

### 3

---

## Playing and Being Played by the Research Impact Game

Michael Power

In 2014, two philosophy departments, at the universities of Birmingham and Keele in the United Kingdom, came equal top in a league table, each with grade point averages of 3.80 (where 4 is a maximum). The Department of Philosophy at Oxford University was ranked tenth with a grade point average (GPA) of 3.40.<sup>1</sup>

Ranking systems are widespread, not least in the field of education (see Kehm, this volume, chapter 6). For example, Espeland and Sauder (2007) examine the rankings of US law schools and their effects on the behavior of key organizational participants, such as deans, who are compelled to pay attention to them despite being doubtful of their worth. Furthermore, while small differences in GPA calculations can amplify differences in rank ordering, these crude snapshots of relative performance provide easy and popular comparability for nonspecialist publics. However, there is something particularly distinctive about the ranking of UK philosophy departments described above: it is based on an evaluation of the *impact* of their research.

By impact in this context, one would ordinarily imagine journal citations and other demonstrable measures of quality within the field of academic philosophy. Such bibliometrics have attracted considerable attention from analysts (e.g., Gingras, 2016). Yet this would be wrong. Impact in this UK setting means the social and economic beneficial impact *outside* academia. In other words, the departments of philosophy at Birmingham, Keele, and elsewhere in the United Kingdom were graded and ranked in terms of the social and beneficial impact of their research. In fact, all subject areas in UK universities were evaluated for this kind of impact as part of a major evaluation of research quality, the Research Excellence Framework (REF2014 hereafter, which is the successor to the Research Assessment Exercises of previous decades). UK universities made 1,911 submissions across all subject areas from 52,061 staff who produced 191,150 research “outputs”

of one kind or another. Importantly, for the purposes of this chapter, UK universities submitted 6,975 case studies to demonstrate the impact of their research.<sup>2</sup> As noted above, the Birmingham and Keele departments of philosophy came top of all philosophy departments for their impact.

How is it possible to produce such strange organizational facts as these, and for them not to be regarded as strange? In the next two sections, I provide a brief account of the REF2014 evaluation regime and of the impact case study (ICS) as a new accounting instrument in UK higher education. The argument then focuses on a specific style of evidence gathering for impact adopted by many UK academics—solicited testimony—and uses this feature of REF2014 to suggest that the requirement to demonstrate the impact of research is a *meta-game*, understood as an infrastructure for the production of a certain kind of truth (Foucault, 1980) that also contains the seeds of a kind of “institutional wrongdoing.” Indeed, as noted by Wouters (this volume, chapter 4), a meta-game signifies that it has become increasingly difficult to distinguish between gaming the system and the normal functioning of the system (see also Biagioli, 2016). Two important internal features of this meta-game are highlighted. First, while ICSs are qualitative narratives, they are also reductive in nature and create the platform for commensuration via evaluative metrics. Second, the impact meta-game is constituted by a “logic of auditability” with consequences for academic habits and orientation.

### REF2014 and Impact

The periodic evaluation of research in the United Kingdom every six to seven years is well established and has been replicated in other countries. Governments demand these evaluations in order to demonstrate that money is being well spent on high-quality research and, in theory at least, to allocate scarce resources to the best universities as judged by the evaluation exercise. In practice, compromises are always required in order to avoid a winner-takes-all concentration of reward in a few universities, leading many to question the purpose of these evaluations.

Different policy fields have had a longstanding interest in the impact of interventions, and “impact” and “impactfulness” have emerged as values in many areas, such as environmental impact assessment. In the UK higher education setting, a decisive catalyst for change was the Worry Report (2006), which recommended that universities measure the impact of their research outside the academy. This ambition finally became a reality for REF2014, and it was decided that twenty percent of total funding

for research would be awarded on the basis of such impact (Funding Councils, 2011). Despite initial skepticism and opposition to the crudity of an impact measure for the humanities, UK universities got on with the job of operationalizing the requirement. However, the approach that emerged was very different from the metrics-based regime imagined by the Worry Report. After a period of consultation and experimentation, the regulators of UK universities settled on a case study approach to the demonstration and evaluation of research impact. In effect, while the journal article has long been an “accounting unit” for individual academics (see Wager, this volume, chapter 17; Gingras, this volume, chapter 2), a further new statement was created—the *Impact Case Study* or ICS—to account for the impact of these primary accounting units. And, as noted above, UK universities produced nearly seven thousand of these ICSs.

### The Impact Case Study as an Accounting Unit

Academics have a long history of being involved, to a greater or lesser extent, in the world beyond academia. They may be advisers to government and business; they may develop beneficial technologies in medicine and engineering; they may be public commentators on issues of the day and so on. Yet the UK impact regime has changed the status of all these activities, which we might call *engagement*. Simply put, they have come to be regarded as not necessarily impactful in and of themselves. They have been redescribed as “pathways to impact” and therefore as distinct from ultimate impact in the sense of a beneficial change. Accordingly, UK academics have had to ask themselves two questions: “What has changed (outside academia) as a result of my research?” and, crucially, “How can I demonstrate it?”

UK universities needed to provide resources to support the ICS process and to build an infrastructure to cope with this entirely new requirement (Power, 2015). The workload was mitigated to the extent that not all research was required to demonstrate its impact; a norm emerged that roughly one ICS would need to be produced for every ten members of research-active academic staff (this is not a natural ratio and there is no reason to suppose it will not change). Furthermore, in contrast to the Worry Report, the rules published by the regulator were pluralistic about the kinds of impact that academic research might have. Creating a new life-saving drug might be the gold standard, but critical interventions in public policy debates contributing to change would also count as long as they could be proven. Finally, UK universities were highly motivated to

comply with the requirement since the financial reward for a 4\* ICS—the highest grade—was considerable. Indeed, an official was heard at a conference to confess that the ICSs were “overpriced.” Research impact had in effect become big business for UK universities.

Many UK academics, believing or knowing themselves to be impactful, embraced the impact agenda and got on with the work of producing their case studies. They were advised and quality controlled by committees that were themselves learning the impact game. ICSs would be evaluated for their “reach” and “significance” and there was a prescribed template design to shape and limit the form, content, and length of an ICS. Although ICSs might essentially be narrative in form (i.e., be “stories of impact”), they would have a prescribed structure with maximum word counts for each section. There were also more specific rules—not least about the accounting time window within which impact might be measured—but also quite a lot of pluralism about the kinds of possible impact to consider and also about the forms of evidence that might be used to support claims of impact (Funding Councils, 2011). A particularly significant form of evidence used by scholars at the London School of Economics (LSE), including myself, was “solicited testimony.” Reflecting on this evidence form provides some insight into the workings of the research impact regime.

### **Evidence of Impact and Solicited Testimony**

In the practical field of evaluation, it is accepted that a way to find out if an intervention has had an impact is to ask those people or groups who one would naturally expect to have been “impacted.” In other words, their testimony is solicited. Although this method of data collection has its own epistemological shortcomings, and care must be taken not to lead the respondents, for very pragmatic reasons, it was an attractive form of evidence for social scientists. Many very smart people at LSE encountered problems writing their ICSs largely because of the difficulties of causal attribution coupled with normal scientific caution and modesty. And even where their external impact might be self-evident to themselves, it would not be to others. Because the effects of research, if any, dissipate into the wider social and institutional environment of universities, evidential traces of impact lie outside the organization and are costly to collect. At LSE, the ICS proved to be a distinctive kind of genre that had to be crafted, and the process gave rise to many unexpected difficulties. In such a setting, and with a race against time to prepare the ICSs, the epistemic

weaknesses of solicited testimony were outweighed by its pragmatic, low-cost features.

Terms like “solicited testimony” imply seriousness and scientificity. But the impact game was new, and many academics, in the search for corroborating evidence, rang or emailed those who may have been influenced by their research. Conversations might have gone like this: “Look, you know that piece of work I did for you, based on my research, which you thought was important? Would you mind just putting that in a letter to me or in an email please?” In this way, a trace of external impact was created that could be collected as evidence. Importantly, it was not sitting out there waiting to be found. It was actively constructed by the researcher.

### Meta-Gaming

What is going on here? Were academics like me gaming the system by seeking solicited testimony in this way? Was this a cheap and quick way to build a credible ICS, particularly as the narrative of impact had to be constructed *ex post* and in a hurry? This is undoubtedly partly the case; at LSE, concerns were voiced about the overuse of this evidence type in many ICSs. These concerns were also shared by evaluators. One report on the REF2014 process and the evaluation of ICSs notes that they found it “very hard to assess the significance of an impact where evidence was nuanced and in the form of corroborating testimonials” (Rand Europe, 2015). However, much more is at stake here than gaming by individual academics.

The impact agenda in the United Kingdom represents a rebalancing of two logics or values that have always existed in tension with each other in academic life, namely, the logic of autonomous curiosity and the logic of use-value and economic benefit. Furthermore, different universities and different subject areas will combine these logics in different ways. Fields such as social policy, where action research is common, embraced the turn toward impact. Strangely perhaps, in business schools, where subdisciplines have been on a path to greater academic respectability, the impact agenda created more challenges than might be expected. And for fields such as history and philosophy, the agenda was entirely new and disruptive.

However, while the rise of impact accounting in UK universities exhibits this diversity of reactions, the example of solicited testimony as an evidence form suggests that something systematic is at stake. Following REF2014, new academic habits are visible, supported and routinized by

new infrastructures and databases for the collation and analysis of impact. Universities are creating dedicated roles, such as “impact officers,” and providing support to academics to help them maximize their impact (LSE Public Policy Group, 2012). Research funding bodies are awarding prizes for impact. In short, an entire *impact apparatus* is being created. UK funding bodies also often require a statement of *expected impact* or something similar in grant applications. In effect, this means that applicants must turn the causal pathway model on its head; the economic stakes of being successful in raising funding implicitly requires them to ensure that they are having or will have an impact *before* they do the related research. So the grant application process reveals a variant of Goodhart’s law, whereby an *ex post* outcome measure of impact “flips” to become an *ex ante* target (see also Griesemer, this volume, chapter 5, for more on Goodhart’s law). Rather than impact being a measured outcome of research, as it was for REF2014, it is now research, or a certain style of research, that is becoming the product of an apparatus that *targets* impact (Power, 2015).

Formally not all UK academics are required to prepare ICSs and be impactful. Yet impact is now one of the formal criteria in promotion guidance. Impact has therefore become an established norm of evaluation regardless of nuances of scope. Consequently, as the example of solicited testimony suggests, UK researchers are learning how to engage with possible users of their research, constructing them as good impactees who are actively cultivated and internally represented for the purpose of writing future ICSs.

From this point of view, the use of solicited testimony is not simply the gaming of the impact accounting system by individual academics. Rather, it reveals how the system actually works. The impact accounting regime may indeed be subject to gaming of its rules, but those rules in aggregate, embedded in apparatuses, constitute a *meta-game*, namely rules for the production of a certain kind of truth in Foucault’s (1980) sense, meaning systematized ways of governing what it is possible for individuals to say, and what actions and performances are legitimate. While it is easy to hold specific humans to account and make them visible *within* an accounting system, it is much harder to hold an accounting system to account and to provide a critical account of its operating logic. Below I argue that this operating logic in the case of research impact in the United Kingdom has little to do with social and economic benefit—the originating policy values—but is motivated by a distinctive cultural commitment to evidence gathering and its associated disciplinary power. I call this the “logic of auditability” (Power, 2019).

## Qualitative Commensuration

While the ambition of the original Worry Report was for the measurement of impact, and this is an ambition which has not disappeared, the ICS regime that took shape in the United Kingdom did not take this direction. While metrics (e.g., citations in public documents) supporting impact may be used as forms of evidence, this is within the context of an overall case study narrative. However, if metrics are not a required input of the impact accounting system, they are a consequence of it. The grade point average of 3.8 scored by the departments of philosophy in the universities of Birmingham and Keele noted above are composites of scores given to individual ICSs by a panel of evaluators. So at this point of evaluation, very different ICSs narrating very different kinds of impact become commensurable, are made capable of being compared, and, importantly, ranked (Espeland and Stevens, 1998).

The process of commensuration, the point at which qualities are made into quantities, is a continuing source of interest to scholars and to the emerging “valuation studies” agenda.<sup>3</sup> However, the case of research impact in the United Kingdom suggests that there may not always be a singular ontological jump from quality to quantity. The ICS accounting regime provides an example of how qualitative characteristics become subject to quasi-commensuration in templates and requirements for narrative precision. This is not a direct form of quantification, but rather its conditions of possibility—in this case, it enables grading by an evaluator. In other words, the point of metricization and quantification has its conditions of possibility in the construction of qualitative or hybrid narrative forms that integrate with, and support, quantification. So the ground for commensuration has already been prepared in their construction. The final stage of quantitative commensuration can then be performed by evaluators who make judgments about the quality of each ICS and express them in metrics.

## Impact and the Logic of Auditability

The research impact requirements in the United Kingdom emerged from policy ambitions to reconnect universities to the UK economy and to make this demonstrable. In particular, the ICS infrastructure requires the traces of impact to be constructed and reported in templates that enable them to be evaluated and audited. Such a model is familiar to accountants—it’s the way accounts are produced. Yet, this creation of traces, how we do it

and for what activities, whether they're citations or other fact-bearing objects, reflects a cultural logic of a distinctive kind, a logic of auditability (Power, 2019). Fundamentally, this is a *logic of trace production* rather than measurement specifically. For example, solicited testimony discussed earlier is one form of systematic, nonquantitative trace construction that is interiorized, embodied in ICSs, and then evaluated. Guided by the logic of auditability, the trace does not simply record the impact; it defines and constitutes what impact is.

The logic of auditability as a cultural form of trace production matters when considering the evaluators of the 6,975 impact case studies noted above. It is clear that they found the evaluation process very difficult, despite the discipline of the ICS production process. Evaluators reported that they had very little time or resources to drill down to check (i.e., audit) the underlying sources or traces referenced in the ICSs, and they drew attention to the poor quality of links to underlying evidence (Rand Europe, 2015). Their operational challenges reveal the face of the real “audit society” (Power, 1997). It is far from being a fully transparent and “auditable” society, or even one that's full of very confident inspectors and watchdogs who perform checks. Rather, the logic of auditability is essentially *productionist*; it names a cultural compulsion for organizations to construct and collect traces of activities, and to fabricate audit trails that link accounts of performance, such as ICSs, to an underlying evidence base. That evidence must be shaped, selected, and constructed so that traces of impact are auditable in principle, even if this is not possible in actuality.

The logic of auditability is a cultural value that passes for common sense and is reinforced by the construction of infrastructures for trace production. In this respect, the impact regime is a meta-game for the production of a distinctive kind of facticity. It is a meta-game because it embodies a potential for misconduct that has little to do with gaming metrics at the individual level. Rather, a kind of misconduct is embodied in the infrastructure itself whose deep logic is that of auditability and trace production, whether the traces are citations within academia (Biagioli, 2016) or traces of impact in the “outside” world. The logic of auditability does not itself value users or consumers or indeed economic value as such—the value is in the trace (Power, 2019). To understand the power of this logic is to understand why there seems to be so much investment in evaluation systems that are very costly and seem to produce little economic benefit.



## Conclusions: From Traces to Metrics

The chapter began with the example of UK philosophy departments playing the game of research impact. Of course we know that philosophy has an impact; the influence of Plato and Aristotle on Western culture and cognitive assumptions is well known. However, had Plato and Aristotle lived today, I think they would have been hard pressed to construct an ICS and to collect traces of the kind required by the REF2014 exercise. Ironically, this is because their impact has been so great and so diffuse that even the most helpful impact support officer would struggle to support them in producing the evidence. This is another paradoxical feature of the impact regime in the UK: the more impact you actually have, the harder it is to account for it.

I have suggested that the regime for accounting for research impact in the United Kingdom is a kind of meta-game in the form of elaborate rules for the production of a new kind of truth about academic research. Apparent individual gaming of this regime, such as the use of solicited testimony discussed above, is in fact a feature of the way the regime works. Following Foucault (1980), I also allude to a growing “impact infrastructure” that regulates the production of this truth in the form of acceptable “accounting” statements about impact. This impact infrastructure is in turn itself permeated by a systemic logic of auditability that demands the production of traces (not just numbers), which, when gathered into the ICS template, can be evaluated and then metricized. As regimes of impact truth production, these infrastructures have also created new forms of deviance. Strangely, this deviance does not always take the traditional form of academic fraud or gaming the system. Rather it also involves carrying on as before without regard for the cultural imperative to produce precise traces of activity that can “travel.” From this point of view, being an intellectual and writing long books are forms of deviance relative to, and brought into existence by, the research impact accounting system in the United Kingdom. But in a system whose logic would define Plato and Aristotle as deviant, we should probably ask where the ultimate misconduct really lies.

## Notes

1. <https://www.timeshighereducation.com/sites/default/files/Attachments/2014/12/17/k/a/s/over-14-01.pdf>. Accessed on May 9, 2016. A grade point average is simply what it says (i.e., an overall average of grades given on individual units of evaluation, usually calibrated from 1 to 4). The impact of all subject areas

was aggregated to give a score for impact at the individual institutional level. The UK Institute of Cancer Research was top in the rankings overall.

2. See <http://www.ref.ac.uk/media/ref/content/pub/REF%20Brief%20Guide%202014.pdf>. Accessed on May 9, 2016.

3. For the recently created online journal *Valuation Studies*, see <http://valuationstudies.liu.se/>.

## References

Biagioli, Mario. 2016. "Watch Out for Cheats in the Citation Game." *Nature* 535:201.

Espeland, Wendy, and Michael Sauder. 2007. "Rankings and Reactivity: How Public Measures Recreate Social Worlds." *American Journal of Sociology* 113:1–40.

Espeland, Wendy, and Mitchell Stevens. 1998. "Commensuration as a Social Process." *Annual Review of Sociology* 24:313–343.

Foucault, Michel. 1980. "Truth and Power." In Colin Gordon (ed.), *Michel Foucault Power/Knowledge: Selected Interviews and Other Writings*. Brighton, UK: Harvester Press.

Funding Councils. 2011. *Decisions on Assessing Research Impact*. Bristol, UK: HEFCE.

Gingras, Yves. 2016. *Bibliometrics and Research Evaluation: Uses and Abuses*. Cambridge, MA: MIT Press.

LSE Public Policy Group. 2012. *Maximizing the Impacts of Your Research: A Handbook for Social Scientists*. London: London School of Economics Public Policy Group.

Power, Michael. 1997. *The Audit Society*. Oxford: Oxford University Press.

Power, Michael. 2015. "How Accounting Begins: Object Formation and the Accretion of Infrastructure." *Accounting, Organizations and Society* 47:43–55.

Power, Michael. 2019. "Modelling the Microfoundations of the Audit Society: Organizations and the Logic of the Audit Trail." *Academy of Management Review*. <https://journals.aom.org/doi/pdf/10.5465/amr.2017.0212>.

Rand Europe. 2015. *Assessing Impact Submissions for REF 2014: An Evaluation*. Cambridge: Rand Corporation.

Warry Report. 2006. *Increasing the Economic Impact of Research Councils: Advice to the Director General of Science and Innovation from the Research Council Economic Impact Group*. London: Department for Business, Innovation and Skills.

This is a section of [doi:10.7551/mitpress/11087.001.0001](https://doi.org/10.7551/mitpress/11087.001.0001)

# Gaming the Metrics

## Misconduct and Manipulation in Academic Research

Edited by: Mario Biagioli, Alexandra Lippman

### Citation:

*Gaming the Metrics: Misconduct and Manipulation in Academic Research*

Edited by: Mario Biagioli, Alexandra Lippman

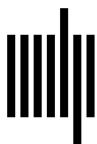
DOI: [10.7551/mitpress/11087.001.0001](https://doi.org/10.7551/mitpress/11087.001.0001)

ISBN (electronic): 9780262356565

Publisher: The MIT Press

Published: 2020

This title is freely available as an open access edition thanks to the TOME initiative and the generous support of the University of California, Davis. Learn more at [openmonographs.org](https://openmonographs.org)



The MIT Press

© 2020 Massachusetts Institute of Technology

This work is subject to a Creative Commons CC BY-NC-ND license.



Subject to such license, all rights are reserved.

This title is freely available as an open access edition thanks to the TOME initiative and the generous support of the University of California, Davis. Learn more at [openmonographs.org](http://openmonographs.org).

This book was set in Sabon by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data

Names: Biagioli, Mario, 1946- editor. | Lippman, Alexandra, editor.

Title: Gaming the metrics : misconduct and manipulation in academic research / edited by Mario Biagioli and Alexandra Lippman.

Description: Cambridge, MA : MIT Press, [2020] | Series: Infrastructures | Includes bibliographical references and index.

Identifiers: LCCN 2019010150 | ISBN 9780262537933 (pbk. : alk. paper)

Subjects: LCSH: Scholarly publishing—Corrupt practices. | Learning and scholarship—Corrupt practices. | Research—Corrupt practices. |

Communication in learning and scholarship—Moral and ethical aspects.

Classification: LCC Z286.S37 G36 2020 | DDC 070.5—dc23

LC record available at <https://lccn.loc.gov/2019010150>

10 9 8 7 6 5 4 3 2 1