6

Your Brother's Gatekeeper: How Effects of Evaluation Machineries in Research Are Sometimes Enhanced

Peter Dahler-Larsen

Introduction

The influence of evaluation machineries (such as bibliometric indicators) upon researcher practices is a much debated issue. One of the key points is that no indicators give a full picture. Some indicators do not show sufficient attention to publications in other languages than English (Archambault et al., 2006; Dahler-Larsen, 2018); to differences across fields and subfields in science, social science and humanities (Leydesdorff & Bornmann, 2016); to the consequences of research evaluation machineries upon the choice of research themes among researchers (López Piñeiro & Hicks, 2015); to the citations of publications rather than just publications themselves (Harzing & Mijnhardt, 2015); and, if social impact of research is taken into account, to the variations in definitions of such impact (Penfield et al., 2014), including the voices of various stakeholders throughout the research process.

P. Dahler-Larsen (✉)

University of Copenhagen, Copenhagen, Denmark

e-mail: pdl@ifs.ku.dk

It seems that there is no indicator which finally closes the “evaluation gap,” meaning a distance between what is measured and the values associated with research (Wouters, 2017). As a consequence, many have articulated advice about how to curb the influence of such machineries and secure they are used responsibly (Hicks et al., 2015). Advice include: “never use one indicator alone,” “be specific about which indicators are used for which purposes” and “always use indicators as support, not as a substitute for human judgment.” In a broader perspective, these warnings resonate with warnings against “automation bias,” a cognitive failure where the human mind places too much trust in technological algorithms providing information, a phenomenon also leading to “death by GPS,” where people drive into lakes if their GPS tells them to (Bridle, 2018).

Back to research evaluation. The underlying assumption is that somehow human judgment can and should function as a bulwark against unintended and constitutive effects of evaluation machineries. While this assumption may hold in some situations, it fails to consider the intimate interaction between human and non-human elements in research evaluation. “Peer review may already be ‘informed’ by metrics, albeit perhaps not in the systematic and expert led way the proponents of informed peer review would have wished for,” says Wouters (2017, p. 110).

Therefore, there is good reason to study the formal and informal practices of gatekeeping which take place in this interactive space where minds and machineries are woven together and associations are made (Latour, 2005). I suggest there are many of these practices. They do not only include official decisions about publications or promotions, but a range of daily-life activities where people engage in conversations about what is and what is not likely to pass in the light of various kinds of metrics, and, importantly, in the light of not what these metrics are, but what they are becoming. If the future is uncertain, one must act with caution. For those who (claim to) know the future, it can be in one’s interest to use this knowledge to position oneself and to make others act accordingly.

Empirical observations reported in this chapter come from a case study of the Danish Bibliometric Research Indicator (BRI; an indicator built upon an earlier Norwegian version of a similar nature) (Schneider, 2009). These observations reveal that quite a lot of social action takes place around this indicator based on interpretations and imaginaries.

As people take into account how they imagine future metrics, and act upon these imaginaries now, and bring others into action, too, it becomes possible to understand why human judgment in some situation helps make evaluation machineries such as bibliometrics even *more* influential than they would otherwise be.

By developing grounded hypothesis about how and why this multiplication of effects happens under some circumstances, this chapter contributes to an understanding of how the interactions between “human” and “machine-like” forms of evaluation contribute to constitutive effects of evaluation systems in research.

A key ingredient in these situations is anticipation—and co-construction—of a not-yet-constructed reality. Gatekeepers who “know the future” or anticipate a coming future play a key role, of course in combination with a range of situational factors. There is no guarantee that these acts of performativity are always successful (Butler, 2010). In fact, one of the reasons why quite a lot of “fuzz” around the BRI is found in the case study is that perhaps there is no easy, direct and linear way to a predictable form of use of the indicator. So many attempts are made by different actors with different perspectives and purposes (de Lancer Julnes, 2011, p. 67). In turn, this makes the social life of the BRI a dynamic one, but also one that remains ambiguous.

The chapter proceeds in the following way. First, a theoretical argument is provided for ambiguity and interpretability as key concepts in recent studies of the use of metrics. Secondly, the BRI and the case study of it are introduced. Then follows a number of incidents presented as vignettes illustrating interpretations and actions in relation to the BRI among Danish researchers and institutions. Finally, a short conclusion.

Metrics and Ambiguity

The concept of ambiguity refers to situations where a phenomenon can meaningfully be interpreted in multiple ways. Theoretically, the concept plays different, but not irreconcilable roles in various frameworks. In relation to the social construction of reality, ambiguity is an indication of some degree of “opening” of the future (Best, 2008). In practice studies,

ambiguity is a sign of “multiple orders of worth” which actors handle in concrete situations by developing a variety of strategies (Stark, 2009).

In recent years, we have seen a number of empirical studies of metrics which resonate with these and similar theoretical orientations. The production of numbers is itself a complicated and demanding social accomplishment (Porter, 1995; Desrosières, 2011). Dambrin and Robson (2011) have shown that perfect validity of an indicator is not a requirement for practical use. Instead, ambivalence and opacity can be conducive to the implementation of otherwise flawed measures.

There is sometimes uneven implementation of the same indicator system across institutions in the same country (Hammarfelt et al., 2016; Lind, 2019; Mouritzen et al., 2018). Institutions that integrate the system most directly into their management systems are not necessarily those institutions where professional values are most consistent with the spirit of the indicator system (Lind, 2019; Mouritzen et al., 2018). The opposite may be the case if managers use the new system to induce change.

Even when a system is adopted, struggles over documentation practices sometimes remain unsettled or ambiguous over fairly long periods (Mouritzen et al., 2018; Kaltenbrunner & de Rijcke, 2017). Sometimes, people under evaluation are in position to influence the design of evaluation systems (Pollock et al., 2018), but since such influence is highly ambivalent, participatory processes run far from smoothly and may be perceived with suspicion and mixed feelings (Jensen, 2011). These observations may be particularly pertinent among researchers who usually cherish autonomy and peer review as sacred professional principles.

Once an evaluation system is in place, it may suffer from “mission drift,” so that it over time serves other purposes than its original ones (Kristiansen et al., 2017). For example, when an evaluation system is connected with money streams, its main function may change from provision of information to resource allocation. Some even suggest the existence of a “runaway effect” (Shore & Wright, 2015). The potential runaway or “mission drift” may sometimes be paradoxical, however, since such effect, as we shall see, does not always hinge on financial and material implications alone. The imaginary aspect is also important.

Interpretations and behaviors among researchers themselves may in fact set in motion a “runaway effect.” Managers may promise that they do

not intend to use particular indicators at the level of individual researchers, but only at collective levels such as departments or research groups. This practice of “buffering” may be seen as a good ethical practice. However, to the extent that individual scores are publicly available or can be computed by individual researchers themselves, they may use their own scores for promotion or marketing purposes (Fochler & de Rijcke, 2017). The original promise of buffering is broken, but it is researchers themselves who break it.

Researchers can also take precautionary actions against potential future use of the evaluation system, thereby setting in motion a new set of effects. Given differences between original purposes and emerging purposes of evaluation systems, and the differences between the design of such systems and imaginations of their functions, it becomes clear that the official promises and declarations about the purposes of evaluation systems cannot be trusted to predict the future use of such systems, even if such declarations are honestly made (Dahler-Larsen, 2013).

Evaluation systems such as bibliometrics and rankings can have dramatic consequences for institutions, especially if strong alliances around these institutions exert pressure on managers to act upon the scores (Espeland & Sauder, 2007). In other situations, managers can intentionally pursue a definition of reality that is constituted only by what is made visible by particular evaluative machineries, thereby reducing the complexity they are dealing with while leaving the difficult interface between evaluation and reality to others in the organization (Roberts, 2017).

As research evaluation based on quantification of publications and citations apparently increase in importance, researchers will potentially change their practices accordingly. Some of these practices may be unfortunate (such as producing more publications of lower quality, focusing on safe but trivial research questions, and slicing projects into several publications) (Osterloh & Frey, 2010). Other practices, however, may be even more problematic, unethical or illegal, including misrepresentation and misconduct (Biagioli et al., 2019). For this reason, institutions see an increasing need to sharpen their regulations of ethical research conduct, documentation practices and more. One of the side effects of these endeavors may be to cast a shadow of suspicion on normal practices which merely happen to not present themselves neatly in relation

to the new regulations and detailed documentation guidelines (Dahler-Larsen, 2017).

In other words, while evaluation systems produce some forms of clarity and transparency, they also produce their own ambiguities. Furthermore, these studies show that the function of evaluation systems is not a physical property inherent in such systems. A more productive focus is on the activities of people in and around these systems (Becker, 1998). If the evaluative systems have “functions,” they are produced by these activities. However, and that is the point, all these activities may be based on neither clear nor consensual understandings of the metrics and their meanings. If ambiguity is ever-present, and the social construction of metrics is unfinished, we can expect quite a lot of interpretive activity directed toward guessing what the metrics will bring in the future. This activity may be a constructive factor itself.

A Case Study of the Danish Bibliometric Indicator

The Danish Bibliometric Indicator (BRI) was politically decided in 2009 and came into effect in 2010. The alleged purpose was defined in terms of a “healthy competition” about resources for research (Mouritzen et al., 2018, p. 17). The BRI is basically a mechanism for distribution of bibliometric points.

Appointed groups of researchers in all disciplines and subdisciplines divide all publication outlets into two levels, while only 20% of the world production is allowed to be placed at level 2, the finest level. Level 1 is intendedly more inclusive, although only peer-reviewed publications count. All forms of publications such as articles, monographs and book chapters at levels 1 and 2, respectively, are then given a particular number of bibliometric points. Finally, a proportion of all state funding of research is reallocated across research institutions depending on how many points they scored. In Denmark, the redistribution takes place only across institutions, not across fields. Over the years, depending on a change in complicated mathematical formulae, the financial value of a BRI point has increased.

In the following case study, daily-life incidents in research institutions in which the BRI played a role will be reported in the form of short narrative vignettes. The vignettes are based on personal observations of the author (although “a research institution” is not necessarily the author's present employer).

One problem with the methodology used here is subjectivity regarding the selection and reporting of the incidents. On the other hand, however, a strength of the same methodology is its ability to capture incidents, arguments and interactions as they unfold in daily life. Furthermore, the specific purpose of this chapter is to highlight the interpretable and interpreted nature of metrics. The methodology is therefore consistent with the aim of this chapter, and it resonates with the orientation toward social practice theory, which characterizes several studies cited above.

Vignettes

Incident 1. Soon after the introduction of the BRI, there is a research symposium where a small group of senior and junior researchers at a department discuss research papers. One of the seniors claim that the level of expectations in international publication has increased in recent years. In discussion of a particular methodology commonly used, “you will not get published in the good journals,” he says, unless particular requirements regarding that methodology are met. He explains what these requirements are. Presumably, junior researchers must follow his advice if they wish to hope for a future in academia.

Although his statement is not causally linked to BRI as such, it helps create a context of rising expectations. It is not specified what “good journals” means more specifically. Nevertheless, the incident exemplifies how local gatekeepers and “wise men” can use the broader context of publication pressure to channel the energy and focus of younger researchers into particular directions. Although nobody can exactly *know* what the future brings and nobody has mapped all the criteria used in all editing decisions and future promotion decisions, there are “wise” men who offer an “authoritative” view of “what is required.” Ambiguity is transformed into advice about choices of paradigm and methodology. Again, although the

BRI is not referred to directly, its very existence suggests that from now on it may have more serious consequences if you do not do “what is required” since the BRI is part of university management.

Incident 2. Soon after the introduction of the BRI, managers at one university in Denmark decided to use the principles of the BRI in their internal allocation of resources across departments, in other words to reinforce the internal pecuniary repercussions of the BRI. At the same time, they defined a minimum threshold of BRI points expected from each researcher over a period of time. They declared that this was a way to prepare their university for the future. Over the years it turned out that this particular university gained from the inter-institutional reallocation of BRI funds as compared to the situation before BRI was introduced. Several other institutions did not draw any implications of the BRI for their internal allocation of resources. Some research groups ignored the BRI because they thought that the existing academic reputations of various journals provided a more serious and nuanced assessment of their value than the simple two-tiered BRI system. Perhaps they assumed that their own prestige would be strong enough to withstand any pressure from the BRI system, and they assumed that the BRI would not survive in the long run.

Incident 3. On a normal day at the department, a professor with a very good reputation talks about the qualities of a recently hired PhD student. Already, the student has got an article accepted at level 2, it is said. The example shows that although researchers often refuse to accept the BRI as a reflection of true academic value, they nevertheless use BRI terminology in some of their descriptions of great achievements.

Incident 4. One department established a system according to which all publications at level 2 release a financial reward to the authoring researcher. Over the years, this system is believed to have contributed to a significant rise in the quality and quantity of publications. The BRI system has contributed to this order of things by providing an externally defined list of journals which relieved the researchers of the otherwise painful task to internally agree on a list of what the “best” journals would be.

In a recent external evaluation, the evaluation committee recommends the elimination of the reward system in the department. The argument is

that now that the system has helped raise the level of achievement, it can now be taken for granted that everybody knows the importance of high-level publications. The recommendation is implemented, although one might ask: If it is acknowledged that a given financial incentive has worked, will its disappearance not make a difference? But the evaluators and the managers assume they know the future.

Incident 5. One researcher at one institution is invited to contribute a chapter to a book edited by a researcher at another university. The editor supplies the invitation with a remark saying that of course, all authors will be given BRI points for their contribution. The invited researcher accepts the invitation.

Incident 6. In order to compensate for the somewhat rough distribution of all publications into two levels in the BRI system, and in order to sharpen the focus on the very best journals, it is suggested to introduce a level 3 in the BRI system. A consultation process is designed. A research committee discusses the proposal. (The research committee is a departmental committee responsible for strategic and practical issues related to research. It consists of leaders of all research groups and research centers in a department). The field consists of different subdisciplines, and since only a small fraction of journals (5%) can be placed in level 3, not all subdisciplines are likely to be represented there. This will create tensions between the subgroups. It is also foreseen that when it comes to the exact identification of journals to be placed in level 3, there will be intense discussions and maybe conflicts. When a committee member proposes to base the decision on an objective criterion such as journal impact factor, another member answers that based on the literature on research evaluation, the journal impact factor is not regarded as an unproblematic criterion. It also cannot be used without normalization across subfields. After lengthy discussions, the issue is brought back to the national BRI committee. It turns out that there will be no level 3, because the other research groups in the same field in the country are against the idea.

Incident 7. A well-respected researcher returns to Denmark after having worked abroad for some years. The researcher is astonished about the fuzz related to the Danish BRI system. The researcher believes that it is silly that Danes establish their own system which does not reflect the exact status that various publications have in the international world.

Another researcher argues that funds paid out through the BRI system comes from Danish taxpayers, so Danes have the right to decide whatever principles they are pleased with, as long as they are financing the consequences of their decisions. Furthermore, perhaps on a more serious note, it is suggested that there is no such thing as a single, uniform, authoritative and undebatable determination of the international reputation of all publication outlets. In addition, geography actually plays a role. For example, in Denmark it may be legitimate to prioritize Scandinavian studies or EU studies higher than they would be, for example, in the US.

Incident 8. A researcher claims to have identified a number of phony publication outlets in the BRI system. All researchers are therefore asked to participate in an official cleaning process. In the local research committee, everybody supports the removal of phony publications, but then the consensus ends. One researcher argues that the committee should not spend much energy on BRI. Intense discussions are only likely to lead to spur dissensus, but not support to major changes in the design of the BRI. In addition, the financial impact of changes in the BRI are likely to be minimal and not implemented at the department level.

Other members of the committee argue that despite the limitations of the BRI, it is likely to remain an important factor in research evaluation. For example, one never knows whether research committees in the future are likely to look at BRI scores for individual researchers in promotion and hiring situations. Even if committee members themselves would not attach much weight to BRI scores themselves, they can be instructed to do so through the terms of references given to them by university managers. For this reason, it is argued, the importance of the BRI may increase in the future, so it is important not to ignore it.

One view is that the included journals should reflect “the core of our field.” Another view is that the original idea in the BRI is to be inclusive, at least at level 1, thereby stimulating plurality and diversity. It is felt that the BRI is again used to enhance a definition of the field, which is in fact not everybody’s definition. The discussion is inconclusive, but the committee decides to return to the issue in future meetings.

Incident 9. As part of the clean-up process mentioned above, the Ministry initiates a review of a large selection of registrations made in the BRI system. To that purpose, it uses a new set of regulations hitherto

unknown among researchers. These regulations clarify that only research publications are allowed to count. This is more complicated than it appears to be, because, for example, in social science, some books have overlapping functions, such as being a research publication and a book used in teaching, or a research publication that is also used to stimulate public debate. But researchers are insistently asked to make sure that they are clear about the primary purpose and primary audience of all their registered publications. It is also reiterated that only publications subject to peer review can be given BRI points. As a result of this clean-up process, the author contributing a chapter to the book mentioned in *incident 5*, is contacted and asked to change the registration of the book chapter. The argument is that the book is “perhaps not a research book” as it might be used in teaching. Although the official message is that within the framework of the new regulation, the responsibility to determine the type of each of his/her publications ultimately rests with the author, the researcher decides in this particular case to change the registration of the book chapter into “teaching.” Paradoxically, however, the researcher thinks it deserves to be mentioned that at the institution where the editor and other colleagues work, their contributions to the same book remain registered as “research.” The researcher therefore continues to believe that some element of ambiguity remains inherent in the very practices of documentation and registration which are crucial to the credibility of the BRI.

Incident 10. The Ministry of Research and Innovation finds it is time to evaluate the BRI. In order to start thinking about relevant issues and evaluation questions, the Ministry invites key academics and evaluators to a meeting. Many issues are discussed, among others whether the present version of the BRI prioritizes quantity over quality. It is also discussed whether the fuzz among researchers over the BRI is paradoxical given the fairly limited redistribution of funds caused by the BRI. However, as a counterargument, it is mentioned that once the BRI is in operation, it is easy to increase its financial impact by a simple change of an algorithm in a spreadsheet in the ministry. A few months later, it is announced that the Ministry does not wish to move forward with an evaluation of the BRI.

Incident 11. A researcher publishes a book with results of a longitudinal research project on how researchers have responded to the BRI

(Mouritzen et al., 2018). Survey data in several rounds are supplemented with interview data and bibliometric data. It is shown that the BRI has an effect on publication patterns, but only in particular fields and only in particular universities. It is also suggested that there is a correlation between use of the BRI and stress among some researchers at those institutions where it is implemented zealously.

At other institutions, researchers know very little about the BRI, the monetary consequences of BRI points, and who gets the money. For these reasons, the BRI presumably plays a very limited role in their daily life. The book also raises a number of issues about documentation practices. When the book is debated in the central committee responsible for the BRI system, its methodology is criticized and the minutes state that “the book cannot stand alone” as an assessment of the BRI system. A member of the committee disagrees with this view, arguing that the methodological weaknesses in the book are not extraordinary, and that the book contributes to the generation of relevant knowledge about the BRI and its effects.

In an interview in public media, the author of the book claims that perhaps the days of the BRI system are numbered.

Incident 12. As a new head of department is hired, a new departmental strategy is developed. The strategy includes a couple of new focus areas. One ambition is to strengthen the social impact of research. Another has to do with finding new types of funding due to financial challenges. These two areas are not totally in line with a focus on BRI points. For the most part social impact of research in Denmark is facilitated through communication channels in Danish. Research demonstrates a fairly clear trade-off between publishing in Danish and having many citations (Dahler-Larsen, 2018). Publications with many citations are often international ones. And publications on level 2 in the BRI are almost exclusively international ones. Trade-offs and compromises can be made but scoring BRI points at level 2 and having social impact are clearly different things.

Next, regarding finances, what are the realities which researchers at the department face? How can they best help the department out of its financial predicament? How much is gained from the BRI system as compared to, for example, externally funded research projects? One researcher asked

himself, for example, if he wanted to earn as much money to his university through the BRI system as he got from his most recent externally funded project, for how many years would he have to publish one single single-authored monograph with a good international publisher more than he would otherwise publish? (And the question is asked under the strict assumptions that researchers at other universities do not increase their production more than they would otherwise do and that all the money gained from his publication activity goes directly to his department.) The answer is 120 years. In other words, there is no realistic way in the world that a reasonable extra individual effort via the BRI system can have any substantial effect on the financial situation at the department compared to other much more effective forms of funding.

Incident 13. The Ministry of Innovation and Research announces that time has come for a political revision of the overall financial model for publicly funded research in Denmark. A committee with international and national experts is put together with the task of describing various models and giving political recommendations. The Ministry argues that present models, including the BRI, put too much emphasis on the quantity of the research production at the expense of quality. As a consequence, the life of the BRI may take another turn.

At a meeting in the expert committee it is debated whether the BRI could be revised so that only a given number of publications every year for each researcher (say three or four) could release BRI points. In this way, researchers would focus on their best pieces of work, not on massive production. Rumors say, however, that the BRI has gotten many enemies among institutions who are not benefitting from the redistribution of funds, which follows from it. Others argue that the BRI is basically flawed and alternatives are needed. For example, some argue that peer review should play a stronger role in the research system, rather than “automatic” bibliometrics such as the BRI. In turn, some say that if peer review takes the form of expert panels visiting each research milieu at regular intervals, this model will be expensive and bureaucratic. Furthermore, in a small country like Denmark, it is not possible to recruit a sufficient number of independent experts to panels evaluating research published in the national language, which remains important at least in social science and the humanities. Another suggestion is to channel more funding

through research councils, so that funding would depend more on competition among proposals. Rumors say that not only do the experts disagree about what characterizes the best model, there is also fundamentally different institutional interests at stake, because no matter which model is chosen, the choice itself has financial implications for them. Others predict that even if the spirit of the times may not be favorable to the BRI system, the Ministry itself may have an interest in maintaining it, because a lot of resources was spent constructing and maintain the whole institutional and technical apparatus that makes the BRI possible.

Incident 14. The research committee mentioned above again discusses its standpoint regarding publications and BRI. One of its members say that rumors claim that as a result of the deliberations in the national expert committee, maybe the whole BRI will not survive. The research committee decides to debate the issue again at a later point in time.

Incident 15. At the department, it is recommended that all researchers carry out a Personal Research Review at regular intervals. This is basically a consultation where a researcher discusses his or her publication strategy with a respected colleague. Before the consultation, the individual researcher prepares a document describing his or her existing production and as well as ideas for the future. One researcher explains that the normal plan, all other things being equal, is to publish articles at level 2 in the BRI whenever possible. But the main part of the consultation focuses on what can be done to increase the number of Google Scholar citations (as a sort of proxy for impact). Nevertheless, the incident shows that the BRI continues to play a role, albeit perhaps not a dominant role, in the considerations and practices among researchers in their daily life.

Discussion and Conclusion

A lot of activity is going on with regard to understanding, interpreting and sometimes influencing the BRI. There are many people playing the role of interpreters. As attempts to subject the BRI and its trajectory to sensemaking are socially distributed (Weick et al., 2005), so are the resulting gatekeeping functions.

It can be legitimately argued that the 15 incidents reported above are not only subjectively reported, they are also poorly connected as a narrative. Even rumors are reported. However, precisely the *lack* of tight connection between the incidents, and their lack of foundation in objective truth, is an important part of the story, since it opens wide spaces for interpretation. Perhaps so much activity is going on in the reported incidents exactly *because* even fairly insightful people are not in position to predict what the BRI will be like, which documentation practices will count, and which implications the BRI will have on managerial and professional practices.

Many of the incidents reported seem to include activities that have no particular finality. As stated by Butler (2010), social construction is not always successful. There seems to be much wasted energy around the BRI. Nevertheless, it is too early to tell which activities are successful given the instability of the system as such. Stories with finality in them can only be told in retrospect, when we see what the “outcome” of social processes was (Castoriadis, 1987).

Logically, because of the open-endedness of the BRI story as it goes on, there is therefore also a lot of activity that has to do with bringing oneself into a position where the BRI has at least been sufficiently taken into account so that it does not become a total surprise. We can call this *precautionary* or *preemptive use* of the BRI. Especially under complex and dynamic circumstances, precautionary or preemptive use of evaluation may be an important mechanism contributing to constitutive effects of evaluation machineries. When people take action based on what they perceive might be a reality in the future, they in fact help create a particular kind of social order (Hanson, 2000). It may contribute to this mechanism that researchers are sensitive to factors influencing their reputation (Hicks et al., 2015). So, if some see BRI points as a source of reputation-building, it is important to watch out because reputation is a positional good, and researchers are not only colleagues, but also competitors. In this context, it is an important observation that researchers are often actively using bibliometric measures even if they are also critical about the validity of such measures (Fochler & de Rijcke, 2017).

Perhaps the most critical and theoretically interesting point about the preemptive or precautionary use of the BRI system is that people who

have interpretations of what the BRI will mean in the future implicate the actions of others as a consequence of these imaginations. The (imaginary construction of the) BRI leads to imperatives such as: "You have to write in this way." "We have to define the core of our discipline." "We have to establish amongst ourselves a common understanding of the hierarchy of publications in our fields." "We have to discuss publication strategies." "We have to take precautions regarding the role of the BRI scores in assessment work." "We have to collectively take a standpoint regarding registration practices." In these many ways, researchers implicate each other both as competitors, as gatekeepers and as colleagues. The fact that the social relations in which the BRI is implicated are sometimes competitive, sometimes controlling and sometimes cooperative in no way reduces the total constitutive effects of the BRI.

In all these ways, and sometimes paradoxically in the midst of confusion and fuzz, the BRI is used both as an implicit sign and as an explicit argument to incite particular understandings of research and collective action based on these understandings. Thus, a particular contribution of this chapter has been to explain that sometimes imaginations of the future of BRI constitute a key ingredient in the social construction of the effects of the BRI. We know they are imaginations because our observations show that different groups hold different views about the meaningfulness, use and future of the BRI.

These imaginaries bring in a number of other agendas with them, such as how to promote particular methodologies or particular definitions of what constitutes the "core" of particular disciplines or the "quality" of research, not to mention what constitutes the very definition of research.

This mechanism may help explain why an evaluation system such as the BRI which has fairly limited financial effects nevertheless has an effect upon minds, mentalities, debates and practices. Just because the BRI is not financially critical for individuals at the present moment in time, it does not mean that it cannot be more financially critical in the future. The precautionary or preemptive logic here contributes to understanding why a system with limited financial effects in the present can still create much fuzz. The point is not that materiality does not count. The point is rather that imagined materiality does count, and may be theoretically and practically very important. Given the strange ramifications of the imaginaries illustrated in this chapter, it has been shown that "human

judgment" in all generality is not enough as a vaccine against constitutive effects of evaluation machineries.

Further research might contribute to understanding the role of imagination in the social construction of effects of evaluation systems. In the meantime, at the practical level, practitioners and gatekeepers of many kinds should perhaps reflect upon their own role in sometimes enhancing and multiplying the effects of evaluation systems. It is not enough here to merely enhance "human judgment" in contradistinction to evaluation machines, because it is "human judgment" which produces the imaginations of the evaluation systems as described in this chapter. It has got to be a more complex and well-reflected kind of human judgment. Or perhaps it is just an idea to stop and breathe for a while in the recognition that effects of evaluation systems are sometimes produced by gatekeepers who imagine they are conquering the future.

The observations presented in this chapter offer a perspective on how to meaningfully respond to the pressures from evaluation machines. Just like most fake news are spread on the internet by individuals who pass them on, the effects of evaluation machines hinge on thousands of small individual actions, which form a large network of social consequences.

As this chapter has shown, individual reactions to evaluation machines sometimes enhance the social implications of these machineries. In research you are not only your brother's keeper. You are in fact your brother's gatekeeper. You may want to take this into account when you deal with evaluation machineries.

One positive implication can be negatively articulated: Do not inadvertently use your human judgment and imagination to multiply and increase the effects of evaluation machineries. People's reactions to performance measurement are part of the construction of the political effects of performance measurement (Johnsen, 2008). However, if open protest may not be fruitful, as people may think that protesters speak up because they are themselves not able to produce good metric scores, then tacit inertness may be a meaningful practical strategy. Perhaps there is wisdom in not elevating "human judgment" to the point where one knows what the future brings and what must therefore be done. Perhaps it is better to enjoy the relative freedom in the present, the freedom that comes with ambiguity. Perhaps it is better to imagine what deserves to be published rather than what deserves to be counted.

References

- Archambault, É., Vignola-Gagné, É., Côté, G., Larivière, V., & Gingrasb, Y. (2006). Benchmarking scientific output in the social sciences and humanities: The limits of existing databases. *Scientometrics*, 68(3), 329–342.
- Becker, H. S. (1998). *Tricks of the trade, how to think about your research while you're doing it*. The University of Chicago Press.
- Best, J. (2008). Ambiguity, uncertainty, and risk: Rethinking indeterminacy. *International Political Sociology*, 2, 355–374.
- Biagioli, M., Kenney, M., Martin, B. R., & Walsh, J. P. (2019). Academic misconduct, misrepresentation and gaming: A reassessment. *Research Policy*, 48(2), 401–413.
- Bridle, J. (2018). *New dark age: Technology and the end of the future*. Verso Book.
- Butler, J. (2010). Performative agency. *Journal of Cultural Economy*, 3(2), 147–161.
- Castoriadis, C. (1987). *The imaginary: Creation in the Social-historical Domain*. Stanford University Press.
- Dahler-Larsen, P. (2013). Constitutive effects of performance indicators—Getting beyond unintended consequences. *Public Management Review*, 16(7), 969–986.
- Dahler-Larsen, P. (2017). *The new configuration of metrics, rules and guidelines creates a disturbing ambiguity in academia*. LSE Impact Blog.
- Dahler-Larsen, P. (2018). Making citations of publications in languages other than English visible: On the feasibility of a PLOTE-index. *Research Evaluation*, 27(1), 212–221.
- Dambrin, C., & Robson, K. (2011). Tracing performance in the pharmaceutical industry: Ambivalence, opacity and the performativity of flawed measures. *Accounting, Organizations and Society*, 36, 428–455.
- Desrosières, A. (2011). How real are statistics? Four possible attitudes. *Social Research*, 68(2), 339–355.
- Espeland, W., & Sauder, M. (2007). Rankings and reactivity: How public measures recreate social worlds. *American Journal of Sociology*, 113(1), 1–40.
- Fochler, M., & de Rijcke, S. (2017). Implicated in the indicator game? An experimental debate. *Engaging Science, Technology, and Society*, 3, 21–40.
- Hammarfelt, B., Nelhans, G., Eklund, P., & Åström, F. (2016). The heterogeneous landscape of bibliometric indicators: Evaluating models for allocating resources at Swedish universities. *Research Evaluation*, 25(3), 292–305.
- Hanson, F. A. (2000). How tests create what they are intended to measure. In A. Filer (Ed.), *Assessment: Social practice and social product* (pp. 67–81). Routledge Falmer.

- Harzing, A.-W., & Mijnhardt, W. (2015). Proof over promise: Towards a more inclusive ranking of Dutch academics in economics & business. *Scientometrics*, 102(1), 727–749.
- Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, 520(7548), 429–431.
- Jensen, C. B. (2011). Making lists, Enlisting scientists: The bibliometric indicator, uncertainty and emergent agency. *Science Studies*, 24(2), 64–84.
- Johnsen, Å. (2008). Performance information and educational policy making. In W. Van Dooren & S. Van de Walle (Eds.), *Performance information in the public sector. Governance and public management series* (pp. 157–173). Palgrave Macmillan.
- Kaltenbrunner, W., & de Rijcke, S. (2017). Quantifying ‘output’ for evaluation: Administrative knowledge politics and changing epistemic cultures in Dutch Law faculties. *Science and Public Policy*, 44(2), 284–293.
- Kristiansen, M. B., Dahler-Larsen, P., & Ghin, E. M. (2017). On the dynamic nature of performance management regimes. *Administration & Society*, 00(0), 1–23.
- de Lancer Julnes, P. (2011). Performance measurement beyond instrumental use. In W. van Dooren & V. de Walle (Eds.), *Performance information in the public sector. How It Is Used*. Palgrave Macmillan.
- Latour, B. (2005). *Reassembling the social. An introduction to actor-network-theory*. Oxford University Press.
- Leydesdorff, L., & Bornmann, L. (2016). The operationalization of “fields” as WoS subject categories (WCs) in evaluative bibliometrics: The cases of “library and information science” and “science & technology studies”. *Journal of the Association for Information Science and Technology*, 67(3), 707–714.
- Lind, J. K. (2019). The missing link: How university managers mediate the impact of a performance-based research funding system. *Research Evaluation*, 28(1), 84–93.
- López Piñero, C., & Hicks, D. (2015). Reception of Spanish sociology by domestic and foreign audiences differs and has consequences for evaluation. *Research Evaluation*, 24(1), 78–89.
- Mouritzen, P. E., Opstrup, N., & Pedersen, P. B. (2018). *En fremmed kommer til byen, ti år med den bibliometriske forskningsindikator*. Syddansk Universitetsforlag.
- Osterloh, M., & Frey, B. S. (2010). *Academic Rankings between the “Republic of Science” and “New Public Management”*. Working Paper. Zurich: CREMA—Center for Research in Management, Economics and the Arts.

- Penfield, T., Baker, M. J., Scoble, R., & Wykes, M. C. (2014). Assessment, evaluations, and definitions of research impact: A review. *Research Evaluation*, 23(1), 21–32.
- Pollock, N., D’Adderio, L., Williams, R., & Leforestier, L. (2018). Conforming or transforming? How organizations respond to multiple rankings. *Accounting, Organizations and Society*, 64, 55–68.
- Porter, T. M. (1995). *Trust in numbers: The Pursuit of objectivity in science and public life*. Princeton University Press.
- Roberts, J. (2017). Managing only with transparency: The strategic functions of ignorance. *Critical Perspectives on Accounting*, 55, 53–60.
- Schneider, J. W. (2009). An Outline of the bibliometric indicator used for performance-based funding of research institutions in Norway. *European Political Science*, 8(3), 364–378.
- Shore, C., & Wright, S. (2015). Audit culture revisited: Rankings, rating and reassembling of society. *Current Anthropology*, 56(3), 421–444.
- Stark, D. (2009). *The sense of dissonance. Accounts of worth in economic life*. Princeton University Press.
- Weick, K. E., Sutcliffe, K. M., & Obstfeld, D. (2005). Organizing and the process of sensemaking. *Organization Science*, 16(4), 409–421.
- Wouters, P. (2017). Bridging the evaluation gap. *Engaging Science, Technology, and Society*, 3, 108–118.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.