RESEARCH GOVERNANCE IN ACADEMIA:
ARE THERE ALTERNATIVES TO ACADEMIC RANKINGS?

by
Margit Osterloh\textsuperscript{a,c,*}, Bruno S. Frey\textsuperscript{b,c}

\textsuperscript{a}University of Zurich
Institute of Organization and Administrative Sciences
Universitätsstrasse 84
CH-8006 Zürich, Switzerland
Email: osterloh@iou.uzh.ch

\textsuperscript{b}University of Zurich
Institute for Empirical Research in Economics
Winterthurerstrasse 30
CH-8006 Zürich, Switzerland
Email: bsfrey@iew.uzh.ch

\textsuperscript{c}CREMA - Center for Research in Management, Economics and the Arts, Zurich
Gellertstrasse 18, CH-4052 Basel, Switzerland

\* Corresponding address: Prof. Dr. Dr. h.c. Margit Osterloh, Institute of Organization and Administrative Science, Universitätsstrasse 84, CH-8006 Zürich, Switzerland.
Tel.: +41 44 634 28 41; fax: +41 44 634 49 42
E-mail: osterloh@iou.uzh.ch.
RESEARCH GOVERNANCE IN ACADEMIA:
ARE THERE ALTERNATIVES TO ACADEMIC RANKINGS?

Abstract

Peer reviews and rankings today are the backbone of research governance, but recently came under scrutiny. They take explicitly or implicitly agency theory as a theoretical basis. The emerging psychological economics opens a new perspective. As scholarly research is a mainly curiosity driven endeavor, we include intrinsic motivation and supportive feedback by the peers as important determinants of scholarly behavior. We discuss whether a stronger emphasis on selection and socialization offers an alternative to the present regime of academic rankings.

(80 words)

Key words: peer reviews, rankings, research governance, agency theory, psychological economics, new public management, economics of science, control theory.
Peer reviews and academic rankings are generally considered the backbone of research governance in academia. The recent, lively discussion about the quality of peer reviews (e.g., Abramo, Angelo, & Caprasecca, 2009; Frey, 2003; Starbuck, 2005, 2006) and academic rankings (e.g., Adler & Harzing, 2009; Lawrence, 2002, 2003; Weingart, 2005) focused mainly on the issues of method and how to improve it. However, the question was not raised on which theoretical background peer reviews and rankings are based and whether this background is adequate for research governance in academia. Until today, there is no discussion about whether there are alternatives to the dominant principles of academic research governance.

We argue that these principles explicitly or implicitly follow mainly from the principal agent view, which in the literature on corporate governance has come under fire due to corporate failures and scandals (e.g., Benz & Frey, 2007; Daily, Dalton, & Canella, 2003). A corresponding, critical discussion in the field of research governance in academia is lacking. In our paper, we contribute to this discussion by confronting two different perspectives, agency theory and the newly emerging psychological economics.¹ The latter approach builds on psychologically informed economics. In line with an understanding of scholarly research as a mainly curiosity driven endeavor, it includes intrinsic motivation as a major determinant of scholarly behavior. We combine this approach with managerial control theory based on the work of Ouchi (1977, 1979) and discuss the implications derived from the two different perspectives. While agency theory
counts on the refinements of indicators and the measurement process, psychological economics builds mainly on the careful selection and socialization of scholars, on supporting feedback, and on symbolic benefits like awards.

We begin by analyzing the theoretical basis of the current dominant governance system for academic research, namely agency theory combined with the economics of science. The second section presents empirically based findings on present research governance, discussing the advantages and disadvantages of its backbones, namely peer reviews and academic rankings. Psychological economics then is suggested as a theoretical basis for academic governance, appreciating the unique features of research in academia. The fourth section considers the implications of the two perspectives for research governance in academia. The last section concludes by arguing that psychological economics presents a promising, novel avenue for research in academic governance.

THEORETICAL BASIS OF THE PRESENT GOVERNANCE SYSTEM FOR ACADEMIC RESEARCH

Over the past years, universities have increasingly adopted the idea that the governance of academic research should be subjected to the same governance as for-profit enterprises. This is reflected in procedures transferred from private companies. The most prominent examples are pay-for-performance for scholars according to output measures like rankings, ratings, and competitive fundraising. Overall, the reforms are aimed at the
establishment of an “enterprise university” (e.g., Bok, 2003; Clark, 1998; Donoghue, 2008; Marginson & Considine, 2000; for business schools, see Khurana, 2007).

This concept is based on new public management and economics of science. The proponents of new public management draw on the principal agent view (Kaboolian, 1998) as proposed by Jensen and Murphy (1990). This view dominates new public management (Burgess & Ratto, 2003) in the same way it dominates corporate governance (Daily et al., 2003). Economics that “has won the battle for theoretical hegemony in academia and society as a whole” (Ferraro, Pfeffer, & Sutton, 2005: 10) has come to dominate the analysis of all spheres of life, for instance, the family, art, sport, and religion (Becker, 1976). Today this approach is also applied to academia, either implicitly (e.g., Worell, 2009) or explicitly (e.g., Deem, 2004; Schimank, 2005). According to agency theory, scholars have to be monitored and sanctioned in the same way as managers. The underlying assumption is that control and correctly administered pay-for-performance schemes contain the potential for opportunistic behavior, boost productivity, and lead to an efficient allocation of resources (Lavy, 2007; Swiss, 2005).

According to the economics of science in academia the evaluation by the market has to be substituted for the evaluation of peers in the self-governed “republic of science” (Polanyi, 1962). This is the case because of two fundamental characteristics of science, its high uncertainty and its public nature (Dasgupta & David, 1994; Nelson, 1959, 2004; Stephan, 1996).
The fundamental uncertainty of scientific endeavors is due to the fact that success in academia is reflected by success in the market often only after a long delay or sometimes not at all (Bush, 1945; Nelson, 2006). In addition, research often produces serendipity effects; that is, it provides answers to unposed questions (Stephan, 1996). As it is often not predictable how useful a particular research endeavor produces is and whether it ever will be marketable, peers instead of the market have to evaluate whether a piece of research represents an advance.

The public nature of scientific discoveries has been intensively discussed by Arrow (1962) and Nelson (1959, 2006). A discovery must be communicated by scholars as quickly as possible to the community of peers in order to be recognized as the discoverer of a new scientific idea (Dasgupta & David, 1994). In contrast, in profit-oriented enterprises, incentives to transform scientific results into a public good are normally absent.\(^2\)

As a consequence of these characteristics of research in academia, the “priority rule” has been established as the main success criterion (Dasgupta & David, 1994; Merton, 1957; Stephan, 1996; Gittelman & Kogut, 2003). Only peers can establish scientific priority. Consequently, the peer review system is taken to be the founding stone of academic research evaluation. Instances are awards, honorary doctorates, or membership in prestigious academies (Stephan, 1996; Frey & Neckermann, 2008). Its main form for the majority of scholars consists of publications and citations in
professional journals with high impact factors. Such indicators are provided by academic rankings, based on peer-reviewed publications, citations, and the impact factor of journals like Thomson Scientific’s Impact Factor (JIF) (see Garfield, 2006, for a historical review) and the recent h-index (Hirsch, 2005).

Indeed, a well-designed governance system based on peer reviews and academic rankings seems to combine perfectly an output-oriented evaluation of researchers, as postulated by new public management, on the one side, with the requirements of a peer-based evaluation system, as postulated be the economics of science on the other side. Therefore, today these measures are adopted almost universally in academia for most things that matter: tenure, salary, postdoctoral grants, and budget decisions.

However, in recent times a broad discussion arose about the quality of peer reviews (e.g., Starbuck, 2005, 2006) and academic rankings (e.g., Adler, Ewing, & Taylor, 2008; Adler & Harzing, 2009; Lawrence 2002, 2003). It focused mainly on the issues of method, while the theoretical background on which these measures are based was not questioned.

EMPIRICALLY BASED FINDINGS ON THE PRESENT RESEARCH

GOVERNANCE IN ACADEMIA

Findings on Qualitative Peer Reviews

Peer reviews are the backbone of the research governance and evaluation system in academia. However, in recent times, it has been argued that the present peer review
system has major problems (e.g., Abramo et al., 2009; Bedeian, 2004; Campanario, 1996; Frey, 2003; Gillies, 2005, 2008; Starbuck, 2005, 2006; Tsang & Frey, 2007; Wenneras & Wold, 1999).³

**Low inter-rater reliability**

There is an extensive literature on the low extent to which reviews conform to each other (Cole, 1992; Miner & MacDonald, 1981; Weller, 2001). The correlation between the judgments of two peers falls between 0.09 and 0.5 (Starbuck, 2005).⁴ The correlation is higher for papers rejected than for papers accepted (Cichetti, 1991). This means that peer reviewers are better able to identify academic low performers; that is, it is easier to identify papers that do not meet minimum quality standards than those of high performers and those that are a result of excellent research (Moed, 2007). The reliability thus is particularly low with regard to the opinion of peers among published papers in top journals (Lindsey, 1991).

**Low prognostic quality**

The reviewers’ rating of manuscript quality is found to correlate only 0.24 with later citations (Gottfredson, 1978). According to Starbuck (2006: 83–84), the correlation of a particular reviewer’s evaluation with the actual quality as measured by later citations of the manuscript reviewed is between 0.25 and 0.3; this correlation rarely rises above 0.37. Although there is evidence that higher prestige journals publish more high-value articles (Judge, Cable, Colbert, & Rynes, 2007), there is much randomness in editorial selections
(Starbuck 2005). As a result, one editor even advises rejected authors to “Just Try, Try Again” (Durso 1997). However, this strategy overburdens reviewers and tends to lower the quality of reviews. For example, reviewers have neither enough time nor the incentive to check the quality of the data and of the statistical methods employed, as some striking examples in economics demonstrate (Hamermesh 2007).

**Reviewers’ biases**

Many rejections in highly ranked journals are documented, even regarding papers that later were awarded high prizes, including the Nobel Prize (Campanario, 1996; Gans & Shepherd, 1994; Horrobin, 1996; Lawrence, 2003). Reviewers find methodological shortcomings in 71 percent of papers contradicting the mainstream, compared to only 25 percent of papers supporting the mainstream (Mahoney, 1977).

**Findings on Bibliometrics**

**Advantages of bibliometrics**

As a reaction to the criticism of qualitative peer reviewing, bibliometric methods, that is, rankings and ratings based on the number of publications, citations, and impact factors have become more prominent. This procedure is expected to produce several advantages over qualitative peer reviews (e.g., Abramo et al., 2009).

First, it is more objective because it is based on more than the three or four evaluations typical for qualitative approaches. Although it is based on qualitative peer
reviews because the articles counted must have passed peer evaluation, there may occur a balance of reviewers’ biases by the aggregation of many reviewers’ evaluations by scientific statistical methods (Weingart, 2005).

Second, the influence of the old boys’ network may be avoided. An instrument is provided to dismantle unfounded claims to fame. Rankings can serve as fruitful, exogenous shocks to some schools and make them care more about the reactions of the public (Khurana, 2007: 337).

Third, it is cheaper than qualitative reviews, at least in terms of time. It admits updates and rapid intertemporal comparisons.

Fourth, outsiders to the scientific community, for example, politicians, administrators, journalists, and students, may get a transparent and easy to comprehend picture of scholarly activity. The evaluation process is externalized and has been said to have unlocked the “secrets of the world of research” (Weingart, 2005: 119). In particular, politicians and deans consider rankings an objective measure to allocate resources and to provide compensation packages (e.g., Worrell, 2009). Scholars themselves use them to assess the research quality of their peers.

However, in recent times, the disadvantages of bibliometric methods have been hotly discussed (Adler & Harzing, 2009; Adler et al., 2008; Butler, 2007; Donovan, 2007b; Weingart, 2005). There are three groups of problems. Until now, mainly technical
and methodological problems were highlighted (van Raan, 2005). The third group, the dysfunctional reaction of scholars and institutions has been discussed less.

**Technical problems**

Technical problems consist of errors in the citing-cited matching process, leading to a loss of citations to a specific publication. First, is estimated that this loss amounts on average to 7 percent of the citations. In specific situations, this percentage may even be as high as 30 percent (Moed, 2002). Second, there are many errors made in attributing publications and citations to the source, for example, institutes, departments, or universities. In the popular ranking of the Shanghai Jiao Tong University, these errors led to differences of possibly 5 to 10 positions in the European list and about 25 to 50 positions in the world list (Moed, 2002). The most important impact factor, Thomson’s ISI Web of Science, is accused of having many faults (Monastersky, 2005; Taylor, Perakakis, & Trachana, 2008). It is unlikely that the errors are distributed equally. Kotiaho, Tomkin, & Simmons (1999) find that names from unfamiliar languages lead to a geographical bias against non-English speaking countries. Third, it has been shown that small changes in measurement techniques and classifications can have large effects on the position in rankings (Ursprung & Zimmer, 2006).

**Methodological problems**
Methodological problems of constructing meaningful and consistent indices to measure scientific output recently have been widely discussed (Adler & Harzing, 2009; Adler et al., 2008; Lawrence, 2002, 2003; Frey, 2003; Frey, forthcoming). Therefore, we briefly mention the main problems discussed in the literature.

First, there are selection problems. Only journal articles are selected for incorporation in the rankings, although books, proceedings or blogs contribute considerably to scholarly work. Other difficulties include the low representation of small research fields, non-English papers, regional journals, and journals from other disciplines even if they are highly ranked in their respective disciplines. Hence, collaboration across disciplinary boundaries is not furthered.

Second, citations can have a supportive or rejective meaning or merely a halo or herding effect. The probability of being cited is a function of previous citations according to the “Matthew effect” in science (Merton 1968). Simkin and Rowchowdhury (2005) estimate that, according to an analysis of misprints turning up repeatedly in citations, about 70–90 percent of scientific citations are copied from the list of references used in other papers; that is, 70–90 percent of the papers cited have not been read. Consequently, incorrect citations are endemic. They are promoted by the increasing use of meta-analysis, which generally does not distinguish between high and low quality analyses (Todd & Ladle, 2008). In addition, citations may reflect fleeting references to fashionable “hot topics.”
Third, using the impact factor of a journal as a proxy for the quality of a single article leads to substantial misclassification. Singh, Haddad, & Chow (2007) and Starbuck (2005) found that in management research many top articles are published in non-top journals, and many articles in top journals generate very few citations (see for economics Laband & Tollison, 2003; Oswald, 2007; for the journal Nature Campbell, 2008). A study of the “International Mathematical Union” even concludes that the use of impact factors can be “breathtakingly naïve” (Adler et al., 2008: 14) because it leads to large error probabilities.

Fourth, there are difficulties comparing citations and impact factors between disciplines and even between subdisciplines (Bornman, Mutz, Neuhaus, & Daniel, 2008).

However, even if these technical and methodological problems could be resolved, there are problems caused by the dysfunctional reactions of scholars and institutions. Even more and better indicators could not overcome these kinds of problems (Osterloh & Frey, 2009). They will occur even if the bibliometric system were to work perfectly.

**Individual dysfunctional reactions.**

The dysfunctional reactions of individual scholars consist of goal displacement and counterstrategies to “beat the system.” Goal displacement (Perrin, 1998) means that people maximize indicators that are easy to measure and disregard features that are hard to measure. This problem is also discussed as the multiple-tasking effect (Holmstrom &
Milgrom, 1991; Ethiraj & Levinthal, 2009). There is much evidence of this effect in laboratory experiments (Gilliland & Landis, 1992; Ordonez, Schweitzer, Galinsky, & Bazerman, 2009; Schweizer, Ordonez, & Douma, 2004; Staw & Boettger, 1990). For example, Fehr and Schmidt (2004) show that output-dependent financial incentives lead to the neglect of non-contractible tasks. This problem is avoided when principals are offered a fixed wage.

Empirical field evidence of goal displacement in academia is the “slicing strategy” (Weingart, 2005: 125) whereby scholars divide their research results into a “least publishable unit” by breaking them into as many papers as possible. This reaction is amplified when funding is dependent on the quantity of published papers. This was demonstrated in a study for Australia (Butler, 2003). The mid-1990s saw a linking of the number of peer-reviewed publications to the funding of universities and individual scholars. The number of publications increased dramatically, but the quality as measured by citations decreased. It could be argued that a remedy to this problem consists of resorting to citation counts. While this remedy overcomes some of the shortcomings of publication counts, it is subject to the technical and methodological problems mentioned.

Counterstrategies are more difficult to observe than goal displacement (Butler, 2007). They consist of altering research behavior itself in order to “beat the system” (Moed, 2007). Numerous examples can be found in educational evaluation (e.g., Haney,
The following behaviors are of special relevance in academia.

Scholars distort their results to please, or at least not to oppose, prospective referees. Bedeian (2003) finds evidence that no less than 25 percent of authors revised their manuscripts according to the suggestions of the referee although they knew that the change was incorrect. Frey (2003) calls this behavior “academic prostitution”.

Authors cite possible reviewers because the latter are prone to judge papers more favorably that approvingly cite their work, and these same reviewers tend to reject papers that threaten their previous work (Lawrence, 2003: 260). Authors willingly adapt to editors who pressure them to cite their respective journals in order to raise their impact rankings (Garfield, 1997; Smith, 1997; Monastersky, 2005).

To meet the expectations of their peers—many of whom consist of mainstream scholars—authors may be discouraged from conducting and submitting creative and unorthodox research. (Armstrong, 1997; Gillies, 2008; Horrobin 1996; Prichard & Wilmott 1997).

Institutional dysfunctional reactions

Dysfunctional reactions on the institutional level are manifold. Most importantly, the ranking system based on self-organized peer evaluation paradoxically results in an intensified control from outside the “republic of science,” in particular, by administrators
and politicians. “Managers are stealing power from scientists” (Lawrence, 2003: 259). Quantitative output indicators give politicians and administrators a handy instrument to manage academia from outside without knowing the process and content of research. However, managerial control theory suggests that such output control is inefficient when outputs are ambiguous and subject to change, and there are information asymmetries between the controller and the controlled. In these cases, only clan control is appropriate (Ouchi, 1977, 1980; Eisenhardt, 1985; Turner & Makhija, 2006), which is exerted by peers. Therefore, the intervention of administrators and politicians into academic self-governance produces unintended effects on the academic system.

First, in academia, output control systems handling research from outside the system create a lock-in effect. Even those scholars and academic institutions that are aware of the deficiencies of bibliometrics and the administrative interventions based on them do well not to oppose them. If they do, they are not only accused of being afraid of competition, but also of not wanting to increase the prestige and resources of their department or university. Therefore, it is a better strategy to follow the rules and to play the game. A self-enforcing cycle sets in. For example, in several countries, highly cited scientists are hired immediately before the evaluation of departments and programs are scheduled to take place in order to raise publication and citation records. Such “stars” are highly paid although they often have little involvement with the respective university. (Brook, 2003; Stephan, 2008).
Second, a negative walling-off effect sets in. Scholars themselves are inclined to apply output criteria to evaluate each other in order to gain more resources for their research group or department. In addition, it is easier to count the publications and citations of colleagues than to evaluate the content of their scholarly contributions. By doing this, scholars delegate their own judgment to the counting exercise behind bibliometrics, although, by using such metrics, they admit their incompetence in that subject (Browman & Stergiou 2008). This practice is defended by arguing that specialization in science has increased so much that even within disciplines it is impossible to evaluate the research in neighboring fields (Swanson, 2004; van Fleet, McWilliams, & Siegel, 2000). However, this practice in turn reinforces specialization and furthers a walling-off effect between disciplines and subdisciplines. By using output indicators instead of communicating on the contents, the knowledge in the various fields becomes increasingly disconnected. This widens the gap between theory and practice and hampers the ability to create radical innovations that often cross disciplinary borders (Amabile, Conti, Coon, Lazenby, & Herron, 1996; Dogan, 1999).

Third, research is increasingly homogenized. Research endeavors tend to lose the diversity that is necessary for a creative research environment. This consequence was pointed out for business schools by Gioia & Corley (2002). For economics, Great Britain provides an example: the share of heterodox, not strictly neoclassical economics sank drastically since the ranking of departments became based mainly on citation counts.
Heterodox journals have become less attractive for researchers due to their smaller impact factor when compared to mainstream journals (Lee, 2007; see also Holcombe, 2004).

Fourth, it is argued that a positional competition or a rent-seeking game takes place instead of an enhancement of research quality by the increased investment by universities and journals in evaluating research (Ehrenberg, 2000). It has been shown that the percentage of “dry holes” (i.e., articles in refereed journal which have never been cited) in economic research during 1974 to 1996 has remained constant (Laband & Tollison, 2003), though the resources to improve the screening of papers have risen substantially.

Despite the various criticisms, there seems to be a consensus that there is no alternative to bibliometrics and rankings as the main basis for research governance in academia. The theoretical basis for this system is not discussed. In particular, it is not questioned whether new public management and agency theory really do match the conditions of scientific work. Therefore, the dominant view emphasizes that a strong effort must be made to improve the present system (Adler & Harzing, 2009; Albers, 2009; Butler, 2007; Moed, 2007; Starbuck, 2009).
PSYCHOLOGICAL ECONOMICS AS THE THEORETICAL BASIS FOR ACADEMIC GOVERNANCE

In research on corporate governance in recent times, agency theory as the dominating theoretical frame of corporate governance has been questioned (Daily, Dalton, & Canella, 2003; Hillman & Daizel, 2003; Sundaramurthy & Lewis, 2003). The same is true for new public management (Bogh-Andersen, 2007; Giauque, 2003; Osterloh, Frey, & Homberg, 2008). Yet, in the present governance system of academic research, agency theory and its implications are still in place, and its theoretical foundations are not discussed.

To do this, we draw on the newly emerging field of psychological economics. It has emerged because of criticism about the assumptions of homo economicus, the standard economic model of human behavior (for reviews, see Fehr & Falk, 2002; Rabin, 1998). It seeks to reintroduce psychology into economics after standard economics had driven it out (Bruni & Sugden, 2007). Psychological economists investigate deviations from homo economicus in three main directions (Frey & Benz, 2004). First, individuals are boundedly rational. They are not able to maximize their expected utility. Second, individuals are boundedly self-interested. Depending on the circumstances, persons are not driven only by external rewards, but also by intrinsic motivation and prosocial preferences. Third, the utility concept of homo economicus is bounded. Psychological economists investigate subjective well-being or happiness as a measure for utility that
goes beyond financial income (Frey & Stutzer, 2002a, 2002b). In our analysis, we focus on the aspect of bounded self-interest, since this is arguably the most contested aspect of the agency paradigm.

Agency theory is based on the idea that individuals respond systematically in a self-interested way on extrinsic incentives from outside. This view disregards intrinsic preferences. People do not only react in an instrumental way, but they also act for their own sake or because of reasons lying within their own person. Examples are compliance with civic virtues, social or professional norms, self image (Ajzen, 1988), a flow experience in a fascinating activity (Cikszentmihalyi, 1975), or curiosity.

It is generally acknowledged that for academic research intrinsically motivated curiosity is of decisive importance (Amabile, 1996, 1998; Stephan, 1996). In standard economics and agency theory, these kinds of preferences are assumed to be a given and can be treated as constant. However, there exists considerable empirical evidence in psychology and psychological economics that this is not the case. Rather, there is a crowding-out effect of intrinsic motivation by externally imposed goals and incentives as well as by perceived unfair treatment provided that intrinsic motivation exists in the first place (Deci, Koestner, & Ryan, 1999; Falk & Kosfeld, 2006; Frey, 1992, 1997; Ordonez et al., 2009; for a survey of the empirical evidence, see Frey & Jegen, 2001). According to self-determination theory (Gagne & Deci, 2005), intrinsic motivation relies on two preconditions, autonomy and a supportive feedback helping to enhance ones
Crowding-out intrinsic motivation is explained by a reduction of autonomy and a controlling instead of a supportive feedback.

From the point of view of psychological economics, output oriented rankings have four disadvantages that cannot be managed by improving the present governance system in academic research based on agency theory. First, intrinsically motivated curiosity to do research tends to be crowded out and is in danger of being substituted by extrinsic motivation to score high in rankings. Content loses importance (Kruglansky, 1975). Autonomy can be reduced by quantitative output measurements, in particular, if they are linked to incentives. A supportive feedback is not provided by quantitative output measurements because in contrast to qualitative peer reviews they do not tell scholars how to improve their research.

Second, if intrinsic motivation is crowded out and extrinsic motivation prevails, the dysfunctional reactions of scholars like goal displacement and counterstrategies are enforced because they are not constrained by intrinsic preferences. The inducement to “game the system” in an instrumental way may get the upper hand.

Third, a negative self-selection effect takes place, in particular, when monetary rewards are linked to the position in rankings. According to Merton (1973), in academia, there exists a special incentive system called “taste for science”. It is characterized by a relatively low importance of monetary incentives and a high importance of peer recognition and autonomy. People are attracted to research for which, at the margin, the
autonomy to satisfy their curiosity and to gain peer recognition is more important than money. They value the possibility of following their own scientific goals more than financial rewards. These scholars are prepared to trade-off autonomy against money, as empirically documented by Stern (2004): scientists pay to be scientists. The preference for autonomy to choose their own goals is important for innovative research in two ways. It leads to a useful self-selection effect, and autonomy is the most important precondition for intrinsic motivation, which in turn is required for creative research (Amabile, 1998; Amabile et al., 1996; Mudambi, R., Mudambi, S., & Navarra, 2007).

Fourth, a negative self-fulfilling prophecy of agency theory sets in by institutional designs (incentive system, measurement practice, selection process), social norms (obeying the norm of self-interest not to appear as foolish or naïve), and language (evoking a gain frame instead of a community frame) (Ferraro et al., 2005). If intrinsic motivation is crowded out, only extrinsic rewards work—the assumption of agency theory has become true (Ghoshal, 2005; Ghoshal & Moran, 1996; Gibbons, 1998).

**IMPLICATIONS FOR THE IMPROVEMENT OF ACADEMIC RESEARCH GOVERNANCE**

The two approaches lead to different implications of how improvements of academic research governance can be achieved. Agency theory builds on ever more refined measurements to monitor and control academic researchers. In contrast, psychological
economics builds on careful selection and socialization to academic research, as well as on supporting the intrinsic motivation to undertake meaningful and creative research. These implications are discussed in turn.

**Implications from Agency Theory**

The proponents of the principal-agency theory for academia are well aware of some of its shortcomings. Three proposals are made to improve the present governance system in academia within the conventional paradigm, in particular, to improve rankings as the backbone of this system.

First, a temporary moratorium of rankings is suggested “until more valid and reliable ways to assess scholarly contributions can be developed” (Adler & Harzing, 2009: 72). As is the case for most authors, they believe that the identification of particular shortcomings should serve as a stepping stone to develop a more reliable research evaluation system (see also Abramo et al., 2009; Starbuck, 2009).

Second, it is suggested that bibliometric indicators should not be used as ready-to-go indicators lacking the competence to understand what is being measured (van Raan, 2005). Therefore, standards of good practice for the analysis, interpretation, and presentation of bibliometric data should be developed and adhered to when assessing research performance. This needs a lot of expertise (Bornmann et al., 2008), which
constrains considerably the responsible use of rankings as a handy instrument for politicians and journalists to assess academic performance.

Third, a combination of qualitative peer reviews and bibliometrics, so-called informed peer reviews, should be applied. It is argued that they can balance the advantages and disadvantages of these two methods (Butler, 2007; Moed, 2007; Weingart, 2005).

While the three proposals may help to avoid some of the disadvantages of bibliometrics, they cannot avoid or balance strategic reactions in the form of goal displacement and counterstrategies of scholars and institutions. This applies even if qualitative and quantitative measures worked perfectly (Osterloh & Frey, 2009).

**Implications from Psychological Economics**

The application of psychological economics to the governance of academic research is in its infancy; it is therefore only possible to outline some implications in need of more theoretical and empirical analysis.

From the point of view of psychological economics, the following aspects are to be considered. Intrinsic motivation is necessary for academic research but it is undermined by rankings because they curtail autonomy and give no supportive feedback. According to the “taste of science” (Merton, 1973) extrinsic motivation mainly in the
form of peer recognition is important. Monetary compensation plays a role, though a secondary one. Two implications follow, which should be further analyzed.

First, instead of treating scholars as agents who have to be monitored permanently, it should be considered whether it is more appropriate to carefully socialize and select aspiring scholars in order to downplay the controlling role of peer reviews and rankings. The main idea is to find out whether scholars master the state of the art and are creative and intrinsically motivated for research—and then trust that he or she will indeed perform well. This approach to research governance was emphasized by the famous President of Harvard University James Bryan Conant (Renn, 2002):¹¹ “There is only one proved method of assisting the advancement of pure science—that is picking men of genius, backing them heavily, and leaving them to direct themselves.” This view is still part of the “Principles Governing Research at Harvard,” which states:¹² “The primary means for controlling the quality of scholarly activities of this Faculty is through the rigorous academic standards applied in selection of its members.” Such a system may lead a limited number of researchers, after having received tenure, to misuse their autonomy. However, it may be the price that has to be paid for creative research to flourish.

Though autonomy is taken to be essential in this approach, it still requires to some extent informed peer reviews in spite of their deficiencies. This applies during restricted periods, for example, the selection and socialization process and whenever scholars apply
to a new position or for a grant, or submit a paper. However, there is a great difference between being under pressure to publish permanently and being submitted to control during a certain phase, knowing that once this phase is over one will enjoy a wide range of autonomy. If the pressure to publish is low, peer reviews change their role. They can be perceived as supportive instead of controlling and thus will further intrinsic motivation instead of undermining it.

Such governance principles also are employed in other professions characterized by a low degree of observable outputs, such as in the life-tenured American judiciary (Posner, forthcoming). These ideas are in accordance with empirical findings in psychological economics. They show that intrinsically motivated people do not shirk when they are given autonomy (Frey, 1992; Gneezy & Rustichini, 2000; Fong & Tosi, 2007). Instead, they raise their efforts when they perceive that they are trusted (Falk & Kosfeld 2006). This is of decisive importance for knowledge work (Kogut & Zander, 1996; Osterloh & Frey, 2000).

Second, since researchers also are motivated extrinsically, awards may serve as an externally mediated recognition (Frey, 2007; Frey & Neckermann, 2008). In contrast to variable pay for performance, awards are not perceived as controlling. Instead, they are of a symbolic nature that gives supportive feedback. Empirical research suggests that symbolic rewards do not crowd out intrinsic motivation (Heckhausen, 1991). In addition,
criteria for awards are usually not clearly specified ex-ante and thus provide considerably lower incentives for goal displacement.

As already mentioned, psychological economics is in its infancy and needs to be further developed theoretically and empirically. Systematic applications to issues of research governance in academia to our knowledge have not been undertaken. For example, little is known to what relative extent the “taste for science” (Merton, 1973) in different stages of a scholar’s career contains intrinsic elements, the desire for peer recognition, and monetary interests. Another open issue is the implications for the allocation of resources for research. Gillies (2008) suggests that each research unit that has passed the rigorous selection processes should be allocated basic funds sufficient to do meaningful research. Horrobin (1996) argues that the present concentration of resources to huge “centers of excellence” or “research empires” only rarely achieves more than would be possible had the same funds been distributed to small research units. This is in accordance with the considerations that giving more and more resources to a few “research empires” may hinder outsiders from participating in the resource allocation (Burris, 2004; Viner, Powell, & Green, 2004) and cause a decreasing marginal effect of additional research resources. While there exists some empirical work in this regard (Etzkowitz & Leydesdorff, 2000; Jansen, Wald, Frenke, Schmoch, & Schubert, 2007), this issue must be further elaborated.
CONCLUSION

Our paper contributes to governance research, in particular, to the neglected field of governance of academic research. The theoretical foundations of the present dominant view of research governance have rarely been seriously analyzed. Implicitly or explicitly, agency theory and its application in the form of new public management are taken for granted as a theoretical basis, as has long been the case in corporate governance. This approach relies on monitoring and sanctioning to govern agents’ behavior. We identify the major shortcomings of the present research governance in academia, in particular, peer reviews and rankings, and confront this view with an alternative approach. We suggest that the new, emerging psychological economics presents a fruitful avenue for research governance. In contrast to agency theory, psychological economics extends the motivational foundations beyond extrinsic preferences. In line with an understanding of scholarly research as a mainly curiosity driven endeavor, we include intrinsic motivation as a major determinant of scholarly behavior. In addition, following Merton (1973), the recognition by peers and supportive feedback is an important part of the motivational bundle of researchers that he aptly calls “taste for science.”

We also confront the different implications of the two approaches. Agency theory counts on the refinements of indicators and the measurement process. While these refinements may help to improve research governance, they cannot avoid the strategic
reactions of scholars and institutions. They also lead to a negative lock-in effect and a self-fulfilling prophecy. In contrast, psychological economics counts on a broader theoretical foundation of behavior including scholars’ intrinsically motivated curiosity, as well as the desire for peer recognition. In accordance with managerial control theory based on the work of Ouchi (1977, 1979), this alternative approach emphasizes selection and socialization of scholars and symbolic benefits in order to downplay the impact of rankings.

We believe the theoretical ideas presented here provide a useful foundation for future research in a number of areas. In particular research governance should be extended to a general academic governance including teaching which we have not dealt with in this paper. Future research in academic governance could also link up to the discourse on professionalization recently directed to managers (Khurana 2007). It may be that agency theory has contributed to a de-professionalization of the scholarly community, possibly leading to an erosion of professional codes of ethics. In view of the experiences with the recent behavior of some managers, the possibility of such a development is an issue of concern.
REFERENCES


Adler, R., Ewing, J., & Taylor, P. 2008. *Citation statistics, A report from the joint committee on quantitative assessment of research* (IMU, ICIAM, IMS). A report from the International Mathematical Union (IMU) in cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics (IMS).


Browman, H. I., & Stergiou, K. I. 2008. Factors and indices are one thing, deciding who is scholarly, why they are scholarly, and the relative value of their scholarship is something else entirely. *Ethics in Science and Environmental Politics*, 8: 1–3.


Washington DC: National Science Foundation.


Posner, R.A. Forthcoming. From the new institutional economics to organization economics: Four applications to corporate governance, government agencies, and legal institutions. *Journal of Institutional Economics.*


ENDNOTES

1 We prefer the expression “psychological economics” rather than “behavioral economics” for two reasons. First, economists had already examined human behavior before this new field emerged. Second, Simon (1985) points out that the term „behavioral” is misleading since it may be confounded with the „behaviorist” approach in psychology.

2 Patents should fulfil the task of transforming the public good “discovery” into a private good and at the same time to communicate the discovery. Patents on the one hand provide an incentive to invest in innovations by a legally enforced temporal monopoly, and on the other hand force to disclose the patent specification. However, it is questionable whether they really fulfil this task. Many discoveries are not patentable or the cost of disclosing are greater than the gains attainable form patenting. Moreover, there is an extensive discussion that today patenting in some fields, in particular university patenting, might have negative impact on the rates of innovation, see Nelson (2004, 2006); Dosi, Marengo & Pasquali (2006).

3 See also the special issue of Science and Public Policy (2007) and the Special Theme Section on “The use and misuse of bibliometric indices in evaluation scholarly performance” of Ethics in Science and Environmental Politics, 8, June 2008.

4 The most discussed study of peer reviewing was conducted by Peters and Ceci (1982). They resubmitted 12 articles to the top-tier journals that had published them only 18 to 32 months earlier, giving the articles fictitious authors at obscure institutions. Only three out of 38 editors and reviewers recognized that the articles had already been published. From the remaining nine articles, eight were rejected.

5 For example the British government decided to replace its Research Assessment Exercise based on qualitative evaluations with a system based on bibliometrics. Interestingly, the Australian
Government, which has used mostly bibliometrics in the past, plans in the future to introduce qualitative peer review methods (Donovan, 2007a).

6 Locke and Latham (2009) in a rejoinder provide counterevidence to Ordonez et al. (2009). However, they disregard that goal setting may well work for simple but not for complex tasks within an organization. For the latter case, see Earley, Connolly, & Ekegren (1989) and Ethiraj & Levinthal (2009).

7 Such problems of sabotage in tournaments have been extensively discussed in personnel economics, see Lazear & Shaw (2007).

8 “Pay for performance” which has become scrutinized in new public management e.g. by Perry 2006; Osterloh, Frey & Homberg (2009) as well as in the for-profit management field, e.g., Osterloh & Frey (2004) in academia in some countries like UK and Australia prevail or has recently been introduced like in Germany or Austria.

9 The crowding-out effect does not always takes place, e.g. Gerhard & Rynes (2003); Locke & Latham (2009), or is contested, e.g. Eisenberger & Cameron (1996). However the empirical evidence for complex tasks and actors intrinsically motivated in the first place is strong, see Deci, Koestner and Ryan (1999); Weibel, Rost & Osterloh (2009).

10 A third precondition is social relatedness, see Gagne & Deci (2005).


12 See http://www.fas.harvard.edu/research/greybook/principles.html.