ground for subsequent more consequential systems (Japan), or be a stage in a path-dependent policy process in which the interest of the government to introduce a strong RES and the opportunity to do so gradually emerge (Netherlands).

There are also weak competitive RES that are not transition stages but will remain with us from now on. The international and national rankings analysed by Weingart and Maasen will not have strong consequences in terms of funding. Produced by various none-state actors for a variety of reasons (some of which are commercial itself), these rankings are easy to challenge when used as a foundation for policy formulation. Weingart and Maasen have demonstrated why universities nevertheless pay attention to all rankings and respond to them in what must be considered rather frantic ways. All rankings pretend to render university research comparable and thus create a competitive situation. By providing the abstract information that enables comparisons of research performance, rankings create and affect competitions for reputation, students, and funding. This is why universities desperately try to get control of their positions or at least of the interpretation of their positions in rankings. The rankings thus contribute to a competitive climate that supports the introduction of strong competitive RES.

Of the two existing strong competitive RES, the less well known Australian system is represented in this volume. It is usually considered to be entirely different from the other, better known strong competitive RES, namely the UK's Research Assessment Exercise (RAE).[7] The Australian RES measures the performance of universities by applying quantitative indicators, while the RAE assesses university departments by peer review. From a social order perspective, however, the two RES show striking similarities. They both arrive at quantitative representations of quality that feed into formulae for the allocation of resources to universities in a zero-sum game where one university's gain is the others' loss. Both RES are public and transparent because they are designed to make research performance comparable. Consequently, both RES create an intense competition between universities and incentives that are characteristic of markets.

This raises the question whether competitive RES can be considered as markets. Since weak competitive RES only provide information about the relative performance, interpreting them as creating an exchange of research performance for reputation would unbearably stretch the metaphor of exchange. The question is more complicated when strong competitive RES are concerned because they exchange promises for future research performance for funding in bilateral exchanges, and compete for the opportunity to conduct this exchange. To decide whether strong RES are markets, we can draw on the eight basic conditions for a market listed by Teixeira et al. (2004: 4). It turns out that the 'providers' in our actor constellation, the universities, do not have all the basic 'freedoms' of market providers. While universities in both countries are largely autonomous and thus have the *freedom to use available resources*, they are less free in other respects. Australian universities have neither *the freedom of entry* nor the *freedom to specify their product* because the funding formula is applied to all research of all universities. Universities in the UK have those freedoms and use them strategically. However, universities in both countries lack the *freedom to determine prices*.

The governments of both countries, which can be considered to act as the 'consumers' of research, are also constrained. The *adequate information on products and quality* is supplied by the evaluation procedures embedded in strong competitive RES. The dependency of competition on such information explains the explosive growth of rankings (Weingart and Maasen, this volume) and of the use of quantitative measurements of research performance (Gläser and Laudel, this volume). But even if we assume that the 'consumers' have adequate information on quality, their

[7] Given that the RAE is the oldest RES, a very strong RES, and has become a paradigmatic case of a RES in policy discussions around the world, the lack of sociological research on its effects is astounding. There is a flood of what might be termed 'stakeholder literature', i.e. of reports commissioned by science policy (e.g. McNay 1997; Evaluation Associates Ltd 1999; Roberts 2003) and comments by scientists and professional organisations on the impact of the RAE on their discipline (e.g. Harley and Lee 1997; Beecham 1998). Some authors have tried to assess the validity of the RAE by comparing its outcome to bibliometric evaluations (Oppenheim 1997; Adams 2002). Henkel (2000) and Morris (2000) provide some comments on changes in the conduct of research. The recent study by Lucas (2006) describes adaptive behaviour at all levels of the university including individual academics but excludes changes in the conduct and content of research. Thus, comparative sociological investigations of the RAE's impact on the production of scientific knowledge are still missing after 20 years of its existence.

freedoms to chose a provider and the product are limited. They use these rights when designing evaluation and resource allocation procedures but are bound by the outcomes of these procedures. Furthermore, governments neither have *adequate information on prices*, nor do they *pay direct and cost-covering prices*.

As a result of these constraints for 'providers' and 'consumers', universities receive funding that is not at all linked to the costs of research, and are unable to negotiate the amount of funding (or the research they undertake). Strong competitive RES make universities compete for a share of the money their government is willing to spend on university research. While research performance is the major criterion in this competition, the cost of the underlying research is simply irrelevant.

The limited freedoms of providers and consumers in strong competitive RES prove that the latter are only quasi-markets, i.e. actor constellations in which service providers compete for government funding (Le Grand and Bartlett 1993; Barr 1998; Dill et al. 2004). But even though we would reject the application of the concept 'market' to the current strong competitive RES, the social order of the market provides a useful tool for analysing them because they follow a similar economic logic (ibid.: 331). Strong RES apply the logic of the market by allocating funds according to the strength of performance. We can assess the effects of this allocation mechanism in particular, by asking whether some causes of market failure can also cause quasi-market failure, which would imply that we need to look for them in both weak and strong competitive RES.

Responses by Universities

All RES have in common that they change the attitude of universities towards research. As long as the state was absent-minded, universities could leave research and its quality to their academics. RES change this situation by signalling that research performance is observed, and that these observations may trigger actions. In order to keep control of their environment, universities need not only to gain control of the interpretation of their evaluation, as described by Weingart and Maasen in this volume. They also need to take control of their research.

Since they are taking control of their research, universities are becoming the most important interfaces that mediate the impact of RES on the production of scientific knowledge. They shape the immediate conditions under which research is conducted, mainly by organising and funding local research efforts. This provision of research opportunities for members of scientific communities by universities is becoming shaped by the latter's perceptions of and responses to the RES. Structural recommendations need to be implemented, and measures for the improvement of a competitive standing must be designed.

The responses of universities to RES are less well researched than the latter's' institutionalisation und mechanisms. However, some of the studies of the RAE in the UK (Harley and Lee 1997; McNay 1997; Henkel 2000; Morris 2002) as well as studies of universities' responses to the 'second episode' evaluations in the Netherlands (Westerheijden 1997), to the emerging German RES (Schiene and Schimank, Lange, this volume) and to the Australian RES (Gläser and Laudel, this volume)

suggest a surprising similarity of responses to vastly different RES. Regardless of the unit of measurement (university versus discipline within the university), the method of measurement (peer review versus quantitative indicators), and the consequences for funding (none versus significant), universities resort to a set of very few measures for improving their research, namely:

- internal mirroring of funding formulae applied to them;
- improving managerial support for research, in particular for external grant applications;
- changing hiring practices (increasing emphasis on research, a preference for academics with established research records rather than early career researchers, and 'head hunting' of highly reputed researchers);[8]
- internal restructuring with an emphasis on 'critical mass' and interdisciplinary collaboration; and
- increasing the pressure on researchers by applying performance management practices based on quantitative indicators.

These strategies are not applied uniformly in all countries mentioned above. For example, German universities have very limited opportunities to hire or fire their academics. If we take into account these national specifics it becomes even more obvious that universities draw their internal strategies from a rather small arsenal. In particular, responses do not seem to differ systematically between intrusive and competitive or between strong and weak RES.

The limited variation and specific content of responses by universities can be explained by the fact that RES make universities improve a part of their work processes, which they cannot fully control. The formulation of tasks, the generation of quality standards for conduct and the integration of contributions into a common product occurs in scientific communities, which form the primary context of work for research (see above). While universities can set tasks for teaching and control outputs in very much the same way as any professional organisation, the fact that researchers formulate tasks and evaluate results in their scientific communities severely limits the control of research by universities. They therefore can only try to create 'successful black boxes' at the levels of organisational subunits and individuals. This involves hierarchical decision-making that is not in all cases informed by peer review in the same way as the intrusive peer-review based RES described in the previous section.[9]

The improvement of performance at the individual level requires the routine application of research performance measurement, which is impossible to achieve by

[8] Changes in hiring practices do not show a consistent trend because they depend on the autonomy and strategy of universities, and are partly offset by central funding programs aimed at supporting early career researchers.

[9] The consequences of these decisions for the quality of research are not necessarily adverse. In-depth studies of funding programmes for interdisciplinary research collaborations (Laudel 1999; Laudel and Valerius 2001) have shown that 'institutional initiatives' can trigger new collaborations involving new combinations of knowledge that would not have occurred spontaneously. By initiating combinations of knowledge that would not have emerged naturally from the 'normal' course of research, 'bribed' or even 'forced' collaborations *may* yield scientific innovations in form of new lines of interdisciplinary research that would not otherwise have emerged.

peer review and is therefore conducted by applying quantitative indicators (see Gläser and Laudel on bibliometrics, this volume). Both unit-level and individual-level strategies create opportunities for hierarchy failures by enabling centralised organisational decisions under insufficient information.

THE INTERACTIONS OF HIERARCHIES AND COMPETITIONS WITH SCIENTIFIC COMMUNITIES

Success and Possible Failures of RES

In the previous section I have sketched how the contributions to this volume position RES in state hierarchies, competitive constellations including quasi-markets, and organisations (universities). RES change the social orders of governance by strengthening hierarchical elements, competitive elements, or both. Since they have consequences for the conduct of research, they also affect the social order of scientific communities. More precisely, they project specific hierarchies and competition onto scientific communities in order to compensate for the latter's relative indifference to quality.

It is difficult to assess the extent to which RES achieve their aims by improving the quality of research in a national university system. This question can only be asked for RES that have been in place for a long time, i.e. for the UK and the Australian RES. Bibliometric analyses have indeed shown the quality of UK's research to increase in comparison to other countries. Unfortunately, the attribution of this improvement to the RAE is a *post hoc ergo propter hoc* fallacy. The UK is neither the only country to improve its relative performance (countries without RES did so as well), nor is the RAE the only factor that affects relative performance. For example, some comparative studies indicate that the absolute amount of money invested in university research is a much stronger predictor of research performance than any RES (see e.g. Liefner 2003, Lange 2005). If we follow this logic, the RAE might still be the cause of improved performance because it legitimated an increase in spending. However, the causal chain would be entirely different from what proponents of the RAE claim it to be.

While the large-scale improvements of research performance are still in doubt, observers agree that the new quasi-markets and hierarchies of RES have produced a stronger overall concern for quality by extending the domain of consequential quality assessment. By either intervening in the internal management of research or creating competition between universities, RES create an additional feedback channel between the quality of previous research and the opportunity to continue that research. Universities under RES need to care for their research.

These are undoubtedly positive effects of RES. However, according to the many complaints about RES they may come at the price of negative side-effects. As is the case with positive effects, the evidence is by no means conclusive. The empirical basis of criticisms remains weak and largely anecdotal (Gläser et al. 2002). In spite of all the concerns about negative side-effects, they remain the least researched

aspects of RES. Since there is not much evidence beyond the various opinions of affected researchers and managers, I will use our distinction between intrusive and competitive RES for a more systematic discussion that might help identifying possible and likely unintended side-effects of RES.

Both types of RES depend on the validity of the measurement of research performance. The allocation of resources to best performers and the implementation of structural change depend on knowledge about research performance. This common concern of both systems is by no means trivial. For example, it has been repeatedly argued in Australia that the indicators used (particularly numbers of publications and of research students) measure quantity rather than quality, and that by implication quality is not appropriately rewarded (e.g. Butler 2002). The more widespread peer review procedure is under less suspicion of being invalid than the simplified quantitative measures that are commonly used. An extensive peer review, which in Lower Saxony was combined with site visits, is believed to produce the best assessment of research performance possible. However, the literature on peer review indicates that even this 'best practice' of peer review is susceptible to error and can thus create information problems. Peer judgements tend to be biased against high-risk, non-mainstream and interdisciplinary research (see e.g. Chubin and Hackett 1990; Horrobin 1996; Berezin 1998), which means that the quality of this research may be systematically underestimated both by hierarchies relying on peer review and by quasi-markets that use peer review to measure the quality of research.

Since RES can be conceptualised as either hierarchies or quasi-markets, we can utilise the 'market failure' and 'hierarchy failure' perspectives for the discussion of possible failures of RES. There is a vast economic literature on market and government/hierarchy failures, which reduces the problem to inefficiencies of resource allocation.[10] Since the sociological perspective advanced here focuses on the social effects that are linked to such failures, we can limit the discussion of possible 'market failures' of competitive RES and 'hierarchy failures' of intrusive RES to the following few problems.

'Competition failure' of Quasi-market RES

Concerning competition and quasi-markets we first need to consider the argument that markets are bad for research per se. This belief is nicely countered by Cozzen's chapter, which demonstrates that markets for research proposals can work beautifully – if there is a variety of overlapping markets, and if there is enough money for them to allocate (see also Geiger 2004). This is also indicated by Kneller's account

[10] 'Market failure' and 'government failure' refer to the failure of these actor constellations to achieve static efficiency or, more precisely, Pareto-efficient resource allocations. Pareto efficiency is defined as a situation in which "it is impossible to make anyone better off … without making someone else worse off" (Bator 1958: 351). For a discussion of market failure, see the classic texts in the collection edited by Cowen (1988) and the important text by Arrow (1962) which is curiously missing from Cowen's book. More recent discussions of market failure can be found in Boadway and Wildasin (1984: 55-82) and Stiglitz (2000: 76-92). Government failure is discussed by Le Grand (1991). Overviews of both market and government failure are provided by Wolf (1988), Levačić (1991), and Lipsey and Chrystal (1999: 285-332).

of the development of ex ante-research evaluation in Japan where the opportunities for scientific innovations are improving. A counterexample is the external funding in Australia where many academics can bid in only one market that allocates very little money. These are conditions under which markets for research proposals support hierarchies in scientific communities in very much the same way as do peer-review based RES. By collectively evaluating *all* externally funded research of its national community, a scientific elite can effectively control that research.

Having established that quasi-markets can solve some allocation problems of research funding, we now need to look for possible causes of 'market failure' or, in the case of quasi-markets, 'competition failure'. The latter refinement already indicates that not all possible causes of market failure apply to competitive RES.[11] The two possible causes of 'competition failure' that need concern us here are negative externalities and public goods. Negative externalities are effects of market transactions that occur outside the market, i.e. effects which are not absorbed by the exchange parties.[12] 'Public goods' are goods that are non-rivalrous (their consumption by one does not affect the consumption by others) and non-excludable (it is impossible or too costly to exclude anyone from the consumption). A classic and extensively discussed example for such a good is academic scientific knowledge (Arrow 1962; Dasgupta and David 1994; Callon 1994; David et al. 1999).

Both the design of the two strong competitive RES (in the UK and Australia) and the available information organisational responses to these systems suggests at least two trends that fit both the description of negative externalities and – in a slightly different perspective – that of the undersupply of public goods. The first trend is a loss of research diversity. Research diversity – the diversity of different approaches to a problem – is a property of scientific fields that cuts across competing universities. In analogy to the biodiversity discussion in ecology (e.g. Purvis and Hector 2000: 216-217) it is believed to facilitate the stability and growth of science, because of interactions between the approaches or because a higher diversity means a higher likelihood that one of the approaches is highly productive. Competitive RES create incentives for universities to favour the mainstream against other approaches and thus contribute to a loss of diversity. The strongest incentive is created when the evaluation procedures of competitive RES define a minimal size of the evaluated unit. This is the case for the RAE in the UK, where assessments of individual academics' publications are synthesised into grades for departments, and its planned small-scale copy in Australia, where the same will happen for research groups of at least five academics (DEST 2006: 16). When such coherent groups of academics need to be submitted for evaluation and funding, universities are under pressure to

[11] We can exclude misallocation due to asymmetric information because the evaluations that establish 'product quality' are conducted by the 'buyer' (the government) who shares them with the universities (the 'sellers'). We can also exclude monopoly power to dictate prices because the 'prices' do not reflect costs but are dictated by the government to begin with – which of course could be considered a use of monopoly power by the state and thus the most general cause of a 'market failure' of competitive RES.

[12] A recent popular example is 'climate change'(global warming), which has been called the 'greatest and widest-ranging market failure ever seen' (Stern 2006, Executive Summary: 1). Economic development has caused greenhouse gas emissions whose (now perceived) immense costs have not been included in the prices of goods.

construct them. More generally, the contributions by Lange, Schiene and Schimank, and Gläser and Laudel suggest that *any* competitive situation motivates universities to create 'critical masses' in order to gain a 'competitive advantage'.

Creating larger coherent units of research requires either identifying them within one field or creating them by amalgamating researchers from several fields. Since research belonging to the mainstream is much more likely to achieve the required critical mass, and since not all fields are equally easy to amalgamate, the procedure of creating coherent research units of a minimal size is necessarily selective in that it tends to crowd out non-mainstream and small 'incompatible' fields or approaches within fields. This trend has been observed for the small 'exotic' disciplines in Germany.[13] Since these disciplines nowhere achieve a 'critical mass' they are in danger of being eliminated by all universities for the sake of 'profile building' and 'creating critical masses'. This could lead to a serious weakening or even the disappearance of these fields at the national level. The German University Rectors' Conference (HRK) recently described this danger:

> Facing an increasing scarcity of public money and the need to create a critical mass for larger interdisciplinary clusters, universities are forced to define priorities and *posteriorities* that particularly affect the small disciplines. Since this priority setting is part of the universities' profile-building processes, which often occur parallel and without supra-regional coordination, the threat to small disciplines increases nationwide. (HRK 2007: 4, my translation)

Another incentive for universities to favour the mainstream is the necessity to increase the likelihood of positive evaluations without being able to evaluate the research themselves. Under these conditions universities are likely to support approaches that conform to general models of 'good research', which contributes to the crowding out of alternative approaches.

These mechanisms may produce the decrease of diversity that Whitley suggests as a possible consequence of strong RES in his introductory chapter. The preceding discussion indicates that a decline of diversity might also be caused by weak competitive RES because the mechanisms operate whenever universities respond uniformly to a competitive situation.[14]

A second general condition of knowledge production that appears to suffer in competitive situations is the stability of scientific careers. The head hunting practiced by British and to some extent by Australian universities and the tendency of Australian universities to hire only academics with a well established research record at entrance positions indicate that all universities try to benefit from successful research careers while the provision of such careers is of lesser concern. The political worries about brain drains and about young academics needing more time to become successful researchers indicate that career patterns might change (Roberts 2002; Bazeley 2003; Åkerlind 2005; Laudel 2005; Laudel and Gläser 2007).

---

[13] I am grateful to Uwe Schimank who alerted me to this effect.

[14] The contribution by Schiene and Schimank in this volume suggests that even non-competitive (i.e. intrusive) RES might trigger these mechanisms when the scientific elites who formulate recommendations uniformly favour the creation of 'critical masses'.

Depending on the perspective applied, a loss of research diversity and the destabilisation of academic careers fit both the negative externalities and the public good problem. If we consider competitive RES as an actor constellation including the government and universities, both consequences of the quasi-market competition between universities are negative externalities because they affect scientific communities and their members, i.e. actors who are not involved in the quasi-market transactions. If we look at scientific communities as actor constellations producing public goods including not only knowledge but also structural prerequisites of knowledge production, the threats to diversity and careers appear as the consequence of an imposed quasi-market's inability to produce these public goods because they cannot be produced by allocating resources to the strongest performers.

'Hierarchy failure' of Intrusive RES

A hierarchy failure (an inefficient allocation of resources by government intervention) occurs because of conflicting objectives or because of information problems (Lipsey and Chrystal 1999: 328-329). The notion of 'conflicting objectives' covers all instances where the evaluation and intrusion might be affected by aims unrelated to the improvement of research quality. For example, an intervention might be aimed at strengthening a field of research that is considered politically important by the government. Intrusive RES are susceptible to this failure. An example has been provided by the government of Lower Saxony, which tried to close a department of sociology before the evaluation report was submitted and against its recommendations (Rehberg 2003). This hierarchy failure of RES is erratic by nature and inherently difficult to legitimize because it is judged against the recommendations of the peer review. It is therefore unlikely to produce systematic distortions of research – at least as far as its occurrence in intrusive RES is concerned.

Insufficient information can cause hierarchy failures as well because a central decision-maker might not have all the necessary information about local conditions (Hayek 1945) or because the central decision-maker is not flexible enough to respond to changes in the conditions for resource allocation (Lipsey and Chrystal 1999: 328). Decision-making under insufficient information may occur in RES if the state cannot validly assess the research performance of universities (see above). This information problem and an accompanying hierarchy failure is even more likely to occur within organisations, which cannot afford the necessary extensive performance assessments in their everyday decision-making and therefore must rely on the expertise of colleagues who are not peers, or on radically simplified quantitative measures. They appear commonly to resort to the latter, thereby achieving comparability across fields but sacrificing validity and forfeiting information about context (conditions of work) or potential. Universities are thus more or less forced to base their decisions about research on insufficient information.

While peer review does not lead to hierarchy failures in state-university interaction, it creates control hierarchies where none have been before, namely in scientific communities. Scientific communities are characterised by a strong stratification (Cole and Cole 1973) and have always featured informal hierarchies in

which members of the elite act as gatekeepers for projects, positions, and publications. However, this relationship between 'average' community members and their elite was pluralistic insofar there was rarely a case of an intellectually and socially coherent elite controlling all research of its community. The integration of elites in RES has the potential to change that. By being integrated into RES, the elite for the first time has the opportunity to comparatively evaluate all research of its national community and to influence the financial support for it. Schiene and Schimank concluded from their analysis of the peer review procedure in Lower Saxony that this constellation empowers the elite and limits the autonomy of the rest of the community. The resulting more systematic approach to the quality of a community's research and to the conditions under which it is conducted is advantageous for those who fund the research because decisions on funding can be both guided and legitimised. It is also advantageous for those who fit the established view of what quality is, including the elites who create the definition. At the same time, this constellation makes it much more difficult for members of a scientific community to oppose the dominant views (dominant because held by the elite or by the majority). In my discussion of the mode of production of scientific communities, I have identified the variety of perspectives that can be brought to bear on a problem simultaneously, as one of the advantages of the communal mode of production. As Whitley has pointed out in the introduction to this volume, the investment of an elite with new evaluative and distributive powers has the potential to reduce this diversity. While peer-review based RES don't create hierarchy failures in state-university relationships because appropriate information is sought, they may initiate hierarchy failures in scientific communities because a control hierarchy is projected onto scientific communities where only a much weaker and incoherent one is supposed to exist.

Similar to the quasi-markets of competitive RES, intrusive RES project their social order onto scientific communities and thereby change the conditions of knowledge production. Also similar to the competitive RES, the relationship that is projected is not entirely alien to scientific communities. Both competition and hierarchies are intrinsic to scientific communities. However, the projection of the social orders of RES tends to amplify these relationships and thereby limits the capacity of scientific communities to utilise their anarchic progress as a means to compensate for market and hierarchy failures.

INCENTIVES FOR FUTURE RESEARCH

The discussion about the role of RES in the social orders that govern science identified some ways in which institutions and practices of governance are linked to the production of scientific knowledge. These links occur at two levels. At the micro-level, academics respond to their funding environment and to performance management practices of universities. They try to find a compromise between these signals and a fruitful and manageable 'research trail' (Chubin and Connolly 1982). At the macro level, the aggregate effects of these individual responses, the hierarchy and market failures of RES, and the projection of these failures onto scientific communities change social structures of and processes in scientific communities. As a result

of both micro-level and macro-level changes, scientific knowledge is produced under different conditions, which means that different knowledge is likely to be produced.

With the notable exception of the US, many highly developed countries are currently institutionalising RES for their university systems. RES have the potential to become an ubiquitous background of knowledge production similar to the ex-ante peer review of grant proposals. What researchers can do, and how they can do it, will then depend on those two systems. The consequences of this emerging situation for the conduct and content of knowledge production are difficult to envision. Their investigation needs to be comparative across countries and fields, and needs to address both the knowledge construction by individual actors and the conditions of action that mediate between RES and the individual level. This is not only an interesting opportunity to link constructivist and institutional analyses and to meet the theoretical and methodological challenges posed by such a link. It is also an opportunity to demonstrate the political relevance of our insights in the social construction of scientific knowledge.

REFERENCES

Adams, J. (2002), 'Research Assessment in the UK', *Science*, 296, 805-805.

Åkerlind, Gerlese S. (2005), 'Postdoctoral Researchers: Roles, Functions and Career Prospects, and Development', *Higher Education Research & Development*, 24, 21-40.

Arrow , Kenneth J. (1962), 'Economic Welfare and the Allocation of Resources for Invention', in National Bureau of Economic Research (ed.), *The Rate and Direction of Investment Activity: Economic and Social Factors*, Princeton: Princeton University Press, pp. 609-625.

Barley, Stephen R. and Gideon Kunda (2001), 'Bringing Work Back In', *Organization Science*, 12, 76-95.

Barr, Nicholas (1998), *The Economics of the Welfare State*, London: Weidenfeld and Nicholson.

Bator, Francis M. (1958), 'The Anatomy of Market Failure', *Quarterly Journal of Economics*, 72, 351-379.

Bazeley, Pat (2003), 'Defining 'Early Career' in Research', *Higher Education*, 45, 257-279.

Beecham, Linda (1998), 'Medical Academics Criticise Research Assessment Exercise', *British Medical Journal*, 316, 481.

Benkler, Yochai (2002), 'Coase's Penguin, or, Linux and The Nature of the Firm', *Yale Law Journal*, 112, 369-446.

Berezin, Alexander (1998), 'The Perils of Centralized Research Funding Systems', *Knowledge, Technology & Policy*, 11, 5-26.

Boadaway, Robin W. and David E. Wildasin (1984), *Public Sector Economics*, Boston: Little, Brown & Co.

Butler, Linda (2002), 'A list of published papers is no measure of value - The present system rewards quantity, not quality - but hasty changes could be as bad', *Nature*, 419, 877.

Callon, Michel (1994), 'Is Science a Public Good? Fifth Mullins Lecture, Virginia Polytechnic Institute, 23 March 1993', *Science, Technology & Human Values*, 19, 395-424.

Chubin, Daryl E. and Terence Connolly (1982), 'Research Trails and Science Policies', in Norbert Elias, Herminio Martins and Richard Whitley (eds.), *Scientific Establishments and Hierarchies*. Dordrecht: Reidel, 293-311.

Chubin, Daryl E. and Edward J. Hackett (1990), *Peerless Science: Peer Review and U.S. Science Policy*, Albany, N.Y.: State University of New York Press.

Clark, Burton R. (1983), *The Higher Education System : Academic Organization in Cross-National Perspective*, Berkeley: University Of California Press.

Cole, Jonathan R. and Stephen Cole (1973), *Social Stratification in Science*, Chicago: The University of Chicago Press.

Cowen, Tyler (ed.) (1988), *The Theory of Market Failure: A Critical Examination*, Fairfax: George Maseon University Press.

Cozzens, Susan E., Peter Healey, Arie Rip and John Ziman (eds.) (1990), *The Research System in Transition*, Dordrecht: Kluwer.

Dasgupta, Partha and Paul A. David (1994), 'Toward a New Economics of Science', *Research Policy*, 23, 487-521.

David, Paul A., Dominique Foray and Edward Steinmueller (1999), 'The Research Network and the New Economics of Science: from Metaphors to Organizational Behaviors', in Alfonso Gambardella and Franco Malerba (eds.), *The Organization of Economic Innovation in Europe*, Cambridge: Cambridge University Press, pp. 303-342.

DEST [Department of Education, Science and Training] (2006), *Research Quality Framework: Assessing the Quality and Impact of Research in Australia - The Recommended RQF*, Canberra: DEST.

Dill, David, Pedro Teixeira, Ben Jongbloed and Alberto Amaral (2004), 'Conclusion', in Pedro Teixeira, Ben Jongbloed, David Dill and Alberto Amaral (eds.), *Markets in Higher Education: Rhetoric or Reality*, Dordrecht: Kluwer, pp. 327-352.

Enders, Jürgen (2002), 'Governing the Academic Commons: About Blurring Boundaries, Blistering Organisations, and Growing Demands', in University of Twente (ed.), *The CHEPS Inaugurals 2002*, Enschede: University of Twente, pp. 69-105.

Evaluation Associates Ltd (1999), *Interdisciplinary Research and the Research Assessment Exercise. A Report for the UK Higher Education Funding Bodies*, http://admin.hero.ac.uk/rae/niss/1_99.doc (accessed 31 May 2007).

Geiger, Roger L. (2004), 'Market Coordination of Higher Education: The United States', in Pedro Teixeira, Ben Jongbloed, David Dill and Alberto Amaral (eds.), *Markets in Higher Education: Rhetoric or Reality*, Dordrecht: Kluwer, pp. 161-183.

Gläser, Jochen (2006), *Wissenschaftliche Produktionsgemeinschaften. Die soziale Ordnung der Forschung*, Frankfurt am Main: Campus.

Gläser, Jochen (2007), 'The Social Order of Open Source Software Production', in Kirk St.Amant and Brian Still (eds.), *Handbook of Research on Open Source Software: Technological, Economic, and Social Perspectives*, Hershey, PA: Idea Group Inc. (forthcoming).

Gläser, Jochen, Grit Laudel, Sybille Hinze and Linda Butler (2002), *Impact of Evaluation-based Funding on the Production of Scientific Knowledge: What to Worry About, and How to Find out,* http://repp.anu.edu.au/expertise-glae-lau-hin-but.pdf.

Harley, Sandra and Frederic S. Lee (1997), 'Research Selectivity, Managerialism, and the Academic Labor Process: The Future of Nonmainstream Economics in U.K. Universities', *Human Relations*, 50, 1427-1460.

Hayek, F.A. (1945), 'The Use of Knowledge in Society', *The American Economic Review*, 35, 519-530.

Hayek, Frederick (1991), 'Spontaneous ('Grown') Order and Organized ('Made') Order', in Grahame Thompson, Jennifer Frances, Rosalind Levačić and Jeremy Mitchell (eds.), *Markets, Hierarchies and Networks. The Coordination of Social Life*, London: SAGE, pp. 293-301.

Henkel, Mary (2000), *Academic Identities and Policy Change in Higher Education*, London: Jessica Kingsley.

Horrobin, David F. (1996), 'Peer Review of Grant Applications: A Harbinger for Mediocrity in Clinical Research?', *Lancet*, 348, 1293-1295.

HRK [Hochschulrektorenkonferenz] (2007), *Die Zukunft der Kleinen Fächer: Potenziale – Herausforderungen – Perspektiven*, Bonn: HRK.

Jiménez-Contreras, Evaristo, Félix De Moya Anegón and Emilio Delgado López-Cózar (2003), ‚The Evolution of Research Activity in Spain - The Impact of the National Commission for the Evaluation of Research Activity (CNEAI)', *Research Policy*, 32, 123-142.

Knorr-Cetina, Karin (1995), 'How Superorganisms Change: Consensus Formation and the Social Ontology of High-Energy Physics Experiments', *Social Studies of Science*, 25, 119-147.

Lange, Stefan (2005), ‚Hochschul-Governance im Wandel. Neuere Beiträge der vergleichenden Hochschulforschung', *Soziologische Revue*, 27, 309-321.

Laudel, Grit (1999), *Interdisziplinäre Forschungskooperation: Erfolgsbedingungen der Institution 'Sonderforschungsbereich'*, Berlin: Edition Sigma.

Laudel, Grit (2005), 'Migration Currents among the Scientific Elite', *Minerva*, 43, 377-395.

Laudel, Grit and Jochen Gläser (2007), 'From apprentice to colleague: The metamorphosis of Early Career Researchers', *Higher Education* (in press).

Laudel, Grit and Gabriele Valerius (2001), *Innovationskollegs als ‚Korrekturinstitutionen' im Institutionentransfer? Abschlussbericht zum DFG-Projekt ‚Innovationskollegs als Instrument der Umgestaltung*

der unviversitären Forschung im ostdeutschen Transformationsprozess - Akteure, Strukturen und Effekte', FIT Arbeitsberichte, Frankfurt (Oder): Europa-Universität Frankfurt, Frankfurter Institut für Transformationsforschung.

Le Grand, Julian, (1991), 'The Theory of Government Failure', *British Journal of Political Science*, 21, 423-442.

Le Grand, Julian and Will Bartlett (eds.) (1993), *Quasi-Markets and Social Policy*, Houndmills: Macmillan Press.

Levačić, Rosalind (1991), 'Markets and Government: An Overview', in Grahame Thompson, Jennifer Frances, Rosalind Levačić and Jeremy Mitchell (eds), *Markets, Hierarchies and Networks. The Coordination of Social Life*, London: SAGE, pp. 35-47.

Liefner, Ingo (2003), 'Funding, resource allocation, and performance in higher education systems', *Higher Education*, 46, 469-489.

Lipsey, Richard G. and K. Alec Chrystal (1999), *Principles of Economics*, Oxford: Oxford University Press.

Lucas, Lisa (2006), *The Research Game in Academic Life*, Maidenhead: The Society for Research in Higher Education and Open University Press.

McNay, Ian (1997), *The Impact of the 1992 Research Assessment Exercise on Individual and Institutional Behaviour in English Higher Education: Summary Report and Commentary*, Anglia Polytechnic University.

Morris, Norma (2000), 'Science Policy in Action: Policy and the Researcher', *Minerva*, 38, 425-451.

Morris, Norma (2002), 'The Developing Role of Departments', *Research Policy*, 31, 817-833.

Oppenheim, Charles (1997), 'The Correlation Between Citation Counts and the 1992 Research Assessment Exercise Ratings for British Research in Genetics, Anatomy and Archaelogy', *Journal of Documentation*, 53, 477-487.

Perrow, Charles (2002), *Organizing America. Wealth, Power, and the Origins of Corporate Capitalism*, Princeton: Princeton University Press.

Purvis, Andy and Andy Hector (2000), Getting the Measure of Biodiversity, *Nature*, 405, 212-219.

Rehberg, Karl-Siegbert (2003), *Brief an den niedersächsischen Minister für Wissenschaft und Kultur vom 13. Oktober 2003*, Dresden, http://www.soziologie.de/dokumente/stellung_hannover.pdf (accessed 19 February 2007).

Roberts, Sir Gareth (2002), *Set for Success: The Supply of People with Science, Technology, Engineering and Mathematics Skills*, London: HM Treasury.

Roberts, Sir Gareth (2003), *Review of Research Assessment*, London: Higher Education Funding Council for England, http://www.ra-review.ac.uk/reports/roberts.asp.

Schimank, Uwe (2005), ''New Public Management' and the Academic Profession: Reflections on the German Situation', *Minerva*, 43, 361-376.

Stern, Nicholas (2006), *The Economics of Climate Change*, http://www.hm-treasury.gov.uk./independent_reviews/stern_review_economics_climate_change/stern_review_report.cfm (accessed 31 May 2007).

Stiglitz, Joseph E. (2000), *Economics of the Public Sector*, New York and London: W W Norton & Company.

Teixeira, Pedro, Ben Jongbloed, Alberto Amaral and David Dill (2004), 'Introduction', in Pedro Teixeira, Ben Jongbloed, David Dill and Alberto Amaral (eds.), *Markets in Higher Education: Rhetoric or Reality*, Dordrecht: Kluwer, pp. 1-12.

Westerheijden, Don F. (1997), 'A Solid Base for Decisions: Use of VSNU Research Evaluations in Dutch Universities', *Higher Education*, 33, 397-413.

Whitley, Richard (2000 [1984]), *The Intellectual and Social Organization of the Sciences*, New York: Oxford University Press.

Whitley, Richard (2003), 'Competition and Pluralism in the Public Sciences: The Impact of Institutional Frameworks on the Organisation of Academic Science', *Research Policy*, 32, 1015-1029.

Wolf, Charles (1988), *Markets or Governments: Choosing between Imperfect Alternatives*, Cambridge, MA: The MIT Press.

AUTHOR INDEX

Sociology of the Sciences

1. E. Mendelsohn, P. Weingart and R. Whitley (eds.): *The Social Production of Scientific Knowledge.* 1977 ISBN Hb 90-277-0775-8; Pb 90-277-0776-6
2. W. Krohn, E.T. Layton, Jr. and P. Weingart (eds.): *The Dynamics of Science and Technology.* Social Values, Technical Norms and Scientific Criteria in the Development of Knowledge. 1978 ISBN Hb 90-277-0880-0; Pb 90-277-0881-9
3. H. Nowotny and H. Rose (eds.): *Counter-Movements in the Sciences.* The Sociology of the Alternatives to Big Science. 1979 ISBN Hb 90-277-0971-8; Pb 90-277-0972-6
4. K.D. Knorr, R. Krohn and R. Whitley (eds.): *The Social Process of Scientific Investigation.* 1980 (1981) ISBN Hb 90-277-1174-7; Pb 90-277-1175-5
5. E. Mendelsohn and Y. Elkana (eds.): *Sciences and Cultures.* Anthropological and Historical Studies of the Sciences. 1981 ISBN Hb 90-277-1234-4; Pb 90-277-1235-2
6. N. Elias, H. Martins and R. Whitley (eds.): *Scientific Establishments and Hierarchies.* 1982 ISBN Hb 90-277-1322-7; Pb 90-277-1323-5
7. L. Graham, W. Lepenies and P. Weingart (eds.): *Functions and Uses of Disciplinary Histories.* 1983 ISBN Hb 90-277-1520-3; Pb 90-277-1521-1
8. E. Mendelsohn and H. Nowotny (eds.): *Nineteen Eighty Four: Science between Utopia and Dystopia.* 1984 ISBN Hb 90-277-1719-2; Pb 90-277-1721-4
9. T. Shinn and R. Whitley (eds.): *Expository Science.* Forms and Functions of Popularisation. 1985 ISBN Hb 90-277-1831-8; Pb 90-277-1832-6
10. G. Böhme and N. Stehr (eds.): *The Knowledge Society.* The Growing Impact of Scientific Knowledge on Social Relations. 1986 ISBN Hb 90-277-2305-2; Pb 90-277-2306-0
11. S. Blume, J. Bunders, L. Leydesdorff and R. Whitley (eds.): *The Social Direction of the Public Sciences.* Causes and Consequences of Co-operation between Scientists and Non-scientific Groups. 1987 ISBN Hb 90-277-2381-8; Pb 90-277-2382-6
12. E. Mendelsohn, M.R. Smith and P. Weingart (eds.): *Science, Technology and the Military.* 2 vols. 1988 ISBN Vol, 12/1 90-277-2780-5; Vol. 12/2 90-277-2783-X
13. S. Fuller, M. de Mey, T. Shinn and S. Woolgar (eds.): *The Cognitive Turn.* Sociological and Psychological Perspectives on Science. 1989 ISBN 0-7923-0306-7
14. W. Krohn, G. Küppers and H. Nowotny (eds.): *Selforganization.* Portrait of a Scientific Revolution. 1990 ISBN 0-7923-0830-1
15. P. Wagner, B. Wittrock and R. Whitley (eds.): *Discourses on Society.* The Shaping on the Social Science Disciplines. 1991 ISBN 0-7923-1001-2
16. E. Crawford, T. Shinn and S. Sörlin (eds.): *Denationalizing Science.* The Contexts of International Scientific Practice. 1992 (1993) ISBN 0-7923-1855-2
17. Y. Ezrahi, E. Mendelsohn and H. Segal (eds.): *Technology, Pessimism, and Postmodernism.* 1993 (1994) ISBN 0-7923-2630-X
18. S. Maasen, E. Mendelsohn and P. Weingart (eds.): *Biology as Society? Society as Biology: Metaphors.* 1994 (1995) ISBN 0-7923-3174-5
19. T. Shinn, J. Spaapen and V. Krishna (eds.): *Science and Technology in a Developing World.* 1995 (1997) ISBN 0-7923-4419-7

20. J.Heilbron, L. Magnusson and B.Wittrock (eds.): *The Rise of the Social Sciences and the Formation of Modernity.* Conceptual Change in Context, 17501850. 1996 (1998) ISBN 0-7923-4589-4
21. M. Fortun and E. Mendelsohn (eds.): *The Practices of Human Genetics.* 1998 ISBN 0-7923-5333-1
22. B. Joerges and T. Shinn (eds.): *Instrumentation: Between Science, State and Industry*. 2000 ISBN 0-7923-6736-7
23. B. Joerges and H. Nowotny (eds.): *Social Studies of Science and Technology: Looking Back Ahead*. 2003 HB ISBN 1-4020-1481-3; Pb 1-4020-1482-1
24. S. Maassen and P. Weingart (eds.): *Democratization of Expertise. Exploring* N*ovel* Forms of Scientific Advice in Political Decision- Making. 2005 HB ISBN 1-4020.3753-8; Pb 1-4020-3754-6
25. J. Lenhard, G. Kueppers, T. Shinn (eds.): *Simulation: Pragmatic Construction of Reality.* 2006 ISBN 1-4020-5374-6
26. R. Whitley and J. Gläser (eds.): *The Changing Governance of the Sciences*. The Advent of Research Evaluation Systems. 2007 ISBN 978-1-4020-6745-7