Making Lists, Enlisting Scientists:
The Bibliometric Indicator, Uncertainty and Emergent Agency

Casper Bruun Jensen

The question of how to measure research quality recently gained prominence in the context of Danish research policy, as part of implementing a new model for the allocating of funds to universities. The measurement device took the form of a bibliometric indicator. Analyzing the making of the indicator, the paper engages the literature on social studies of quantification and classification. The analysis proceeds from the inside out, through description of the organizational processes and classificatory disputes through which the indicator was developed. It addresses questions such as: How was the indicator conceptualised? How were notions of scientific knowledge and collaboration inscribed and challenged in the process? The analysis shows a two-sided process in which scientists become engaged in making lists but which is simultaneously a way for research policy to enlist scientists. In conclusion, the analysis offers suggestions for a reorientations of the of study emergent quantification systems. [0]

Keywords: Bibliometric indicator, classification, controversy

The so-called Barcelona objective called on EU member states to increase investment in research to three percent of the GDP by 2010. In the context of Danish research policy, one result of this requirement was the making of a "globalization pool" that was to channel 39 billion DDK of additional funding into university research over a six year period starting in 2007. How would this happen? The model for allocating these funds, according to the university output measures, had four components: it was based on university teaching, attraction of external funds, production of PhDs, and, centrally for this paper – publication *quality*, which would therefore have to be *quantified*. This quantification would be facilitated by means of a *bibliometric indicator*. However, this indicator did not exist, it had to be made. To accomplish this feat of quantification, over a few years, hundreds of Danish researchers joined 68 disciplinary groups to make lists that classified and ranked thousands of journals.

The Social and the Quantitative

An abundant literature from the history and sociology of science and from the evaluation research makes the point that there are tight interrelations between policy and quantification. Sociologists Wendy Espeland and Michael Sauder note that Max Weber saw the "peculiarity of modern cul-

ture" as closely related to "calculability" (Espeland & Sauder, 2007: 4). Philosopher Ian Hacking has described the "avalanche of numbers" proliferating through modern society (Hacking, 1982). And historian of science Theodore M. Porter has elucidated the processes through which *Trust in Numbers* (Porter, 1995) is established.

A growing literature has also focused on quantification within research and higher education. Behavioural economists Margit Osterloh and Bruno S. Frey note that "academic rankings are generally considered the backbone of research governance" (Osterloh & Frey, 2010: 2). They link this development with the adoption of new public management in universities. New public management is enabled through the collection of "huge amounts of quantitative information" which is used to "produce critical numbers in order to compare and evaluate different activities" (Hansson, 2006: 162). This "proliferation of quantitative performance measures" (Espeland & Sauder, 2007: 5) has been affiliated with the term audit society since the seminal works of Michael Power (e.g. 1997).

Focusing on the organizational consequences of rankings, Espeland and Sauder (2007) proposed a framework for "investigating the reactivity of social measures". Reactivity is at stake in the many situations where people "change their behaviour in reaction to being evaluated" (Espeland & Sauder, 2007: 1; cf. Hacking, 1995). Sandra van Thiel and Frans L. Leeuw (2002) make a similar argument, focusing on what they term "performance paradoxes".

Quantification, as well, *discipline* the actions of subjects and institutions, although this rarely happens without resistance. Institutional theory has argued that resistance often takes the form of "buffering" and "decoupling", which means that organizational actors respond by "symbolically" going along with the demands of quantification while refraining from changing "substantively" (Sauder & Espeland, 2009: 64, cf. van Thiel & Leeuw, 2002: 271).

Quantification Inside-Out

What is shared among these studies is that they focus on the macro level, which is defined as societal, institutional or political and that their temporal orientation is largely retrospective.

Another possibility, however, is enabled by the actor-network theory injunction to enter controversies, technologies or facts, before they are stabilized (cf. Latour, 1987). This entails following quantification measures *through their process of emergence*. Such studies move from the *inside* of quantification systems and *outwards* rather than the reverse; they study quantification in the making rather than ready-made. In the context of the making of the bibliometric indicator, I use the terminology of inside and outside, because it allows me to make a specific analytical distinction.

The central point is that critical organization research generally begins at a point in time where quantitative measurement devices are *in place*. Critical analysis is enabled by this analytical orientation, which is thus both *retrospective* and *detached* from the process of establishing the quantitative system. The alternative afforded by studying such processes *inside out* is not a celebratory or uncritical analysis. Rather the point is to enable a form of analysis more oriented towards the contingencies and complexities of quantification and, in particular, that is more attentive to the processes of *interactive modification between multiple kinds of actors* through which quantification measures are constructed.

Accordingly, the first aim of the paper is to characterize the organizational and

classificatory processes and *controversies*, through which the authority list of journals, that provided input for the bibliometric indicator, was made. How was the indicator conceptualised? How were notions of scientific knowledge and collaboration inscribed in the process of classifying and ranking journals? Who were the involved actors? Which disputes and controversies did these classificatory efforts engender? Only at the end will we be in a position to consider what broader institutional and political consequences might follow.

In dealing with these questions, the paper fuses two central areas within STS, namely, classification studies (e.g. Bowker and Star, 1999; Gieryn, 1999) and controversy studies (e.g. Latour, 1987). However, given that the focus is on a device developed to measure the quality of knowledge, the paper offers a controversy study at a *one level remove*: not of scientific classifications but of classifications of science.[1]

The analysis shows a two-sided process in which research policy makers attempt to *enlist scientists*, with the specific purpose of *making lists*. The notion of enlisting does not indicate a deterministic or linear process. Like Latour's (1987) enrolment, enlisting points to a far more precarious and ambivalent process. On the one hand, researchers were explicitly *enlisted* by the Danish Agency for Science, Technology and Innovation[2], as they joined groups that would make lists of ranked scientific journals. However, this enlistment was to a significant extent formal: many researchers joined the process, without fully accepting its premises, or, indeed, while actively continuing to challenge its content. Enlisting, then, is a complex and uncertain process: in the present case we shall see that it simultaneously failed and succeeded in achieving objectives. Such uncertainties are intimately connected to what I call the *emergent agency* of the authority list – and, consequently – of the bibliometric indicator.

Method

This study began in the late 2009, around the time the indicator was first made public. I conducted a number of preliminary semi-structured interviews with social science colleagues who had been representatives in a variety of the disciplinary groups, including public health, psychology, architecture and urban planning, and science studies and research policy. This provided a general understanding of the development process. Subsequently, I focused more specifically on the group that covered that science studies and research policy. This group, called group 68, was somewhat unusual in the sense that it was constituted specifically as an interdisciplinary group and also in that the head of the group, associate professor Claus Emmeche, turned out to be an outspoken critic of the lists and the indicator. Thus, the story of group 68 cannot be taken, synechdochically, for the whole. Still, this story facilitates analytical insights that go beyond the specific group. For one thing, multiple other groups, *including those that were assumed to be traditionally disciplinary*, in fact turned out to also have interdisciplinary characteristics. For another, much of the documentation that I use in the analysis was sent to all of the groups. Though group 68 provides the central illustrations in what follows, many or most of the issues I address were thus experienced in all of the groups, although the way they were handled, of course, varied.

I contacted all members of group 68 and interviewed three out of five; I had several conversations with two of the members. In connection with these interviews I received a voluminous set of documents that had been circulated to the research

groups by the Danish Agency for Science, Technology and Innovation. Additional documentation was acquired at the webpage of the agency.[3] Much relevant material was also available at a webpage dedicated to group 68.[4] Subsequently, I interviewed the contact person from the agency. I also conducted interviews and exchanged emails with central management at my own institution to get a sense of their views on the indicator. Additionally, I searched public media for information. The empirical material is thus composite, but it draws primarily on researchers' perspectives and mostly on the experiences of members from group 68.

Norwegian Lists into Danish Policy

Since it became clear in 2005-6 that the Danish government was going to infuse the universities with extra funds, the question of how to measure and quantify research quality gained prominence on the political agenda. There were initial propositions to use citation indices directly. However, the humanities and social science disciplines argued that exclusive reliance on impact factor would systematically bias the indicator.[5]

In light of such objections, Danish policy makers looked abroad. The idea of measuring research quality was not a new one. Countries including Australia, the U.K., and Norway already had experiences with different models. The English and Norwegian models were scrutinized with particular interest because they were seen to represent different ideal types. The English RAE model worked with a mix of quantitative indicators and a qualitative evaluation based on an institutional peer-review process. The RAE model, however, was already under pressure from metrics-based evaluation systems (cf. Barker, 2007). The model faced additional difficulties in the Danish context because the research environment is much smaller than in the U.K. A RAE-like model would require small groups of researchers, who are also competitors, to repeatedly evaluate each other, while using a large amount of international evaluators was seen as prohibitively costly.

In contrast, the Norwegian model suspended with qualitative peer-review and was based solely on listing and ranking journals. It worked from a large number of smaller *disciplinary lists* that classified all journals as normal or excellent. These lists were integrated into one comprehensive *authority list* that ranked all scientific journals. Translated into the bibliometric indicator, each publication would get a number of points depending on its excellence; these points would be used to calculate distribution of funds. However, although it was quantitative, the Norwegian indicator was perceived as relatively fair, because it did not rely exclusively on impact factor. Thus, different contextual and disciplinary factors might be used to determine the level of excellence of a journal (cf. Schneider, 2009).

In 2006, the political momentum was swinging towards using the Norwegian model. By 2007 the Norwegian indicator was about to be adopted as part of what the Ministry of Science, Technology and Innovation saw as a routine procedure. However, this generated considerable consternation from different universities. Specifically, Danish Universities,[6] a stakeholder organization that promotes university collaboration, opposed the implementation of the Norwegian model. Thus it became evident to policy makers that for the quality measurement system to have legitimacy in the university environment, the Norwegian lists would need Danish input.

What accounted for the opposition to the Norwegian authority list? One discussion related to the implications of having a

single list of journals. Given the differences in how excellence is evaluated in different disciplines, why opt for constructing one general list? Why not have three or four lists, corresponding to the major faculties? However, the humanities in particular were worried by this idea. Their concern was that a model that was based on differences between *general areas* might later be used as a basis for creating hierarchy among entire faculties.

In addition, Danish researchers also found fault with the content of the Norwegian authority list. It had not been validated with enough care: non-peer reviewed journals had snuck in, it was said, and one could find what an interviewee described as "the research equivalent of Donald Duck". Further, the Norwegian indicator was seen as biased towards traditional disciplines. Newer disciplines and especially interdisciplinary research was not well represented. Basically, the Norwegian indicator had parcelled out the entire field of scientific knowledge into three different groups (natural, human and social science) (cf. Schneider, 2009: 371-2). Since interdisciplinary research, such as, for example, science and technology studies, was not represented in the disciplinary structure, it would be little surprise if few STS journals would be ranked as excellent.

These criticisms led to the decision to make a Danish version of the Norwegian authority list. This entailed creating a process for forming *groups of researchers* that would do this classification work. *Danish Universities* created a structure that comprised nineteen humanities groups, eight social science groups, twenty natural and technical science groups, fifteen health science groups, and six interdisciplinary groups, including group 68, which would represent science studies and research policy.

Excellent Representatives

Who were the researchers who would join the groups and do the classifying? *Danish Universities* contacted the universities to recruit potential members. These members were required to be *excellent representatives* of their fields. At some universities, management nominated researchers were seen as particularly suited. Other institutions involved librarians in analysing publication patterns, to ensure that researchers would be represented in the most locally important groups. In other cases researchers were recommended if they expressed personal interest. One consequence of these various principles of selection was that the extent to which the groups were able to represent and evaluate *all relevant knowledge* in their fields was unclear from the outset. Yet, based on the recommendations, the groups were established as a *gentleman's agreement*, according to which all the nominees would be included but no university could have more than one member in a group

The outcome of this process was the creation of 68 groups consisting of five to twelve members. They would communicate with the Danish Agency for Science, Technology and Innovation[7], part of the ministry, which, in turn, created an organizational structure comprising a disciplinary and a technical committee.

Presumptions of List-Making

> A *list* is a record of things or abstract statements that have been removed from their context and written down one after another as *facts*. The classification system and selection principle according to which the facts are chosen is not included in the list itself. Usually, the principle used in ordering the facts in the list is linked neither to the origi-

> nal context of the facts nor to their selection principle, but instead to the logic of the list itself. (Rottenburg, 2009: 137.)

In January 2008, 68 groups had been composed. Together they had to create a set of lists of journals that would be combined in a Danish authority list. These journals would then get bibliometric points depending on how they were classified. These points would be used to distribute funds to the Danish universities. Such was the basic (though not yet explicated) logic of the bibliometric indicator. The authority list to be made was thus a crucial actor in this set-up.

But what precisely was the task of the groups? A guideline from the agency explained that each group would make two lists of "publication channels", one for journals and one for book publishers.[8] The point, then, was that "organizations only get points for publications in journals or with publishers that are on the lists. It is thus very important that the channels through which Danish researchers publish are listed if they live up to the requirements of peer-review as defined by the disciplinary committee".

The guideline further specified two tasks. The first was to make suggestions for the "publication channels" that should be represented on the lists made by the groups. Second, these channels would have to be classified as level one or two, according to their "*excellence within the disciplinary area represented by the group*".

The literal mandate of the disciplinary groups was thus to *make lists*. Although this task was articulated as a matter of creating representational devices that would render research quality amenable to quantification, the following analysis suggests that the task involved the *creation of a new actor*: namely, a general authority list that allowed for the transformation of journal publications into points in the bibliometric indicator.

Although I have just characterized the list as an emerging actor, it was not in fact made *de novo*. As a point of departure, each group received a number of excel sheets comprising *other lists* from which they would compose their own. The most important were the Norwegian lists. These were categorized into the broad areas of "humanities", "social science", "natural/technical sciences" and "medicine/health". Additionally, groups were sent assistance lists. One was the NSI (National Science Indicators) list. The humanities groups also received a draft list from the European Reference Index for Humanities (ERIH). Multiple lists at hand, the task of each group was to add and remove journals, thereby creating their own. The deadline was set for mid-March 2008.

Now, if the mandate of the groups was to make lists, the guideline also defined *how* they should do so. On the one hand, it described the required work of listing and classifying; on the other hand, it specified a particular frame for *enlisting researchers*. Here I pause to briefly consider the presumptions that were inscribed into the organizational process for making these lists. Each of these presumptions can be read from the ways in which the groups were put together and from how the relations between groups were imagined.

First, in spite of the somewhat haphazard way in which the groups had been put together, the group members were literally expected to be *excellent representatives* of their area and *therefore* capable of surveying and evaluating all relevant knowledge within it. Second, the *collegiality of scientists* was also taken for granted: groups were assumed to be able and willing to create *comprehensive and fair* lists, rather than privileging their own fields of specialization. Third, the organizational set-

up assumed that groups would be able to *reach agreement* (both within each group and among the groups) about the quality of journals, since, unable to reach agreement, the groups would be unable to make the lists.

Aside from these rather explicit presumptions, I suggest, the making of authority lists embodied a discernible, though more oblique, idea about *scientific knowledge* itself. This idea can be called the *puzzle-metaphor* of scientific knowledge. It had two aspects. First, that the collection of scientific knowledge could be *distributed in a non-overlapping way among groups*, each of which, second, would *guarantee excellence within its own domain*. Thus, each group would represent *one piece* of the jigsaw puzzle of scientific knowledge, and the combined set of groups would represent *the entire picture* of scientific knowledge.

Each of these presumptions would be challenged during the complex process of list making. It is thus not the case that these inscribed presumptions determined all that followed. Rather, the case shows a process of reciprocal transformation, in which diverse actors *engaged with* these presumptions, both by going along with them and by challenging them. The *emergent agency* of the list is the (still uncertain) result of this combined set of engagements and challenges.

Boundary Conditions

> A number of central questions related to scientific quality arise: What constitutes quality and how is quality measured? Who decides, and what is the role of the scientific community and its organization into disciplines? (Hansson, 2006: 168.)

Before setting to work, groups still had to settle several issues. They included questions relating to group identity, representativeness and the classification of knowledge. For many groups, including group 68 but also, for example, group 10 (media and communication, comprising e.g. journalism, film science, rhetoric), or group 28 (geo-sciences and climate, comprising e.g. geophysics, geo-information, meteorology, polar research) these were difficult questions because they covered multiple, divergent disciplines.

Consider the case of group 68. As noted, this group was in charge of the interdisciplinary area of science studies and research analysis. Since it could conceivably find relevant journals in all major areas, the five members of the group needed to inspect all of the assistance lists. This work-intensive endeavour raised both identity, and classification issues, requiring members to ask questions such as "who are we?" as a group (and thus, which areas are we *able to evaluate*?); and "who are we supposed to represent?" as a field (and thus, which areas are we *supposed to cover*?). Such questions established the boundary conditions for the list-making activities (cf. Gieryn, 1999).

According to members, research policy and "clear-cut" science and technology studies obviously belonged to group 68. However, since group 13 dealt with philosophy, including "of science", such journals were excluded. The history of science, on the other hand, was seen central to group 68 (unsurprisingly, since three group members were affiliated with history of science broadly conceived).

The initial steps of making the list were not marked by significant controversy. For group 68, as apparently for many other groups, the ideal was to make a very open and comprehensive list. No one had an interest in delimiting the publication opportunities of colleagues by excluding

journals. That this logic was in fact *both* sensible and deeply problematic became evident only later.

When the groups reported back to the committee a few months later, the list of group 68 contained 426 journals. In the cover letter, the head of group 68 wrote that "the group has been assigned a troublesome and in principle an undecideable task. The group agreed to delimit the task so that sociology of knowledge, history and science studies falls under 68, whereas epistemology and philosophy of science belongs to philosophy." Yet, it had been "very difficult to limit the number of journals", given that many publish across the defined areas.

Classification Criteria

> The quantification of qualities is as much an administrative accomplishment as an intellectual one (Porter, 1994: 404).

Having created a list *in principle* comprising all relevant journals, the next task was to classify them as either level one or level two. Level two journals were defined as those "broadly considered to be absolutely leading in the discipline" and as "publishing the most important papers from researchers in different countries". The importance of this ranking was now also made explicit: translated into the bibliometric indicator, publication in level two journals would give three points against only one point for publication in a level one journal. Thus, the indicator quantified the difference in excellence according to a 3:1 ratio. This ratio would influence how many points a university would receive, and thus the allocation of funds. The ratio also introduced a particular kind of external, political incentive to publish in precisely the journals that the authority list defined as highest quality. This aimed to counter the possibility that researchers would churn out many mediocre (level 1) articles, rather than publish fewer excellent (level 2) ones.

Still, while it is easy to imagine that deciding what counts as "leading quality" might lead to classification disputes, such controversy could in principle be avoided. As noted, group 68 initially adopted a "live and let live" strategy in order to accommodate different specializations represented by the group. Concretely this meant that the group was initially quite liberal in assigning journals to the "excellent" level two. However, an additional criterion made this approach unviable.

This second criterion specified that "the combined set of level two journals must add up to a maximum of 20% of the worldwide production of scientific articles" within a disciplinary group's field. The requirement meant that a very significant portion of the list *had to* be categorized as the lower level one. In this situation, the previously adopted strategy of broad inclusiveness came to make unintended strategic sense, since enlarging the gross list of journals would enable more journals to be categorized as level two.

Another consequence related to the distribution of journals at different levels *within each of the lists*. The case of *Science* vividly exemplifies this point. It is, of course, widely accepted that *Science* is an excellent journal. But the central attribute of *Science* in the context of the evaluation criteria was not its recognized quality. It was rather that *Science* is published *often* and publishes *many* short pieces. If one relates these properties to a measurement based on world-production of articles it becomes clear that *Science takes up a lot of space*.

In the case of group 68, *Science* accounted for approximately 3.58% of the combined world-production (compare with e.g. *Social Studies of Science* which accounted for 0.15% or *Science Studies* that

covered 0.06%). Now if group 68 assigned *Science* to level two status this would mean that 3.58% out of 20% would be covered by this single journal. As a consequence, numerous smaller journals would have to be excluded (for example, 24 journals with a world-production equal to *Social Studies of Science* or approximately 60 *Science Studies*).

Consider, on the other hand, the possibility of including *Science* on the gross list but assigning it to the lower level one (arguing, for example, that it is *generally* relevant for the field but that it is not crucial because it is not *specialized*). Obviously, this would add significantly to the combined world-production from which the 20% would be taken, thus enabling the inclusion of several smaller journals. On the first set of submitted lists, groups consistently ranked *Science* as a level one journal.[9] This example shares characteristics with what Michéle Lamont has called "strategic voting" in the context of academic evaluation panels: the "practice of giving a lower rank than would otherwise be justified to some proposals, in order to increase the likelihood that other proposals will win" (Lamont, 2009: 122).

When group 68 submitted its initial ranked list, it included 585 journals of which 15.02% the world-production was represented as level 2. Explaining the decision making procedure, the head stated that the list excluded journals that were found "in this context less relevant". The head continued that "note that we do consider *Science*, *Acta Sociologica*, *Explorations in Economic History, Language* and *Nature* to be extremely important, just not for science studies. As you are aware there are 67 other groups dealing with the rest of the cosmos."[10]

Classification Controversies

> When the first bibliometric based evaluations of research institutions were carried out ... the reaction of the scientists concerned was predictable. They challenged the possibility of the enterprise on methodological grounds, and they threatened to take the analysts to court because they feared that the results would have adverse effects. (Weingart, 2005: 117-8.)

Some months later, the general authority list was made public. According to the head of group 68, this gave rise to "wild rage" in many groups. What had happened between October 2008 and February 2009?

At this point, each of the groups had made their own list, by choosing from a set of approximately 20.000 journals and classifying these journals as level one or two. Now, to start out the groups had not been told to consider the potential issue of journal duplication. Consequently, while each group defined its own list, it might contain journals that were also on the lists of other groups. An unproblematic situation at first, this turned into an administrative problem when the lists were returned, because the task of the committee was to integrate the input into one authority list. The committee now faced a serious issue: what to do with cases, in which a journal was classified as both level one and two – perhaps by several different groups?

In all likelihood the committee had underestimated the problem of duplication. In reality, however, around 4.000 journals were registered on more than one list. Thus, the committee drew the perhaps inevitable conclusion that journals had to be located exclusively in one group. However, due to pressure from the ministry that wanted the bibliometric to be usable as soon as possible, the committee

also made quick decisions about *where* and at *which level* to place duplicates. Without offering much detail, the committee explained that this had been a "substantial" task.

Nevertheless the result was the aforementioned "wild rage". In the case of group 68 more than a third of the original level two journals had been downgraded; the head characterized the returned list as "massacred". Meanwhile other journals had received upgrades that contradicted group preferences. For group 68, the damaging consequences of these upgrades quickly became clear. For one thing, *Proceedings of the National Academy of Sciences* and *Nature* had been upgraded. Since they accounted for 16% of the world-production of journals on the decimated list, this enforced reclassification would necessitate the downgrading of approximately 30 other journals. While this situation re-opened the question of quality measures, it also highlighted the *variety of quality criteria in different groups*.

Thus, while group 68 was keen to get rid of *PNAS* and *Nature, in spite of* their high-impact, other groups complained that impact factors were *neglected*. In a letter to the committee, the head of group 34 explains how he was contacted "by a highly esteemed plant biologist," who asked him to account for the criteria used to classify journals in his field. The head explained that the biotech group had used an ISI impact factor of 4.2 as a cut-off point for admission to level two. The plant biologist asked why five out of the ten highest ranked journals in the field were then level one. The head "could not think of an answer". Another biologist added that: "science measures as exactly as possible. That means objective numbers collected according to described criteria (such as ISI impact factor) can be used, whereas intuitions cannot". Paradoxically, it seemed that the process had managed to alienate both humanists and natural scientists.

The head of group 37 (transport and traffic research) as well made a number of critical remarks. He thought it made sense to refrain from relying exclusively on impact factor, so that "smaller sub-disciplines" could also be represented at level two. However, he noted that: "If only 50% of journals nominated as level 2 end up in the group you might as well have tossed a coin." He was particularly frustrated to see that *Nordic Transport Research,* which had been rejected by the transportation group, had re-appeared on the list, although it was "neither scientific, independent nor reviewed," consisting of "one-page abstracts where 2/3 of the space is taken by photos of roads".

As well, he pointed to another contentious issue. Journals such as the *Journal of Transport Economics and Policy* and *Networks and Spatial Economics* held relevance both for economists and for transport researchers; yet, they might not have the same *importance* for both. In a context of prevalent interdisciplinarity, the "in-principle undecideability" in assigning unequivocal quality levels to journals remained.

However, this was the case not only for "interdisciplinary groups". Medical researchers counted that immunology journals were spread out over ten groups, while group 62 (basic medical disciplines) contained more than ten disciplinary clusters, covering everything from physiology and genetics, to neuroscience and clinical chemistry.

Rather than clear boundaries, researchers thus found interdisciplinary clusters *distributed across groups* and *fragmented within groups* that were assumed to be "disciplinary". Compared with the presumption of internal coherence (in the groups) and external complementarity (among the

groups) these discussions give witness to a much more complicated ecology of scientific knowledge.

In any case, however, the agency had by now succeeded in making a list. Clearly it had also, in a practical sense, managed to enlist scientists in this process. It had, however, failed to effectively enlist their support for the outcome. Indeed, an increasing number were vocal in their opposition to the authority list.

Re-Enlisting Scientists, Re-Negotiating Journals

> Reactivity on the institutional level takes several forms. First, if rankings are used as measure to allocate resources and positions they create a lock-in effect. Even those scholars and academic institutions that are aware of the deficiencies of rankings do well not to oppose them. (Osterloh & Frey, 2010.)

In spite of the negative atmosphere, the heads of group were generally in agreement that "constructive criticism" rather than mutiny was the best way ahead. Given that the authority list would be made irrespective of how unhappy participants might be, compromise was required. Meanwhile the agency and disciplinary committee had not apparently anticipated the hostility and frustration generated by their handling of the duplication issue. They, too, were interested in reconciliation. Circulated emails and calls for meetings admitted room for improvement and recognized the need for adjustment.

At a general conciliatory meeting, high-ranked research administrators did their best to appease the critical heads of groups. They were met with overall criticism and presented with estimates of negative consequences. The meeting proceeded to discuss ways of negotiating disagreements between groups concerning the rights to specific journals that had been relocated. In effect, this entailed re-classifying journals for a third time.

During this phase, different groups requested a mixed bag of journals from group 68; it included *Daedalus*, *M3 – Man, Medium, Machine*, *History of Science, Gender, Technology and Development* and *Science and Society*. While some of these were given away more or less freely, others gave rise to renewed boundary considerations, as indicated in this mail from one group member:

> Under no circumstances can we accept to let *Research Policy* and *Scientometrics* disappear. They are completely central both for science studies and research analysis. I also think we should keep *Technology Analysis and Strategic Management* as well as *Technovation* but in those cases I am not quite as certain; perhaps we could accept giving up *Technovation.*

The business economy group, though, had identical preferences. Although it might also be willing to give up *Technovation,* it would under no circumstances lose *Research Policy*, an "all-important business economic journal". Arguing that "our list will look strange if we don't have the leading journal within innovation management", a negotiation ensued that, again, resemble aspects of what Lamont (2009) found in the context of research councils.

By "horse-trading", Lamont referred to actions that enable "the realization of other panelists' objectives, in the hope that they will reciprocate" (Lamont, 2009: 122). Indeed, the business economy group offered just such a deal. They would promise that *Research Policy* would be nominated to level two *if only* it could be included in their list. Faced with this offer, and the need to downgrade numerous journals, group 68 gave up the journal.

One member insisted that: "if the business economic list looks strange without it, ours looks even stranger." However, this disciplinary logic was outdone by what another referred to as the "logic of disciplinary egotism". In the end, it did not matter where the journal went, as long as it was certain it would be ranked as excellent.

Uncertain Measures

While jostling for journals, the groups also engaged the committee in a discussion about the rule that the world production of articles in level two journals could add up to no more than 20% of the world-production of journal articles on the list.

As part of the Norwegian consideration about how to establish quality measures, the 20% rule was originally articulated in a 2004 report that outlined the new system for documenting scientific publications. The report stated that "based on level, a number of publication channels will be chosen for a list representing level 2 and representing approximately a fifth of the publications within the field".[11] The issue was crucial, among other things, because the wrong measures might lead to "perverse incentives" (van Thiel and Leeuw, 2002: 271), such as slicing research into the least publishable unit.[12]

The Norwegian report suggested that incentives that would encourage higher quality publications should be introduced. After having dismissed various solutions as inadequate, the report defined its own two-tiered model as part of a discussion of different publication patterns in the different faculties. This is where the 20% rule appears:

> In group A (natural science) one will have a draft list where it can be ensured that level 2 will represent a fifth of world article production. In the two other groups one will have a more random basis where level 2 will certainly represent a fifth of the "world production" (within ISI) but where the journals are more dominated by the US (and are thus far from representing a fifth from a Norwegian or European perspective).

The statement is noteworthy for different reasons. First, the Norwegian report operated with the assumption that different faculties have different publication patterns. Second, what counts as a fifth is *relative* to national perspective and discipline; thus, if social sciences are more dominated by US publications than natural sciences, then the 20% rule makes less sense for Norwegian humanists than for e.g. physicists. The third crucial point, however, is that the basis of the 20% rule itself is not explained but simply stated.

Nevertheless the rule did not fall from the sky. Instead, it claims legitimacy by referencing what is known in bibliometrics as *Bradford's distribution*. Roughly stated, it implies that within a given field the central articles tend to concentrate in a limited number of journals (approximately 20 %), whereas the remaining less important articles will be distributed in a many (approximately 80 %). It is this distribution that reappears in the Norwegian and Danish guidelines. It does so, however, in the form of a *normative principle*. This translation from the empirical to the normative is not innocent. Because rather than working with an *actual* distribution of influential and less influential articles, the *rule* moves in the opposite direction: it defines a limit for what will be *allowed to count* as good quality at the journal level. Unsurprisingly this was consequential.

Estimating World-Production

When the lists were returned, most groups were shocked to see that their percentage of level 2 journals had increased drastically.

The list of group 68 was submitted with 15.02% of the world-production of articles at level two; when it came back the percentage had increased to 36.53%.

This was a general experience. In 55 of 67 groups the percentage of level 2 journals exceeded 20%. Gradually it became clear that these increases related to an *estimation* of the world-production of articles.

Now why would one need to estimate world-production? For ISI journals there were no problems because precise figures were available. Still, in most cases they were not. For thousands of journals the committee thus had to somehow determine the amount of published articles. Thus began a set of manual check-ups, which turned out to be far from flawless. In a graphic example, a theologian refers to the:

> *Journal of Ecclesiastical History*, which is a level 2 journal in our group. It is listed with 911 (!) articles over three years. Even though the number of church historians is quite considerable, such fabulous productivity seems remarkable.

The reason for this "fabulous productivity" could be found at the journal webpage, which stated that: "each volume includes about twenty articles and roughly three hundred notices of recently published books relevant to the interests of the journal's readers".[13] If one adds these notices, then "911 articles actually fit quite well".

In most cases, however, estimated production was *literally* estimated. Realizing what an enormous task it would be to check world production for each journal, level 2 journals were simply assumed to publish 40 articles per year.

In August 2009, multiple examples of error arising from the estimation of world-production were collected and sent to the committee by many heads of group. It was argued that the authority list should not be finalized until these errors had been corrected. The committee responded by again acknowledging the need for continued learning; indeed, it expressed satisfaction that the groups were interested in improving data. Thus, it reinterpreted what many researchers saw as fundamental flaws in the listγthat would automatically spill over into the indicator γas issues that could receive a technical fix.

Yet, although the committee admitted to have "found instances where the number of articles was misleading", it insisted that it would be impossible to check every journal. Instead, it proposed that groups mark: "cases in which you suspect there may be serious errors and which significantly impact the calculation of the level 2 percentage".

Even so, many researchers failed to see why they should correct the committee's errors; as well, they felt that they had already spent an inordinate amount of time on the list. Accordingly, a debate ensued once again among some heads of groups about whether to resign. But this proposal was again rejected as too radical. In the eventual compromise, groups with less than 50% ISI-registered journals would be allowed to downgrade based to a limit of 20% of journals rather than world production of articles.

At this point, the list of group 68 contained 225 journals of which 77 were level 2. Even with the new principle, it still had to downgrade dozens of journals. Hence, a dual strategy was adopted, which involved "close examination" of journals with no impact-factor and of journals in which no Danish researchers had published. The list was finally resubmitted with an accompanying standard disclaimer also included in the submissions by the majority of the other groups[14]:

> We would like to emphasize that the revised categorization proposed for our field is to be considered as a poor solu-

> tion (and perhaps not even the 'least poor'). We do not find the conditions satisfying or appropriate and we do not vouch for the overall model. Our role has been strictly consultative and our general advice is to not use the model in practice before all consequences and problems have been further examined and the quantitative data has been improved.

The list had now finally been made, although clearly not to everyone's satisfaction. Soon after the authority list was publicized by the agency. From then on the life of the bibliometric indicator began in earnest.

Contextualizing Indicators

Several explanations are available with which to interpret the social and political effects of new quantitative systems. However, in the context of the authority list and the bibliometric indicator, two are particularly pertinent.

First, a strong "indigenous" critique was brought forward by academics that read the indicator as *a tool of neoliberal ideology*. This line of criticism can also be found in critical organizational studies and sociology.

A second line of critique, affiliated with institutional theory, emphasizes how organizational members respond to quantification measures by "decoupling"; that is, by attempting to detach everyday work from the detrimental consequences of systems, while paying lip service to them. In what follows, I refer to this line of critique as focusing on *bibliometric rituals*. In this view, new quantification systems are certainly far from harmless, but the harm may be mitigated by members' on-going efforts to work around the system. Indeed, the fact that actors continuously spend energy to *stay detached* from these systems and to remain immune from their *intended* effects may be among the most harmful *indirect consequences* of new quantification systems. Nevertheless, *symbolic* adherence to the demands of indicators is important in order to ensure institutional legitimacy.

How do these analytical forms hold up in the case of the bibliometric indicator?

Bibliometric Rituals?

> A new, contentious and largely uncontrollable external pressure like rankings would seem to create a situation ripe for buffering, one likely to cause symbolic rather than substantive reactions (Sauder & Espeland, 2009: 64).

The making of authority lists was embroiled in classification controversies. Yet, for many of the most vocal opponents these "internal" struggles were secondary to more general problems relating to quantification: the translation of the authority list into the bibliometric indicator.

One way of looking at these issues is to return to the "reactivity of measures" argument. Sociologists Sauder and Espeland (2007) argue that new quantitative measures are often strongly resisted by organizational members. However, since direct opposition is likely to be counterproductive, involved actors tend to engage in indirect diversionary tactics. They defuse what is perceived to be the harmful consequences of the measures by pretending to go along with the official game, while decoupling from its substance to the degree possible: this gives rise to "performance paradoxes" of the kinds outlined by van Thiel and Leeuw, such as "mimicking the outward appearance of prestigious universities" (van Thiel & Leeuw, 2002: 270) rather than working to improve teaching conditions.

To the extent that such reactive responses are successful, quantification is

"symbolic"; and its effects predominantly ritual. By symbolic, Sauder and Espeland (2009: 64) refer to "relatively superficial changes" that can be put on display for public purposes. Although the language is slippery, they warn against viewing the symbolic as unreal or inconsequential (2009: 79). Symbolic responses, too, have real effects; they are just not the ones quantifiers hope for. This is why such responses are often characterized as "perverse". As a reiterated form, symbolic responses may take on a ritual character, as exemplified by Michael Power's (1997) studies of "rituals of verification". The term ritualization is used to describe precisely what happens when verification becomes "empty" (Power, 1997: 45), a "dramaturgical performance" (141), or purely "abstract" (96).

Certainly examples of symbolic reactions and decoupling can be identified in the present material. It shows, for example, in the way groups attempted to get around classification problems by defining as many journals as possible as "excellent" and by defining a maximum of journals as relevant to their own group. However, it must then also be noted that the agency was attentive from the outset to the negative consequences of as it saw it of such manoeuvring. Indeed, the agency never ceased to act on the assumption that decoupling would be attempted. First, the Norwegian model already included attempts to remove "perverse incentives". Second, the Danish agency also emphasized criteria relating to the maximum percentage of excellent journals, in order to ensure that journals were not added as part of a decoupling strategy.

Meanwhile, however, the committee was also committed, in its way to certain forms of ritualization. For one thing, it saw its strictly ad hoc procedure for choosing group members as wholly unproblematic, insisting that the somewhat randomly chosen members were, in fact, both excellent and representative of the domains of knowledge covered in their groups. For another, it persisted in arguing for the scientific basis of the 20% rule, though, as we have seen, its bibliometric basis had been translated beyond recognition.

Symbolic response was thus not one-way; rather it was a game of *mutual response and translation*. This mutual reactivity, however, meant that both decoupling and ritualization remained unstable, only halfway effective, strategies – and this for both researchers and administrators.

Nevertheless, it is tempting to see the bibliometric indicator as a "ritual of verification" at the *general level* of Danish research policy. Officially, it was constantly highlighted that the allocation of funds would depend on the publication of excellent science. Thus, by 2012, publication would account for 25% of distributed funds.[15] However, the increase was in fact far less dramatic than claimed, because the redistribution did not touch the *generic* pool of basic funds. Instead, the percentages referred only to the increase in basic funds brought about by the efforts to live up to the Barcelona goals. Although this was a significant amount (ranging from app. 38.5 M € to app. 81.8 € M in 2009-11) it still accounted for a low percentage of the basic funds as a whole (app. 904M € in 2009). In 2012, the bibliometric indicator would apparently determine less than 2.3% of research funding.

Accordingly, we are in the realm of hyperbole when we read in the press announcement from the ministry of science, technology and innovation stating that: "because of the significantly increased investments in research and an ambitious goal to create world-class Danish universities, the government wants to redistribute basic funds based on quality". The minister further stated that: "the universities that do well will be rewarded. And low quality must have consequences".

The ministerial pronunciation seems like a classic case of ritualization.

This reading has emphasized how the indicator seems quite unlikely to create the effects that were its official rationale. However, the central weakness of the decoupling explanation is that it fails to address why the indicator so engaged researchers' critical faculties. From the governmental side, as well, it poses the question of why hundreds of researchers had to be engaged in a prolonged and contentious process to create an indicator if it "really" did not do much. Something important must be going on, that is *not accounted for* by bibliometric rituals.

A Neoliberal Tool?

> It is worrisome that the top administration is so quick to engage in something that looks like a practical exercise in bibliometric cash thinking for low-level managers (Claus Emmeche, "Freedom of Research" blog, May 18th 2009).

> Schools have not protected their practices from this environmental pressure (Sauder & Espeland, 2009: 64).

We can refer to Sauder and Espeland's (2009) argument one last time, in order to see what is left out of sight by focusing exclusively on symbolic reactions and ritualization. For their analysis of ranking of American universities shows that in spite of attempts, universities have largely *failed* to "decouple" from the new systems. To make this argument, however, Sauder and Espeland is required to change from an analytical approach inspired by new institutional theory to one focusing on the "disciplinary" and "internalizing" effects of quantification – inspired by Michel Foucault. This orientation is much closer to Danish critics of the indicator.

Although Claus Emmeche, head of group 68, is not unaware of the symbolic aspects of the list-making effort, his critical arguments have focused on what he sees as *actually harmful* political consequences. Very broadly stated, his analysis sees the bibliometric indicator as a new New Public Management tool; one used to wedge a neoliberal agenda into the world of scientific research.

Our narrative stopped at the time when the groups submitted their final lists to the committee. But, of course, this was not the actual end of the process. On the contrary, a series of complaints from groups flooded in along with these submissions. At yet another crisis meeting in March 2009, the head of the mathematics group gave a critical presentation that distinguished between different levels of bibliometric effect.

At the first level, the indicator was seen as "a tool for measuring research activities at the university level and a key for distributing basic funds to the universities". It was acknowledged that the list could be used for such purpose "in lack of better tools". However, another level was seen to be far more problematic: the list could be used to "measure research activity and quality at smaller units". Although this possibility had not been part of the rationale for making the indicator, researchers were increasingly worried that it might be used in this way anyway. Thus politics, in the form of threatening departmental and managerial decisions, was seen to enter research through the backdoor of bibliometrics.

This reading received strength in a context of economic crisis, where some universities cut down on research personnel. Indeed, rumours had it that the indicator was put to immediate use to legitimate firing researchers. Around the same time, the blog "Freedom of research," featured a story about the communication unit at

Copenhagen University, which had sent out a memo noting the significant differences in publication patterns among departments. This was important, the memo stated, given that basic funds were now distributed based on numbers of publications. The indicator seemed already to be pushing competition downwards.

Following the implementation of authority lists in Norway, a number of dissident researchers gathered their criticisms in a small volume entitled *Hva skal vi med vitenskap?* (What do we need science for?). In his contribution, media and cultural historian Anders Johansen identified a problem (Johansen, 2006). Faced with sustained criticism of the fact that English journals were systematically ranked higher than Norwegian-language journals, thus hollowing out the basis for using Norwegian as a research language, the Norwegian research agency responded by arguing that this had not been intended, and admitting that it had failed to communicate this clearly. Johansen replied:

> I don't know if I understand this completely, but I think it can be interpreted in this way: Since the change to English is only a *consequence* of the indicator, not the original *intention*, the indicator is in itself language neutral. To relate it to what might follow *in practice* is a complete misunderstanding. To explain this misunderstanding is a prioritized task for the agency: If people would only learn to ignore the consequences, they would realize that the arrangement is unproblematic. (Johansen, 2006: 94.)

Viewing the indicator as a neoliberal tool facilitates two different interpretations of the situation. The first is that the Danish (and Norwegian) agencies simply failed to foresee the consequences of the indicator: later they simply tried to patch up the results of their own ignorance. Contrary to the decoupling argument, there are actual negative consequences, but they are due to the *naivety* of research administrators and policy makers. The second version, in contrast, assumes that policy makers had a hidden agenda all along: however, this agenda is only gradually revealed to researchers through its consequences. Again, the consequences are actual and harmful, but in this interpretation they are due to complicity. Johansen's damning (and funny) indictment seems to waver between these two views.

In the final sections, however, I offer yet another interpretation; one that focuses on the inherent uncertainty of the list-making process and on the emergent agency of the bibliometric indicator.

Uncertainty and Emergent Agency

> Whether successful or not, and however far spread at this time, this kind of practice demonstrates that not only the behavior of individuals but that of organizations may be affected by bibliometric measures in ways that are clearly unintended (Weingart, 2005: 127).

For a few years, Danish researchers became intensively engaged in developing, while criticizing and reflecting upon, the authority list and the bibliometric indicator. They continuously challenged the idea that the list would in fact make scientific excellence visible in a reliable way. As well, they never ceased to question the official claim that the list would do *nothing but* making scientific excellence visible. But although these researchers swiftly tuned in to the potential dangers of the lists they were making, their concerns were predominantly with the neoliberal context in which the bibliometric indicator would be used as a policy tool.

In this paper I have developed a contrasting perspective by studying the consequences of the authority list, starting from the inside of its process of creation, and following the complex trajectory through which it was eventually given a more or less stable form. On the one hand, this has required attention to the way in which journals were classifiedγas belonging to specific groupsγand as belonging to specific levels. On the other hand, it has demanded detailed consideration of the controversies this process gave rise to: in struggles within groups, among groups, and between groups and the agency.

Analytically located in between classification and controversy studies, this study of list-making thus offers a case of how changes in research policy are made up and contested. In particular, this has facilitated a focus on the uncertainty of the list-making process and on the emergent agency, not only of researchers and policy-makers, but, indeed, also of the list itself. Here I draw some conclusions about the organizational and scientific worlds built into the list and their emerging consequences and uncertainties, unfolding as the list gradually grew in its *capacity to act*.

The authority lists articulated a world in which science, policy and their relations were configured in particular ways. What is scientific knowledge and how is it organized? Basically, these are unresolved questions that have generated multiple incongruent answers. The authority list, however, incorporated some particular ones. It inscribed a conception of scientific knowledge as an entity that could be divided *cleanly* between disciplines. Thus, scientific knowledge formed an incomplete puzzle, to which disciplines each provide their pieces by publishing in specialized journals. This conception provided the organizational format through which 68 groups partitioned the 'cosmos' of knowledge.

The people inhabiting these groups were, of course, scientists. But, then, what is a scientist, what does he or she do, and why? Inscribed in the list-making process was a view of a scientist as a disinterested specialist, that shares general norms of excellence with every other scientist and who agree with fellow specialists on specific criteria of excellence within his or her own field.

Finally, we may ask, what role is the authority list itself supposed to play? The answer is that it is simply a general catalogue of excellent journals. From the point of view of the agency, and certainly for policy-makers, the authority list *does not do anything* except classify the excellence scientific contributions. As a neutral standard it makes scientific *quality* transparent. This is why, translated into the point system of the bibliometric indicator, quality can also be made *quantitatively* comparable.

Now in fact the list making process put each of these ideas under pressure. Paradoxically, the complex and diverse set of critical responses *to* the presumptions of the list were *fed into* the list-making process and helped shape it. How, then, were the presumptions of list making challenged? First, the classificatory process made clear that scientific knowledge hardly fit together like pieces in a puzzle. Disciplines overlapped and journals contained knowledge that belonged to several fields at once but were evaluated differently within them. Organizational challenges ensued due to such different evaluations.

And the list multiplied. First, it came to comprise several interdisciplinary clusters. But then, as group 68 exemplified, these interdisciplinary groups required constant negotiation of the boundaries with other groups and different disciplines. Far from homogenous and stable, the "identity" of science was thus enacted as uncertain, multiple and contested.

Additionally, it became clear that scientists were both less disinterested and less homogenous than had been presumed. In a contribution to the *What do we need science for?*, Beate Elvebakk and Vidar Enebak noted that: "the final list with accepted publication channels is made through 'consensus processes' in the disciplinary environments – an expression that must have been made by people with only limited experience with such environments" (Elvebakk & Enebak, 2006: 19). Although efforts to reach consensus were certainly part of the process, the work to classify and evaluate journals was attended by a whole range of discussions and disagreements that pointed to the politics of classification involved in stabilizing a version of scientific excellence.

Finally, the group lists, the authority list and the bibliometric indicator were far more *unruly* entities than anyone had expected. Rather than neutral tools, these lists turned out to be instigators of classification controversies, attractors of competing and divergent interests and vehicles for expressing conflicting views on the politics of knowledge. If the task of involved researchers were to determine the quality of scientific journals, lists were the devices that simultaneously gave structure to this effort, highlighted the multiple stakes involved in classifying knowledge, functioned as negotiation devices and worked as tools for managing the uncertainties of the process.

Making Lists, Enlisting Scientists

The present account enables an alternative perspective on the politics of the bibliometric indicator, because its starting point is tracing the on-going *mutual responses and modifications* not only among researchers and the agency, but also, consequentially, among these groups of people and *other actors*, such as "level 1 and 2 journals", "estimated world-production of articles", "the 20% rule" and, centrally, the lists and the indicator.

We might summarize the general argument by stating that the emergent effects of the indicator is the result neither of a determining political logic, nor of a distance created by effective decoupling, but rather of a set of mutually modifying responses of all involved actors (of course *including* political actors). In other words, this was a context where *no one was quite sure what was going on;* a situation of uncertainty.

Centrally, this uncertainty was also an effect of the *emergent agency of the indicator*. Starting out as a simple document outlining a *process for producing a list* in each of 68 groups, the indicator morphed into 68 lists, integrated into a single one, transmuted into new versions and were shaped and challenged by competing ideas, visions and presumptions about science and scientific quality. Gradually, however, the authority list began to *take form* and doing so its existence became less challengeable, more irreversible. But while irreversibility increased, in the sense that it would no longer be possible to object to the authority list or indicator *tout court,* this did not imply that *its effects* were determinable (cf. Jensen 2010). Thus, whereas there is no doubt that the indicator *has begun to effect* researchers and policy-makers in increasingly diverse contexts; the involved actors are to a significant extent unsure about what its implications are going to be. Paradoxically, the list has emerged on the other side of these processes, in a form that is not controllable, either by those policy makers who inscribed their presumptions of science, scientists and scientific knowledge into the list making process in the first place, or by those scientists, who were first enlisted by the agency, but then soon began to challenge the contents and implications of the authority list *from within*.

It is due to this general context of uncertainty that the *emergent authority* of the bibliometric indicator provides an especially interesting entry point for understanding research policy in action. This same uncertainty also provides the rationale for characterizing the list making process with the ambiguous double designation: *making lists, enlisting scientists*.

Acknowledgments

The author would like to thank the editor and reviewers of Science Studies for their helpful comments. Particular thanks to Claus Emmeche who read and commented on several previous drafts.

Notes

[0] This article was edited and approved for publication by Torben Elgaard Jensen.

[1] This characterization takes at face value the official *quality-oriented* rationale for the indicator. That this rationale *uneasily co-exists* with a *quantity-oriented* interest in establishing an "objective" measure will become clear in what follows.

[2] Forsknings- og Innovationsstyrelsen.

[3] See http://www.fi.dk/viden-og-politik/tal-og-analyser/
den-bibliometriske-forskningsindikator.

[4] See http://faggruppe68.pbworks.com/w/page/6015708/FrontPage.

[5] Note that in subsequent usage the term *scientific* follows the continental tradition: the scientific thus includes what Anglo-American terminology categorizes separately as *humanities*.

[6] See http://dkuni.dk/english/

[7] Forsknings- og Innovationsstyrelsen.

[8] The classification of book publishers was temporarily suspended, due to the difficulties of establishing relevant quality measures that would hold for entire university presses. This paper does not engage that problem.

[9] This caused considerable embarrassment and eventually led to the formation of a special group 100 in which high-profile general science journals such as *Science, Nature* and *PNAS* could be defined as level two.

[10] http://professorvaelde.blogspot.com/2008/09/kosmos-verdensproduktion-og.html.

[11] Report available at: www.uhr.no/documents/Vekt_p__forskning__sluttrapport.pdf

[12] In a presentation from 2007, the head of the Danish steering referred to "goal displacement" and "perverse effects". These perverse effects related to different quality measures and differences in publication patterns. See http://forskningsfrihed.pbworks.com/følgebrev-efter-30102007konf (Consulted April 7th, 2010).

[13] http://journals.cambridge.org/action/displayJournal?jid=ECH, consulted April 7th, 2010.

[14] The disclaimer and details about who submitted it can be found (in Danish) at: http://faggruppe68.pbworks.com/f/Faggruppernes-forbehold-v2.doc (consulted April 7th, 2010).

[15] These funds were to be distributed to universities based on the aggregate of publication points determined by the indicator.

References

Barker, K. (2007) 'The UK Research Assessment Exercise: The Evolution of a National Research Evaluation System,' Research Evaluation 16(1), 3-12.

Bowker, G. C. & S. L. Star (1999) Sorting Things Out – Classification and Its Consequences (Cambridge, MA: MIT Press).

Elvebakk, B. & V. Enebak (2006) 'Kunnskapsløst Kunnskapsløft', in E. B. Hagen & A. Johansen (eds) Hva skal vi med vitenskap – 13 innlegg i striden om tellekantene (Oslo: Universitetsforlaget): 13-26.

Espeland, W. N. & M. Sauder (2007) 'Rankings and Reactivity: How Public Measures Recreate Social Worlds', American Journal of Sociology 113(1): 1-40.

Gieryn, T. F. (1999) Cultural Boundaries of Science: Credibility on the Line (Chicago, IL: University of Chicago Press).

Hacking, I. (1982) 'Biopower and the Avalanche of Printed Numbers', Humanities in Society 5: 279-95.

Hacking, I. (1995) 'The Looping Effect of Human Kinds,' in D. Sperber *et al* (eds) Causal Cognition: An Interdisciplinary Approach. (Oxford: Oxford University Press): 351-83.

Hansson, F. (2006) 'Organizational Use of Evaluations: Governanopen in cs4ce and Control in Research Evaluation', Evaluation 12(2): 159-78.

Jensen, C.B (2010) Ontologies for Developing Things: Making Health Care Futures Through Technology (Rotterdam: Sense Publishers).

Johansen, A. (2006) 'Politikk og kvalitet,' in E. B. Hagen & A. Johansen (eds) Hva skal vi med vitenskap – 13 innlegg i striden om tellekantene (Oslo: Universitetsforlaget): 92-115.

Lamont, M. (2009) How Professors Think: Inside the Curious World of Academic Judgment (Cambridge, MA & London: Harvard University Press).

Latour, B. (1987) Science in Action: How to Follow Scientists and Engineers Through Society. (Cambridge, MA & London: Harvard University Press).

Mackenzie, D., F. Muniesa & L. Siu (eds) (2007) Do Economists Make Markets? On the Performativity of Economics (Princeton, NJ & Oxford: Princeton University Press).

Osterloh, M. & B.S. Frey (2010) 'Academic Rankings and Research Governance', Institute for Empirical Research in Economics Working Paper Series 482. Available at: http://ideas.repec.org/p/zur/iewwpx/482.html.

Porter, T. M. (1992) 'Quantification and the Accounting Ideal in Science', Social Studies of Science 22: 633-52.

Porter, T. M. (1994) 'Making Things Quantitative', Science in Context 7(3): 389-407.

Porter, T. M. (1995) Trust in Numbers: The Pursuit of Objectivity in Science and Public Life (Princeton, NJ: Princeton University Press).

Power, M. (1997) The Audit Society - Rituals of Verification (Oxford: Oxford University Press).

Rottenburg, R. (2009) Far-Fetched Facts: A Parable of Development Aid (Cambridge, MA & London: MIT Press).

Sauder, M. & W. N. Espeland (2009) 'The Discipline of Rankings: Tight Coupling and Organizational Change', American Sociological Review 74: 63-82.

Schneider, J. W. (2009) 'An Outline of the Bibliometric Indicator Used for Performance-Based Funding of Research Institutions in Norway', European Political Science 8: 364-78.

van Thiel, S. & F. L. Leeuw (2002) 'The Performance Paradox in the Public Sector', Public Performance & Management Review 25(3): 267-281.

Weingart, P. (2005) 'Impact of Bibliometrics Upon the Science System: Inadvertent Consequences', Scientometrics 62(1): 117-31.

Casper Bruun Jensen
IT University of Copenhagen
Rued Langgaardsvej 7
2300 Copenhagen S, Denmark
cbje@itu.dk