



BIROn - Birkbeck Institutional Research Online

Moore, Samuel and Neylon, Cameron and Eve, Martin Paul and O'Donnell, Daniel and Pattinson, Damian (2017) Excellence R Us: university research and the fetishisation of excellence. *Palgrave Communications* 3 (16105), pp. 1-13. ISSN 2055-1045.

Downloaded from: <http://eprints.bbk.ac.uk/id/eprint/17617/>

Usage Guidelines:

Please refer to usage guidelines at <https://eprints.bbk.ac.uk/policies.html> or alternatively contact lib-eprints@bbk.ac.uk.



ARTICLE

Received 29 May 2016 | Accepted 12 Dec 2016 | Published 19 Jan 2017

DOI: 10.1057/palcomms.2016.105

OPEN

“Excellence R Us”: university research and the fetishisation of excellence

Samuel Moore¹, Cameron Neylon², Martin Paul Eve³, Daniel Paul O’Donnell⁴ and Damian Pattinson⁵

ABSTRACT The rhetoric of “excellence” is pervasive across the academy. It is used to refer to research outputs as well as researchers, theory and education, individuals and organizations, from art history to zoology. But does “excellence” actually mean anything? Does this pervasive narrative of “excellence” do any good? Drawing on a range of sources we interrogate “excellence” as a concept and find that it has no intrinsic meaning in academia. Rather it functions as a linguistic interchange mechanism. To investigate whether this linguistic function is useful we examine how the rhetoric of excellence combines with narratives of scarcity and competition to show that the hyper-competition that arises from the performance of “excellence” is completely at odds with the qualities of good research. We trace the roots of issues in reproducibility, fraud, and homophily to this rhetoric. But we also show that this rhetoric is an internal, and not primarily an external, imposition. We conclude by proposing an alternative rhetoric based on soundness and capacity-building. In the final analysis, it turns out that that “excellence” is not excellent. Used in its current unqualified form it is a pernicious and dangerous rhetoric that undermines the very foundations of good research and scholarship. This article is published as part of a collection on the future of research assessment.

¹ Kings College, London, UK ² Curtin University, Perth, Australia ³ Birkbeck, University of London, UK ⁴ University of Lethbridge, Canada ⁵ Research Square, London, UK Correspondence: (e-mail: cn@cameronneylon.net)

Introduction: the ubiquity of excellence rhetoric

Excellence” is the gold standard of the university world. Institutional mission statements or advertisements proclaim, in almost identical language, their “international reputation for [educational] excellence” (for example, Baylor, Imperial College London, Loughborough University, Monash University, The University of Sheffield), or the extent to which they are guided by principles of “excellence” (University of Cambridge, Carnegie Mellon, Gustav Adolphus, University College London, Warwick and so on). University research offices and faculties turn this goal into reality through centres and programmes of “excellence”, which are in turn linked through networks such as the Canadian “Networks of Centres of Excellence” or German “Clusters of Excellence” (OECD, 2014; Networks of Centres of Excellence of Canada 2015). Funding agencies use “excellence to recognize excellence” (Nowotny, 2014).

The academic funding environment, likewise, is saturated with this discourse. A study of the National Endowment for the Humanities is entitled *Excellence and Equity* (Miller, 2015). The Wellcome Trust, a large medical funder, has grants for “sustaining excellence” (Sustaining Excellence Awards, 2016). The National Institutes of Health (NIH), the largest funder of civilian science in the United States, claims to fund “the best science by the best scientists” (Nicholson and Ioannidis, 2012) and regularly supports “centres of excellence”. The University Grants Commission of India recently awarded 15 institutions the title of “University with Potential for Excellence” (University Grants Commission, 2016). In the United Kingdom, the “Research Excellence Framework” uses expert assessment of “excellence” as a means of channelling differential funding to departments and institutions. In Australia, the national review framework is known as “Excellence in Research for Australia”. In Germany, the Deutsche Forschungsgemeinschaft supports its “Clusters of Excellence” through a long standing “Excellence Initiative” (OECD, 2014).

As this range of examples suggests, “excellence”, as used by universities and their funders, is a flexible term that operates in a variety of contexts across a range of registers. It can describe alike the activities of the world’s top research universities and its smallest liberal arts colleges. It applies to their teaching, research, and management. It encompasses simultaneously the work of their Synthetic Biologists and Urban Sociologists, their Anglo-Saxonists and Concert Pianists. It defines their Centres for Excellence in Teaching and their Centres of Excellence for Mechanical Systems Innovation (The University of Tokyo Global Center of Excellence, 2016; “USC Center for Excellence in Teaching”, 2016), their multiculturalism (Office of Excellence and Multicultural Student Success 2016) and their athletic training programmes (Excellence Academy, 2016). “Excellence” is used to define success in academic endeavour from Montreal to Mumbai.

But what does “excellence” mean? Is there a single standard for identifying this apparently ubiquitous quality? Or is “excellence” defined on a discipline-by-discipline, or case-by-case basis? Can you know “excellence” before you see it? Or is it defined after the fact? Does the search for “excellence”, its use to reward and punish individual institutions and researchers, and its utility as a criterion for the organization of research help or hinder the actual production of that research and scholarship? Tertiary education enrolls approximately 32% of world’s student age population, and OECD countries spent on average 1.6% of their GDP on University-level teaching and research in 2015; the United States alone spent 2.7% or US\$484 billion (*The Economist*, 2015). Is “excellence” really the most efficient metric for distributing the resources available to the world’s scientists, teachers, and

scholars? Does “excellence” live up to the expectations that academic communities place upon it? Is “excellence” excellent? And are we being excellent to each other in using it?

This article examines the utility of “excellence” as a means for organizing, funding, and rewarding science and scholarship. It argues that academic research and teaching is not well served by this rhetoric. Nor, we argue, is it well served by the use of “excellence” to determine the distribution of resources and incentives to the world’s researchers, teachers and research institutions. While the rhetoric of “excellence” may seem in the current climate to be a natural method for determining which researchers, institutions, and projects should receive scarce resources, we demonstrate that it is not as efficient, accurate, or necessary as it may seem. As we show, indeed, a focus on “excellence” impedes rather than promotes scientific and scholarly activity: it at the same time discourages *both* the intellectual risk-taking required to make the most significant advances in paradigm-shifting research *and* the careful “Normal Science” (Kuhn [1962] 2012) that allows us to consolidate our knowledge in the wake of such advances. It encourages researchers to engage in counterproductive conscious and unconscious gamesmanship. And it impoverishes science and scholarship by encouraging concentration rather than distribution of effort. The net result is science and scholarship that is less reliable, less accurate, and less durable than research assessed according to other criteria. While we acknowledge that it often seems politically necessary to argue for “excellence”, and while we understand that funding and accreditation bodies and agencies must play a *political* as well as *scientific* game, we here present the evidence that the internalization of such rhetoric into the research space can be counter-productive.

The article itself falls into three parts. In the first section, we discuss “excellence” as a rhetoric. Drawing on work by Michèle Lamont and others, we argue that “excellence” is less a discoverable quality than a linguistic interchange mechanism by which researchers compare heterogeneous sets of disciplinary practices. In the second section, we dig more deeply into the question of “excellence” as an assessment tool: we show how it distorts research practice while failing to provide a reliable means of distinguishing among competing projects, institutions, or people. In the final section, we consider what it might take to change our thinking on “excellence” and the scarcity it presupposes. We consider alternative narratives for approaching the assessment of research activity, practitioners, and institutions and discuss ways of changing the “scarcity-thinking” that has led us to our current use of this fungible and unreliable term. We propose that a narrative built on “soundness” and “capacity” offers us the opportunity to focus on *practice* of productive research and on the crucial role that social communication and criticism plays. Where there is more heterogeneity and greater opportunity for diversity of outcomes and perspectives, we argue, research improves.

What is “excellence”?

In her book, *How Professors Think: Inside the Curious World of Academic Judgment*, Michèle Lamont opens by noting that “‘excellence’ is the holy grail of academic life” (Lamont, 2009, 1). Yet, as she quickly moves to highlight, this “excellence is produced and defined in a multitude of sites and by an array of actors. It may look different when observed through the lenses of peer review, books that are read by generations of students, current articles published by ‘top’ journals, elections at national academies, or appointments at elite institutions” (3). Or as Jack Stilgoe suggests: “‘Excellence’ tells us nothing about how important the science is and everything about who decides” (Stilgoe, 2014).

This tallies with the work of others who have considered reforms to the review process in recent years. Kathleen Fitzpatrick, for instance, has also situated the crux of evaluation in the evaluator, not the evaluated. For, as Fitzpatrick notes,

“in using a human filtering system, the most important thing to have information about is less the data that is being filtered, than the human filter itself: who is making the decisions, and why. Thus, in a peer-to-peer review system, the critical activity is not the review of the texts being published, but the review of the reviewers.” (Fitzpatrick, 2011, 38)

The challenge here is that it is not possible to conduct a “review of the reviewers” without some reference to the evaluated material. It is possible to query the *conduct* of reviewers or the *process* they are (supposed to be) applying against another set of disciplinary norms (that is, are the reviewers acting in good faith? Have they provided a useful report? Do they know the field as normatively defined?); but to assess qualitative aspects of reviewers’ judgment of a specific work requires an external evaluation of the work itself—a type of circularity in which a *pre-shared* evaluative culture must exist in order to pass judgment on the evaluation that is its basis: the “shared standards” of which Lamont writes (2009: 4).

Yet despite the anti-foundational nature of this problem, there remains a pressing need, in Lamont’s view, to ensure that “peer review processes [... are] themselves subject to further evaluation” (247). Calls for training in peer review practices as well as calls for greater transparency occur across disciplinary boundaries, but generally without addressing the differences in *practice* that occur on either side of those boundaries. Lamont suggests that current remedies to this problem—which mostly consist of changing the degrees of anonymity or the point at which review is conducted (pre- versus post-filter)—are insufficient and constitute “imperfect safeguards”. Instead, she suggests, it is more important that members of peer-review communities should be educated “about how peer evaluation works,” avoiding the pitfalls of homophily (in which review processes merely re-inscribe value to work that exhibits similitude to pre-existing examples) by re-framing the debate as a “micro-political process of collective decision making” that is “genuinely social” (246–247). As with most problems in scholarly communication, the challenge with peer review is therefore not technical but social.

As Lamont and others show, then, “excellence” is a pluralized construct that is specific to (and conservative within) each disciplinary environment. Yet even the most obvious solution to this challenge—interdisciplinary diversity of evaluators—only leads to further problems. For the differences in practice of review and perceptions of “excellence” across disciplinary boundaries, combined with a lack of appreciation that these differences exist, makes it difficult to reach consensus within such diverse pools of reviewers. This is because, as Stirling (2007b) has noted, “it is difficult indeed to contemplate any single general index of diversity that could aggregate properties [...] in a uniquely robust fashion”. If diversity itself cannot easily be collapsed onto a single measurable vector then there is little hope of aggregating diverse senses of “excellence” into a coherent and universal framework.

This suggests that “excellence” resides between different communities and is ill-structured/defined in each context. Local groups and disciplines may have their own more specific (though sometimes conventional rather than explicit) measures of “excellence”: Biologists may treat some aspects of performance as “excellent” (for example, number of publications, author position, citations counts), while failing to recognize aspects considered equally or more “excellent” by English professors

(large word counts, single authorship, publication or review in popular literary magazines and journals) (O’Donnell, 2015). Finally, as we will go on to show, it is clear that evaluative cultures are operating without even internal consensus beyond a few broad categories of performance.

That said, it remains tempting to argue that such concepts of value, even if they are ungrounded and unshared, can be used pragmatically to foster consensus. This is the point of Wittgenstein’s (2001: section 293) famous “beetle in a box” metaphor, which he uses to exemplify the “private language argument”. For Wittgenstein, the question of unique non-communicable epistemic knowledge (such as pain experience), should actually be framed in terms of public, pragmatic language games/contexts. If we each have an object in a box that is called a “beetle,” but none of us can see each other’s “beetles”, he argues, then the important thing is not what the objects in our boxes actually are but rather how we negotiate and use the term socially to engender intersubjective understanding or action. In such cases, “if we construe the grammar of the expression of sensation on the model of ‘object and designation’, the object drops out of consideration as irrelevant” and designation is all that matters.

We might therefore productively ask: even if “excellence” is a concept that carries little or no information content, either within communities or across them, might it nonetheless be useful as a “beetle”? That is, as a carrier of interpretation or a set of social practices functioning as an expert system to convert intrinsic, qualitative, and non-communicable assessment into a form that allows performance to be compared across disciplinary or other boundaries? Might it, indeed, even be useful given the political necessity for research communities and institutions to present an (ostensibly) unified front to government and wider publics as a means of protecting their autonomy? Could “excellence” be, to speak bluntly, a linguistic signifier without any agreed upon referent whose value lies in an ability to capture cross-disciplinary value judgements and demonstrate the political desirability of public investment in research and research institutions?

In actual practice, it is not even useful in this way. Although, as its ubiquity suggests, “excellence” is used across disciplines to assert value judgements about otherwise incomparable scientific and scholarly endeavours, the concept itself mostly fails to capture the disciplinary qualities it claims to define. Because it lacks content, “excellence” serves in the broadest sense solely as an (aspirational) claim of comparative success: that some thing, person, activity, or institution can be asserted in a hopefully convincing fashion to be “better” or “more important” than some other (often otherwise incomparable) thing, person, activity, or institution—and, crucially, that it is, as a result, *more deserving of reward*. But this emphasis on reward, as Kohn (1999) and others have demonstrated, is itself often poisonous to the *actual* qualities of the underlying activity.

Is “excellence” good for research?

Thus far, we have been arguing that “excellence” is primarily a rhetorical signalling device used to claim value across heterogeneous institutions, researchers, disciplines, and projects rather than a measure of intrinsic and objective worth. In some cases, the qualities of these projects can be compared in detail on other bases; in many—perhaps most—cases, they cannot. As we have argued, the claim that a research project, institution, or practitioner is “excellent” is little more than an assertion that that project, institution, or practitioner can be said to succeed better on its own terms than some other project, institution, or practitioner can be said to succeed on some other, usually largely incomparable, set of terms.

But what about these sets of “own terms”? How easy is it to define the “excellence” of a given project, institution, or practitioner on an intrinsic basis? Even if we leave aside the comparative aspect, are there formal criteria that can be used to identify “excellence” in a *single* research instance on its own terms or that of a single discipline?

Research suggests that this is far harder than one might think. Academics, it turns out, appear to be particularly poor at recognizing a given instance of “excellence” when they see it, or, if they think they do, getting others to agree with them. Their continued willingness to debate relative quality in these terms, moreover, creates a basis for extreme competition that has serious negative consequences.

Do researchers recognize excellence when they see it? The short answer is no. This can be seen most easily when different potential measures of “excellence” conflict in their assessment of a single paper, project, or individual. Adam Eyre-Walker and Nina Stoletzki, for example, conclude that scientists are poor at estimating the merit and impact of scientific work even after it has been published (2013). Post-publication assessment is prone to error and biased by the journal in which the paper is published. Predictions of future impact as measured by citation counts are also generally unreliable, both because scientists are not good at assessing merit consistently across multiple metrics and because the accumulation of citations is itself a highly stochastic process, such that two papers of similar merit measured on other bases can accumulate very different numbers of citations just by chance. Moreover, Wang *et al.* (2016) show that in terms of citation metrics the most novel work is systematically undervalued over the time frames that conventional measures use, including, for instance, the Journal Impact Factor that Eyre-Walker and Stoletzki suggest biases expert assessment.

This is true even of work that can be shown to be successful by other measures. Campanario, Gans and Shepherd, and others, for example, have traced the rejection histories of Nobel and other prize winners, including for papers reporting on results for which they later won their recognition (Gans and Shepherd, 1994; Campanario, 2009; Azoulay *et al.*, 2011: 527–528). Campanario and others have also reported on the initial rejection of papers that later went on to become among the more highly cited in their fields or in the journals that ultimately accepted them (Campanario, 1993, 1996; Campanario, 1995; Campanario and Acedo, 2007; Calcagno *et al.*, 2012; Nicholson and Ioannidis, 2012; Siler *et al.*, 2015). Yet others have found a generally poor relationship between high ratings in grant competitions and subsequent “productivity” as measured by publication or citation counts (Pagano, 2006; Costello, 2010; Lindner and Nakamura, 2015; Fang *et al.*, 2016; Meng, 2016).

As this suggests, academics’ abilities to distinguish the “excellent” from the “not-excellent” do not correlate well with one another even within the same disciplinary environment (there tends to be greater agreement at the other end of the scale, distinguishing the “not acceptable” from the “acceptable,” see Cicchetti, 1991; Weller, 2001). To earn citations or win prizes for a rejected manuscript, after all, authors need to begin by convincing a different journal (and its referees) to accept work that others previously have found wanting.

But this is not something that only Nobel prize winners are good at: as Weller reported in the early years of this century, most (51.4%) rejected manuscripts were ultimately published; in the vast majority of cases (approximately 90%), these previously rejected articles were accepted on their second submission and, in the vast majority of these cases (also approximately 90%), at a journal of similar prestige and circulation (Weller, 2001).

While these statistics have almost certainly changed in the last few years with changes in the demographics of submission and, especially, the development of venues that focus on the publication of “sound science” (Public Library of Science, 2016), the basic sense that journal peer review is a gatekeeper that is frequently circumvented remains.

Articles that are initially rejected and then go on to be published to great acclaim or even just in journals of a similar or higher ranking represent what are in essence false negatives in our ability to assess “excellence.” They are also evidence of terrible inefficiency. The rejection of papers that are subsequently published with little or no revision at journals of similar rank increases the costs for everyone involved without any countervailing improvement in quality. In addition to multiplying the systemic cost of refereeing and editorial management by the number of resubmissions, such articles also present an opportunity cost to their authors through lost chances to claim priority for discoveries, for example, or, even more commonly, lost opportunities for citation and influence (Gans and Shepherd, 1994; Campanario, 2009; Şekercioğlu, 2013; Brembs, 2015; Psych Filedrawer, 2016).

More worryingly, there is also considerable evidence of false positives in the review process—that is to say submissions that are judged to meet the standards of “excellence” required by one funding agency, journal, or institution, but do worse when measured against other or subsequent metrics. In a somewhat controversial work, Peters and Ceci submitted papers in slightly disguised form to journals that had previously accepted them for publication (Peters and Ceci, 1982; see Weller, 2001 for a critique). Only 8% overall of these resubmissions were explicitly detected by the editors or reviewers to which they were assigned. Of the resubmissions that were not explicitly detected, approximately 90% were ultimately rejected for methodological and/or other reasons by the same journals that had previously published them; they were rejected, in other words, for being insufficiently “excellent” by journals that had decided they were “excellent” enough to enter the literature previously.

When it comes to funding, a similar pattern of false positives may pertain: a study by Nicholson and Ioannidis (2012) suggests that highly cited authors are *less* likely to head major biomedical research grants than less-frequently-cited but socially better-connected authors who are associated with granting agency study groups and review panels. Fang, Bowen and Casadevall have discovered that “the percentile scores awarded by peer review panels” at the NIH correlated “poorly” with “productivity as measured by citations of grant-supported publications” (Fang *et al.*, 2016). These suggest a bias towards conformance and social connectedness over innovation in funding decisions in a world in which success rates are as low as 10%. It also provides further evidence of funding-agency bias against disruptively innovative work noted by many researchers over the years (Kuhn [1962] 2012; Campanario, 1993, 1995, 1996, 2009; Costello, 2010; Ioannidis *et al.*, 2014; Siler *et al.*, 2015).

Fraud, error and lies. To the extent that the above are evidence of inefficiencies in the system, some might argue that individual problems in determining “excellence” in specific cases are resolved in the longer term and over large samples. Of course, these examples only show work for which multiple measures of “excellence” can be compared: given their unreliability, this suggests that work that is not measured more than once may be unjustly suppressed or unjustly published, without us being able to tell the difference. On the other hand, it is presumably possible that even such extreme examples of differing perceptions of “excellence” represent honest differences of opinion as to the

qualitative merit of the research or researchers. The same cannot be said, however, of actual fraud and outright errors.

As various studies have concluded, reported instances of both fraud and error (as measured through retractions) are on the rise (Claxton, 2005; Dobbs, 2006; Steen, 2011; Fang *et al.*, 2012; Grieneisen and Zhang, 2012; Yong, 2012b; Chen *et al.*, 2013; Andrade, 2016). This is particularly true at higher prestige journals (Resnik *et al.*, 2015; Siler *et al.*, 2015; Belluz, 2016). If we add to this list of (potentially) “false positives” studies that cannot be replicated, the number of papers that meet one measure of “excellence” (that is, passing peer review, often at “top” journals) while failing others (that is, being accurate and reproducible, and/or non-fraudulent) rises considerably (Dean, 1989; Burman *et al.*, 2010; Lehrer, 2010; Bem, 2011; Goldacre, 2011; Yong, 2012b; Rehman, 2013; Resnik and Dinse, 2013; Hill and Pitt, 2014; Chang and Li, 2015; Open Science Collaboration, 2015). It is the very focus on “excellence”, however, that creates this situation: the desire to demonstrate the rhetorical quality of “excellence” encourages researchers to submit fraudulent, erroneous, and irreproducible papers, at the same time as it works to prevent the publication of reproduction studies that can identify such work.

In other words, erroneous, and especially fraudulent or irreproducible papers are interesting because they represent a failure of *both* our ability to identify and predict actual qualitative “excellence” *and* the incentive system that is used to encourage scientists and scholars to produce the kind of sound and defensible work that should be a *sine qua non* for quality. As Fang, Steen, and Casadevall (2012; cf Steen, 2011 for which the later article represents a correction) have shown, the majority of retracted papers are withdrawn for reasons of misconduct including fraud, duplicate publication, or plagiarism (67.4%), rather than error (21.3%)—although inadvertent error should presumably itself be disqualification from “excellence”. But even these figures may under-represent the true incidence of misconduct. Mistakes and errors made in good faith are a natural and necessary part of the research process. Yet, as focus groups and surveys conducted by various researchers have demonstrated, some forms of error can be misconduct in the form of a (semi-) deliberate strategy for ensuring quick and/or numerous publications by “cutting a little corner” in order to get a paper out before others or to get a larger grant... [or] because... [a researcher] needed more publications that year” (Anderson *et al.*, 2007: 457–458; see also Fanelli, 2009; Tjldink *et al.*, 2014; Chubb and Watermeyer, 2016).

Thus in one small sample of detailed surveys, Fanelli showed that while only a small percentage of scientists (1.97% pooled weighted average, $n=7$) admitted to fabricating, falsifying, or modifying data, a much larger percentage claimed to have seen others engaging in similarly outright fraudulent activity (14.12%, $n=12$). Furthermore, even larger percentages had engaged in (33.7%) or seen others engage in (72%) questionable research described using less negatively loaded language (Fanelli, 2009; the percentage of scientists admitting to explicit misconduct is considerably higher [15%] in Tjldink *et al.*, 2014). As Fanelli concludes: “Considering that these surveys ask sensitive questions and have other limitations, it appears likely that this is a conservative estimate of the true prevalence of scientific misconduct” (2009, 9)—a conclusion very strongly supported by the anecdotal admissions of Anderson *et al.*’s focus groups.

The drive for “excellence” in the eyes of assessors is shown even more starkly in work by Chubb and Watermeyer (2016). In structured interviews, academics in Australia and the United Kingdom admitted to outright lies in the claims of broader impacts made in research proposals. As the authors note: “Having to sensationalize and embellish impact claims was seen to have become a normalized and necessary, if regretful, aspect of

academic culture and arguably par for the course in applying for competitive research funds” (6). Quoting an interviewee, they continue, “If you can find me a single academic who hasn’t had to bullshit or bluff or lie or embellish to get grants, then I will find you an academic who is in trouble with his [sic] Head of Department” (6; “[sic]” as in Chubb and Watermeyer). Here we see how a competitive requirement, perceived or real, for “excellence”, in combination with a lack of belief in the ability of assessors to detect false claims, leads to a conception of “excellence” as pure performance: a concept defined by what you can get away with claiming in order to suggest (rather than actually accomplish) “excellence”.

What is striking about these behaviours, of course, is that they are unrelated to (and to a great extent perhaps even incompatible with or opposed to) the actual qualities funders, governments, journal editors and referees, and researchers themselves are ostensibly using “excellence” to identify. No agency, ministry, press, or research office intentionally uses “excellence” as shorthand for “able to embellish results or importance convincingly”, even as the researchers being adjudicated under this system report such embellishment as a primary criterion for success. Whether it occurs through fraud, cutting corners, or exaggeration, this performance of “excellence” is commonly justified as being necessary for survival, suggesting a cognitive and cultural dissonance between those aspects of their work that the performers feel is essential and those aspects they feel they must emphasise, overstate, embellish, or fabricate to appear more “excellent” than their competitors. The evidence that fraud and corner-cutting are a problem at the core of the research process suggests that the pressure for these performances of “excellence” is not restricted to stages that do not matter. As Kohn argues, reward-motivation affects scientific creativity (the ability to “break out of the fixed pattern of behaviour that had succeeded in producing rewards... before”) as much as it does evidence-gathering or the inflation of results (1999, 44; see also Lerner and Wulf, 2006; Azoulay *et al.*, 2011; Tian and Wang, 2011).

Competition for scarce resources and the performance of “excellence”. So why do researchers engage in this kind of dubious activity? Clearly for both Chubb and Watermeyer’s interviewees, as well as those identified as having committed scientific fraud, it is competition for scarce resources, whether funding, positions, or community prestige. Of course this is not a new issue (Smith, 2006). Taking time away from his work on the difference machine, Charles Babbage published an analysis of what he saw as the four main kinds of scientific frauds in an 1830 polemic, *Reflections on the Decline of Science in England: And on Some of Its Causes*. These included the self-explanatory “hoaxing” and “forging”, in addition to “trimming” (“clipping off little bits here and there from those observations which differ most in excess from the mean and in sticking them on to those which are too small”) and “cooking” (“an art of various forms, the object of which is to give ordinary observations the appearance and character of those of the highest degree of accuracy”) (Babbage, 1831: 178; see Zankl, 2003; and Secord, 2015 for a discussion).

The motivation for these frauds, then as now, involves prestige and competition for resources. Babbage’s typology of fraudulent science was but a minor chapter in a book otherwise mostly concerned with the internal politics of the Royal Society. He attributed the decline he saw in English science to the lack of attention and professional opportunities available to potential scientists. He was, as a result, keenly sensitive to questions of credit and its importance in determining rank and authority. Indeed, as Casadevall and Fang remind us, “Since Newton, science has changed a great deal, but this basic fact has not.

Credit for work done is still the currency of science.... Since the earliest days of science, bragging rights to a discovery have gone to the person who first reports it" (Casadevall and Fang, 2012: 13). The prestige of first discovery always has been a scarce resource. Now that that prestige is measured also through the scarce resource of authorship in "the right journals" and coupled ever more strongly to the further scarce resources of career advancement and grant funding, it should not be a surprise that the competition for those markers has become steadily stronger. The performance of "excellence" has become more marked as a result.

If scandals such as fraudulent articles were the only way in which this overwhelming competitive focus on "excellence" hurt research, it would be bad enough. But the emphasis on rewarding the performance of "excellence" also has a more general impact on research capacity: it is the mechanism by which "the Matthew effect"—that is, the disproportionate accrual of resources to those researchers and institutions that are already well-rewarded—operates in a hyper-competitive research environment, creating distortions throughout the research cycle, even for work that is not fraudulent or the result of misconduct (Bishop, 2013; as its etymology implies, the "Matthew effect" predates today's hypercompetition, see Merton, 1968, 1988)¹: it increases the stakes of the competition for resources and, as a result, encourages gamesmanship; creates a bias towards (non-disruptively) novel, positive, and even inflated results on the part of authors and editors; and discourages the pursuit and publication of types of "Normal Science" (such as replication studies) that are crucial to the viability of the research enterprise, without being glamorous enough to suggest that their authors are "excellent".

Positive bias and the decline effect. Just how destructive this need to perform "excellence" is can be illustrated by the well-known bias towards positive results in scientific publication (for example, Dickersin *et al.*, 1987, 2005; Sterling, 1959; Kennedy, 2004; Young and Bang, 2004; Bertamini and Munafò, 2012; Rothstein, 2014; Psych Filedrawer, 2016). Thus, for example, Fanelli (2011) demonstrated a 22% growth between 1990 and 2007 in the "frequency of papers that, having declared to have 'tested' a hypothesis, reported a positive support for it". This is all the more remarkable given that the late 1980s were themselves not a halcyon period of unbiased science: in an 1987 study of 271 unpublished and 1041 published trials, Dickersin *et al.* found that 14% of unpublished and 55% of published trials favoured the experimental therapy (1987). As Young *et al.* suggest, "the general paucity in the literature of negative data" is such that "[i]n some fields, almost all published studies show formally significant results so that statistical significance no longer appears discriminating" (2008, 1419).

Another artifact of this positive bias is the "decline effect," or the tendency for the strength of evidence for a particular finding to decline over time from that stated on its first publication (Schooler, 2011; Gonon *et al.*, 2012; Brembs *et al.*, 2013; Groppe, 2015; Open Science Collaboration, 2015). While this effect is also well-known, Brembs *et al.* have recently shown that its presence is significantly positively correlated with journal prestige as measured by Impact Factor: early papers appearing in high prestige journals report larger effects than subsequent studies using smaller samples (2013, see Figs. 1b and 1c in this reference).

The bias against replication. Finally, there is a bias against the publication of replication studies in disciplines where such patterns make scientific sense. Indeed, there are currently insufficient structural incentives to perform work that "merely" revalidates

existing studies, fuelled by a focus on novelty in most definitions of "excellence". As Nosek *et al.* note

Publishing norms emphasize novel, positive results. As such, disciplinary incentives encourage design, analysis, and reporting decisions that elicit positive results and ignore negative results. Prior reports demonstrate how these incentives inflate the rate of false effects in published science. When incentives favour novelty over replication, false results persist in the literature unchallenged, reducing efficiency in knowledge accumulation. (2012)

This bias against replication is even more remarkable, however, when it involves studies that invalidate rather than confirm the original result, especially when the original result has a high profile or is potentially field-defining—qualities that one would assume would increase the novelty and interest of the (non) replication itself (Goldacre, 2011; Wilson, 2011; Nosek *et al.*, 2012; Yong, 2012a, b; Aldhous, 2011; for a view from the other side of replication, see Bissell, 2013). This is in part, a function of publishing economics: commercial journals earn money from subscription, access, and reprint fees (Lundh *et al.*, 2010); high profile results and a high prestige reflected by a high Impact Factor help maintain the demand for these journals and hence ensure both a continuing stream of interesting new material and a steady or rising income for the journal as a whole (Lawrence, 2007; Munafò *et al.*, 2009; Lundh *et al.*, 2010; Marcovitch, 2010). Undercutting (or perhaps even qualifying) the high-profile results that help bring in these subscribers, new articles, and attention attacks the very foundation of this success—a journal that publishes high profile but incorrect papers is undercutting its case for subscription and author submissions. One doesn't need to imagine a conspiracy to promote poor science to understand how a conscious or unconscious bias against replication studies might arise under such circumstances.

The reluctance of major journals to publish replication studies embeds this bias in the incentive system that guides authors. As Wilson notes:

[M]ajor journals simply won't publish replications. This is a real problem: in this age of Research Excellence Frameworks and other assessments, the pressure is on people to publish in high impact journals. Careful replication of controversial results is therefore good science but bad research strategy under these pressures, so these replications are unlikely to ever get run. Even when they do get run, they don't get published, further reducing the incentive to run these studies next time. The field is left with a series of "exciting" results dangling in mid-air, connected only to other studies run in the same lab. (2011)

As Rothstein (2014) argues "The consequences of this problem include the danger that readers and reviewers will reach the wrong conclusion about what the evidence shows, leading at times to the use of unsafe or ineffective treatments".

Homophily. Thus far, we have been discussing the negative impact of "excellence" largely in terms of its effect on the practice and results of professional researchers. There is, however, another effect of the drive for "excellence": a restriction in the range of scholars, of the research and scholarship performed by such scholars, and the impact such research and scholarship has on the larger population. Although "excellence" is commonly presented as the most fair or efficient way to distribute scarce resources (Sewitz, 2014), it in fact can have an impoverishing effect on the very practices that it seeks to encourage. A funding programme

that looks to improve a nation's research capacity by differentially rewarding "excellence" can have the paradoxical effect of reducing this capacity by underfunding the very forms of "normal" work that make science function (Kuhn [1962] 2012) or distract attention from national priorities and well-conducted research towards a focus on performance measures of North America and Europe (Vessuri *et al.*, 2014). A programme that seeks to reward Humanists, similarly, by focussing on output in "high impact" academic journals paradoxically reduces the impact of these same disciplines by encouraging researchers to focus on their professional peers rather than broader cultural audiences (Readings, 1996), reducing the domain's relevance even as its performance of "excellence" improves. A programme of concentration on the "best" academics, in other words, can have the effect of focussing attention on problems and approaches in which "excellence" can be performed most easily rather than those that could benefit the most (or provide the greatest actual impact) from increased attention.

Moreover, a concentration on the performance of "excellence" can promote homophily among the scientists themselves. Given the strong evidence that there is systemic bias within the institutions of research against women, under-represented ethnic groups, non-traditional centres of scholarship, and other disadvantaged groups (for a forthright admission of this bias with regard to non-traditional centres of scholarship, see Goodrich, 1945), it follows that an emphasis on the performance of "excellence"—or, in other words, being able to convince colleagues that one is even more deserving of reward than others in the same field—will create even stronger pressure to conform to unexamined biases and norms within the disciplinary culture: challenging expectations as to what it means to be a scientist is a very difficult way of demonstrating that you are the "best" at science; it is much easier if your appearance, work patterns, and research goals conform to those of which your adjudicators have previous experience. In a culture of "excellence" the quality of work from those who do not work in the expected "normative" fashion run a serious risk of being under-estimated and unrecognised (King *et al.*, 2014, 2016; O'Connor and O'Hagan, 2015; University of Arizona Commission on the Status of Women, 2015; this is, in part, an explanation for the systemically underreported and poorly acknowledged and rewarded work of women "assistants" in many of the great scientific discoveries of the twentieth century). There is a clear case to answer that, absent substantial corrective measures and awareness, a focus on "excellence" will continue to maintain rather than work to overcome social barriers to participation in research by currently underrepresented groups.

Homophily is in some senses a variant on Merton's "Matthew effect," discussed above. It is also a variant on the old argument that existing power structures—those populated by those whom it is assumed already exemplify "excellence"—tend towards conservatism in their processes of evaluation. It underpins the calls to reassess the focus of mainstream scholarship, whether this is "great men" history, the "Dead White Male" in literary "canon", or the bias towards the ills of the western male patient in medical research. As Barbara Herrnstein Smith says with respect to literary evaluation:

...[a work that "endures"] will also also begin to perform certain characteristic cultural functions by virtue of the very fact that it *has* endured...In these ways, the canonical work begins increasingly not merely to *survive within* but to *shape and create* the culture in which its value is produced and transmitted and, for that very reason, to perpetuate the conditions of its own flourishing. (Herrnstein Smith, 1988 emphasis in the original)

In other words, the works that—and the people who—are considered "excellent" will always be evaluated, like the canon that shapes the culture that transmits it, on a conservative basis: past performance by preferred groups helps establish the norms by which future performances of "excellence" are evaluated. Whether it is viewed as a question of power and justice or simply as an issue of lost opportunities for diversity in the cultural co-production of knowledge, an emphasis on the performance of "excellence" as *the* criterion for the distribution of resources and opportunity will always be backwards looking, the product of an evaluative process by institutions and individuals that is established by those who came before and resists disruptive innovation in terms of people as much as ideas or process.

Alternative narratives: working for change

If, as we have argued, "excellence" in all its many forms and meanings is both unreliable as a measure of actual quality, and pernicious in the way it promotes poor behaviour and discourages good, what then are the alternatives? Given the political realities that have promoted the use of this rhetoric in defence of science and scholarship, are there other, less damaging ways in which we can evaluate and promote the value of research and its communication?

Because "excellence" is used so ubiquitously across the research space, a complete answer to this question is far beyond the scope of any single paper: there is no single alternative that can replace the rhetoric of "excellence" in scholarly publishing, research funding, government and university policy, public relations, and promotion and tenure practices. In some areas, moreover, technological and economic changes suggest fairly obvious directions in which progress is being made—a prime example being the change from the physical scarcity that characterized print journals, adjudication to the abundance that, technically at least, characterizes a web-based publication infrastructure (for well-known discussions of this, see Shirky, 2010; Nielsen, 2012).

In many ways, however, the greatest challenge is research funding and infrastructure. The continuing competition for government and private funds raises questions of prioritization and adjudication that are unlikely to be rapidly answered by changes in technology or attitudes. A central test of our critique of rhetorics of "excellence" is therefore to ask whether there are any alternatives in this arena. Since funding applications tend to collect examples of "excellence" from other aspects of the research enterprise as a form of justification (success in funding is a function of one's ability to demonstrate "excellence" in different types of performance), it also represents the apex of the problem.

Perhaps because it is so hard, the tendency in policy, at least in the traditional North Atlantic centres of research in the last several decades, has clearly been in a non-distributive direction: for the concentration of resources on "top" institutions (in earlier periods, such as the early space race, for example, the focus was arguably more distributive). The Research Excellence Framework in the United Kingdom (REF) and massive new research centres such as the Crick in London are intended to create a "critical mass" of "excellent" or "world-leading" research. In Canada, which is an outlier internationally in the push towards stratification (Usher, 2016), it remains the case that the "top" universities (which have their own independent lobby group), receive a disproportionate share of research resources when measured, for example, against the percentage of students (including Doctoral students) they educate (U15 Group of Canadian Research Universities/Regroupement des universités de recherche du Canada, 2016). In the much larger U.S. post secondary system, ten universities received nearly 20% of all

government research funds; as Weigley and Hess note, while these universities are among the richest in the country in terms of their endowments, public funding still constitutes the largest part of their R&D funding (2013).

Many have questioned the value of such an inequitable distribution of funds when a less concentrated, or less unequal, distribution could achieve greater outcomes. Dorothy Bishop argues, with respect to the REF that there should be less of a disparity between rewarding research that is perceived to be “the best” and that which is perceived as merely average. Instead, Bishop (2013) argues, all research submitted to the REF should receive *some* funding and the perceived best research should receive a smaller overall proportionate gain. This would have the benefit of decreasing the funding gulf between elite and middle-tier universities and would encourage diversity in the process. Of course such an approach may be politically troublesome for the academy, as long as the criterion it promotes is relative “excellence” rather than, say, “capacity”, “breadth”, “soundness”, “comprehensiveness” or “accessibility”. If funding is allocated on a scattered basis, following the logic that predictive approaches to quality are weak at best, then the authority claims of the university are substantially devalued as long as the rhetoric used to defend them privileges a “winner-take-all” measure of effectiveness.

There is, however, a compelling case to be made for the value of greater redistribution of research funding. Cook *et al.* (2015) showed that for UK Bioscience groups an optimal allocation of fixed resources would involve spreading the money between a larger number of smaller groups. This was the case whether number of publications or number of citations were used as the measure of productivity. A similar conclusion is reached by Fortin and Currie who argue that scientific impact is only “weakly money-limited” and that a more productive strategy would be to distribute funds based on “diversity” rather than perceptions of “excellence” (Fortin and Currie, 2013). Gordon and Poulin argued that, for science funding in Canada through the National Science and Engineering Research Council (NSERC, the main STEM funding agency), it would have cost less at a whole system level simply to distribute the average award to all eligible applicants than to incur the costs associated with preparing, reviewing and selecting proposals (2009; although see Roorda, 2009 for a critique of their calculation). A rough calculation of the system costs of preparing failed grant applications would suggest that they are in the same order of magnitude as research grant funding itself (Herbert *et al.*, 2013).

What this suggests is that “excellence” is not the only policy choice concerning the resourcing of research, nor even, necessarily, the only politically compelling one: from concentrating resources on the most deserving, allegedly “excellent”, institutions and researchers, to distributing them amongst all those that meet some minimum criteria—or even some subset, by lottery (Health Research Council of New Zealand, 2016; Fang *et al.*, 2016), arguments can be made for a variety of different methods of funding research. In the context of scarce resources and a desire to maximize outcomes, indeed, there is even an argument for focussing most attention on the *worst* institutions; those that might most benefit from resources to improve (Bishop, 2013), have the greatest scope for improvement, and would go the longest way to ensuring an increase in basic capacity. In this case, rather than “excellence” appraisers would be looking for some sort of baseline level of qualification, “credibility” (Morgan, 2016), perhaps, or “soundness”. This would be a shift from focussing on evaluation of *outputs* to an evaluation of *practice*.

The challenge with any redistributive scheme is how to engage with politics. While proposing interesting and valuable thought experiments, they do not address the needs of working with

governments who need to account for the distribution of public funds and may fear the optics of a system built on criteria other than “the best”. The narrative and the need for “excellence” (like that of “international competitiveness”) is important as a shared language of externally recognizable symbols that justify funding to government and to wider publics.

As noted earlier, this serves the interests of those who have already “earned” the label. The local construction of “excellence” is inherently conservative, and maintaining its structures serves the interests of those who hold local power. Therefore, narratives arguing for redistribution need to be more than just interesting ideas and more than simply factually correct. They need to be politically as well as intellectually compelling.

Soundness and capacity over “excellence”. This is where a rhetoric built around “soundness” and “capacity” offers opportunities. The idea that “sound research is good research”, and “more research is better than less”—that our focus should be on thoroughness, completeness, and appropriate standards of description, evidence, and probity rather than flashy claims of superiority—presents an alternative to the existing notions of “excellence”. Such a narrative also addresses deeper concerns regarding a breakdown in research culture through hypercompetition. These terms resonate with public and funder concerns for value, and they align with the need for improved communications and wider engagement encouraged by many governments and agencies.

It might be argued in the case of “soundness” in particular that the term is as subjective as “excellence”. Stirling (2007a) has argued that the implication that expert analysis can be free from subjective values in determining something like “soundness” is itself misleading and exclusionary. Certainly “soundness” or “scientificness” rhetorics have been used to give credibility to controversial technologies and to shut a range of perspectives out of public discourse in ways that are similar to uses of “excellence” we have criticized.

But the evaluation of “soundness” is based in the practice of scholarship, whereas “excellence” is a characteristic of its objects (outputs and actors). In this sense “soundness” aligns well with approaches that locate the value of scholarship and evaluation in the nature of its processes (that is, “proper practice”) and its social conduct. While disagreeing on what the outputs of research can actually *mean*, scholars from Fleck, through Merton, Kuhn, Ravetz and Latour have all focussed on how practice in a social context in which norms and ethics are sustained and enforced leads to productive scholarship (Fleck [1935] 1979; Ravetz, 1973; Latour and Woolgar, 1986; Latour, 1987). “Soundness” can be assessed by how it supports socially developed and documentable processes and norms. In contrast assessment of “excellence” depends on how convincing the performance of importance and impact is. Like “excellence” the criteria for “soundness” are not universal qualities distinct from pre-existing socially developed practice; but in contrast to “excellence”, the qualities of “soundness” *can* be benchmarked. They are also more precise: “excellence” in the senses we are discussing is used describe the competitive position of an entire performance in relation to others; “soundness” focusses on details: statistical or bibliographic appropriateness, say, or well-chosen evidence.

Another question about “soundness” involves its cross-disciplinary application. What is “soundness” in the context of the Humanities? Eve (2014, 144) has suggested that “soundness” in a humanities paper might involve the ability to “evinced an argument; make reference to the appropriate range of extant scholarly literature; be written in good, standard prose of an appropriate register that demonstrates a coherence of form and

content; show a good awareness of the field within which it was situated; pre-empt criticisms of its own methodology or argument; and be logically consistent". More recently, Morgan (2016) has suggested that "credibility" may be the humanities equivalent of "soundness". Others have focussed on the term "quality" in the sense in which it used in quality assurance (Funtowicz and Ravetz, 1990; Funtowicz and Ravetz, 2003), as fitness for an explicitly defined purpose. As we have argued above all of these appear to capture the sense that productive scholarship can be defined by allegiance to socially defined research practice as much as performance of success.

Our argument here is not that expanding our boundary for resourcing from "excellence" to "soundness" and "capacity" is all that is necessary to change research culture and improve the distribution of resources; rather, it is that a move from resourcing based on the performance of an ineluctable quality to one based on the demonstration of documentable, socially developed practice, is the first step to solving the problems our rhetoric of "excellence" has created. Soundness appears to be a plausible basis on which to build a new narrative, or rather to combine existing threads into a more consistent rhetorical framework. Such a framework will work to refocus our attention on research that is sufficiently valuable to be worth pursuing. To drive adoption and practice towards making this real, however, will require more than narrative. It will need resources to be redistributed towards supporting a broader class of research activities.

Do soundness and capacity sell? Although we have been focussing on funding, the rhetoric of soundness and capacity, about the idea that the most important quality of research is that it be done and done with care, does resonate with other aspects of the research enterprise.

Some examples of this are the broad area of reproducibility (Burman *et al.*, 2010; Lehrer, 2010; Goldacre, 2011; Yong, 2012b; Rehman, 2013; Chang and Li, 2015; Open Science Collaboration, 2015), reporting guidelines for animal experiments (Kilkenny *et al.*, 2010) and clinical trials (Schulz *et al.*, 2010), and work on registered replication studies in social psychology (Simons *et al.*, 2014). All have been areas of substantial professional and popular discussion and the emphasis on the need for clarity of description and "doing things properly" is consistent. The idea that research must be reproducible, safe, and complete can be at least as compelling an argument as that it must be simply excellent.

Another place where the rhetoric of "soundness" and "capacity" has booked considerable success is the online journal PLOS ONE and the journals that have since begun to follow its approach.² PLOS ONE was launched with the stated aim of publishing any scientific research that was deemed technically sound, regardless of its perceived novelty or impact. This approach was made possible by two developments in academic publishing—the move to fully online publications without the need for print editions, and the growing acceptance of Article Processing Charge (APC)-funded Open Access as a viable publication model. These enabled the journal to consider and publish any manuscript that met its criteria, with no limitations on page space or fixed subscription revenue. As a result, the journal grew very quickly, becoming the largest journal in the world within 5 years of launching (MacCallum, 2011).

The PLOS ONE model has been widely emulated, with almost every major scientific publisher now offering a journal with similar editorial criteria. This has created a competitive landscape with interesting properties. Traditional journals compete by seeking to publish the most "excellent" papers that they can attract and demonstrate this by the number of papers they reject. This also leads authors to self-select for submission to those

journals only the papers they consider most important—avoiding, for example, "wasting" anybody's time by submitting "non-original" work such as replication studies. Over time, success in this venture, its own form of hypercompetition, leads to a differentiated set of ranked journals driven by their own performative targets, or aspirations to join the top ranks. Authors and editors engage in a cycle of performance that reduces the breadth of research journals are willing to publish and authors willing to submit.

PLOS ONE and its competitors also compete, but on quite different terms and in ways that arguably improve rather than imperil the research enterprise. Speed of publication, for example, always features in author surveys, and journals like PLOS ONE often advertise their average turnaround times. They even compete on the basis of journal prestige, reputation and Impact Factor (Solomon, 2014), albeit with a heavier emphasis on soundness and number of publications (that is, capacity) rather than exclusivity and "excellence". Even when the criteria for inclusion is only soundness, membership in the club of authors still provides a prestige benefit: that the doors of the club are more open does not necessarily mean that there is no benefit to membership (Potts *et al.*, 2016).

But PLOS ONE and similar journals also demonstrate that it is not simply enough to create mechanisms that test for soundness and capacity. Even when offered a distributive narrative, researchers often still find it difficult to avoid the concentrating rhetoric of "excellence". A common complaint from the managers of journals such as PLOS ONE, indeed, is that their journals' referees, who are usually made up of previous authors, often seek to reject papers that they feel do not meet their own perceptions of "excellence," instead of focussing on the journal's formal criterion of "soundness". Many anecdotes from PLOS ONE authors, likewise, involve being surprised by how tough the refereeing process was for their articles—a response that signals relative "excellence" that might otherwise not be apparent to the reader (see especially Curry, 2012 and comments). The performance of "excellence", the signalling of relative superiority through an additional line on the CV, is still more important from a career perspective than the science itself: nobody gets tenure for publishing to arXiv, no matter how good the quality of their research. At least that appears to be what most tenure-track academics believe. And while reader attention or online conversation are gaining some currency as indicators of qualities valued in an article, the current discourse indicates that authors need to feel that they have cleared a higher bar than they in fact have.

In other words, initiatives like PLOS ONE will have truly succeeded in changing researchers' own bias towards (ultimately undemonstrable) "excellence" only when their rejection rate is seen to be less important than the evidence that controls are in place to ensure and encourage the recognition of "soundness".

Caveats and further work

The potential scope of the project of this article is huge, and we have only been able to touch on some of its aspects. We have focused on narratives and rhetoric and sought to bring evidence of how existing rhetorics are damaging. What we have not done, as a variety of both anonymous reviewers and non-anonymous commenters have noted, is address the power politics that underlie many of the structures that we are critiquing. Nor have we analysed the degree to which different actors within the system are able to enact change.

Understanding how the changes we propose in narrative and indeed culture can be achieved politically and institutionally is a much larger project, one on which others are already engaged

and one that is critically important in the current political climate. Institutional change is challenging and slow. We hope that alongside the criticism, implicit and explicit of some existing institutions, we have offered some routes forward to be investigated and explored.

We have also not undertaken a historical analysis. While we draw on literature from a range of periods we have not addressed how and when our current narratives developed. While we would argue that it has deep roots, we have neither the expertise nor the space to probe the history through which excellence rhetorics became institutionalized in their current forms. The differing registers and locations of excellence rhetorics over time—poling access to the right clubs, publication in the right journals, career success and contributions to institutional funding—is deserving of further study and would additionally strengthen the political analysis.

Closing the loop: planning for cultural change

In this article, we have advanced an argument that “excellence” is not just unhelpful to realising the goals of research and research communities but actively pernicious. A narrative of scarcity combined with “excellence” as an interchange mechanism leads to concentration of resources and thence hypercompetition. Hypercompetition in turn leads to greater (we might even say more shameless, see Anderson *et al.*, 2007; Fanelli, 2009; Tjink *et al.*, 2014; Chubb and Watermeyer, 2016) attempts to perform this “excellence”, driving a circular conservatism and reification of existing power structures while harming rather than improving the qualities of the underlying activity.

We have also argued that, while many commentaries reviewed throughout this piece lay the blame for this at the feet of external actors—institutional administrators captured by neo-liberal ideologies, funders over-focussed on delivering measurable returns rather than positive change, governments obsessed with economic growth at the cost of social or community value—the roots of the problem in fact lie in the *internal* narratives of the academy and the nature of “excellence” and “quality” as supposedly shared concepts that researchers have developed into shields of their autonomy. The solution to such problems lies not in arguing for more resources for distribution via existing channels as this will simply lead to further concentration and hypercompetition. Instead, we have argued, these internal narratives of the academy must be reformulated.

Finally, we have argued for a more pluralistic approach to the distribution of resources and credit. Where competition does take place it should do so on the basis of the many different *qualities*, plural, that are important to different communities using and creating research. But it should also be recognized that competition is not, in this context, an unalloyed good. In the context of assessing the risks of application of research Stirling and others argue for “broadening out *and* opening up” the technology assessment process (Ely *et al.*, 2014, see also Stilgoe, 2014), that is to say increasing both the set of criteria considered and the range of people who have a voice in its assessment and application. The same approach needs to be applied to research assessment itself.

This leads to our argument for a focus on redistribution instead of concentration, which, we suggest, is necessary for three core reasons. Firstly because “excellence” cannot be recognized or defined consensually, except as a Wittgensteinian “beetle in a box” that no-one has ever seen, and even then, unlike Wittgenstein’s beetle-owners, by researchers who cannot agree even within disciplinary communities on which aspects of “excellence” might matter or be useful. Second because, as we have argued, there is a case to be made for redistribution on its

own merits. Unlike concentration, and the hypercompetition to which it leads, which break down our standards and cultures in systematic, predictable, and negative ways, redistribution enhances capacity and breadth of participation. And thirdly, we have shown that top-loading of research funding based upon anti-foundational principles of “excellence” is likely to hurt the incremental advances upon which research implicitly relies.

The argument for redistribution is a challenging one to advance. The rhetorics of scarcity, of concentration and competition are linked to strong cultural and economic narratives, particularly in the United Kingdom and United States. But as a route towards this goal we have argued that it is possible to build upon existing narratives of “soundness”, “credibility” and “capacity”—which is to say on narratives of reproducibility, transparency, high-quality reporting, and a breadth and diversity of activity—to build a case for strong cultural practices that focus on fundamental standards that define proper scholarly and scientific practice. This focus on the practice *of* research, including its communications, rather than the performance of success *at* research can also be aligned with developing narratives of Responsible Research and Innovation and public engagement. For instance the approach of Post-Normal Science advocated by Funtowicz and Ravetz (2003; 1990), focuses on assessing the quality of the process of research practice, and emphasises the need to effectively communicate the weaknesses of any claims made on the basis of research.

In taking this approach we root the discourse in long-standing traditions and culture, while also engaging with the newer concerns. It is through showing that we *can* recognize sound and credible research and that we *can* build strong cultures and communities around that recognition, that we lay the groundwork for making the case for redistribution. And *that* would be excellent.

Notes

- 1 The name of the Matthew Effect is derived from *Matthew* 13:12: “For whosoever hath, to him shall be given, and he shall have more abundance: but whosoever hath not, from him shall be taken away even that he hath”.
- 2 As noted in the disclosure of competing interests, three of the authors of this article have worked for *PLOS* previously.

References

- Aldhous P (2011) Journal Rejects Studies Contradicting Precognition. *New Scientist*, 11 May, <https://www.newscientist.com/article/dn20447-journal-rejects-studies-contradicting-precognition/>, accessed 19 February.
- Alpher RA, Bethe H and Gamow G (1948) The origin of chemical elements. *Physical Review*; 73 (7): 803–804.
- Anderson MS, Ronning EA, De Vries R and Martinson BC (2007) The perverse effects of competition on scientists’ work and relationships. *Science and Engineering Ethics*; 13 (4): 437–461.
- Andrade R de O (2016) Sharp Rise in Scientific Paper Retractions. *University World News*, 8 January <http://www.universityworldnews.com/article.php?story=20160108194308816>.
- Azoulay P, Zivin JSG and Manso G (2011) Incentives and creativity: Evidence from the academic life sciences. *The Rand Journal of Economics*; 42 (3): 527–554.
- Babbage C (1831) Reflections on the Decline of Science in England: And on Some of Its Causes, by Charles Babbage (1830). To Which Is Added On the Alleged Decline of Science in England, by a Foreigner (Gerard Moll) with a Foreword by Michael Faraday (1831). B. Fellowes: London.
- Belluz J (2016) Do ‘Top’ Journals Attract ‘Too Good to Be True’ Results? *Vox*. 11 January, <http://www.vox.com/2016/1/11/10749636/science-journals-fraud-retractions>.
- Bem D (2011) Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*; 100 (3): 407–425.
- Bertamini M and Munafò MR (2012) Bite-size science and its undesired side effects. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*; 7 (1): 67–71.

- Bishop D (2013) The Matthew Effect and REF2014. *BishopBlog*, <http://deevybee.blogspot.ca/2013/10/the-matthew-effect-and-ref2014.html>, accessed 15 October.
- Bissell M (2013) Reproducibility: The risks of the replication drive. *Nature*; **503** (7476): 333–334.
- Brembs B (2015) The cost of the rejection-resubmission cycle. *The Winnower*. doi:10.15200/winn.142497.72083.
- Brembs B, Button K and Munafò M (2013) Deep impact: Unintended consequences of journal rank. *Frontiers in Human Neuroscience*; **7**, 291.
- Burman LE, Reed WR and Alm J (2010) A call for replication studies. *Public Finance Review*; **38** (6): 787–793.
- Calcagno V, Demoinet E, Gollner K, Guidi L, Ruths D and de Mazancourt C (2012) Flows of research manuscripts among scientific journals reveal hidden submission patterns. *Science*; **338** (6110): 1065–1069.
- Campanario JM (1993) Consolation for the scientist: Sometimes it is hard to publish papers that are later highly cited. *Social Studies of Science*; **23**, 342–362.
- Campanario JM (1995) Commentary on influential books and journal articles initially rejected because of negative referees' evaluations. *Science Communication*; **16** (3): 304–325.
- Campanario JM (1996) Have referees rejected some of the most-cited articles of all times? *Journal of the American Society for Information Science*; **47** (4): 302–310.
- Campanario JM (2009) Rejecting and resisting Nobel class discoveries: Accounts by Nobel Laureates. *Scientometrics*; **81** (2): 549–565.
- Campanario JM and Acedo E (2007) Rejecting highly cited papers: The views of scientists who encounter resistance to their discoveries from other scientists. *Journal of the American Society for Information Science and Technology*; **58**, 734–743.
- Casadevall A and Fang FC (2012) Winner takes all. *Scientific American*; **307** (2): 13.
- Chang AC and Li P (2015) Is Economics Research Replicable? Sixty Published Papers from Thirteen Journals Say 'Usually Not.' 2015-083. Finance and Economics Discussion Series. Washington DC: Board of Governors of the Federal Reserve System. <http://www.federalreserve.gov/econresdata/feds/2015/files/2015083pap.pdf>.
- Chen C, Hu Z, Milbank J and Schultz T (2013) A visual analytic study of retracted articles in scientific literature. *Journal of the American Society for Information Science and Technology*; **64** (2): 234–253.
- Chubb J and Watermeyer R (2016) Artifice or integrity in the marketization of research impact? Investigating the moral economy of (pathways to) impact statements within research funding proposals in the UK and Australia. *Studies in Higher Education*; 1–13.
- Cicchetti DV (1991) The reliability of peer review for manuscript and grant submissions: A cross-disciplinary investigation. *The Behavioral and Brain Sciences*; **14** (1): 119–135.
- Claxton LD (2005) Scientific authorship. Part 1. A window into scientific fraud? *Mutation Research*; **589** (1): 17–30.
- Cook I, Grange S and Eyre-Walker A (2015) Research groups: How big should they be? *PeerJ*; **3** (June): e989.
- Costello LC (2010) Perspective: Is NIH funding the 'best science by the best scientists'? A critique of the NIH R01 research grant review policies. *Academic Medicine: Journal of the Association of American Medical Colleges*; **85** (5): 775–779.
- Curry S (2012) PLoS ONE: From the Public Library of Sloppiness? *Reciprocal Space*, <http://occamstypewriter.org/scurry/2012/04/01/plos1-public-library-of-sloppiness/>, accessed 1 April.
- Dean DG (1989) Structural constraints and the publications dilemma: A review and some proposals. *The American Sociologist*; **20** (2): 181–187.
- Dickersin K (2005) Publication bias: Recognizing the problem, understanding its origins and scope, and preventing harm. In: Rothstein HR, Sutton AJ and Borenstein M (eds.) *Publication Bias in Meta-Analysis*; John Wiley & Sons: Chichester, UK, pp 9–33.
- Dickersin K, Chan S, Chalmers TC, Sacks HS and Smith H (1987) Publication bias and clinical trials. *Controlled Clinical Trials*; **8** (4): 343–353.
- Dobbs D (2006) Trial and Error. *The New York Times* 15 January, http://www.nytimes.com/2006/01/15/magazine/15wvln_idealab.html?_r=0.
- Ely A, Van Zwanenberg P and Stirling A (2014) Broadening out and opening up technology assessment: Approaches to enhance international development, coordination and democratisation. *Research Policy*; **43** (3): 505–518.
- Eve MP (2014) *Open Access and the Humanities: Contexts, Controversies and the Future*. Cambridge University Press: Cambridge, UK.
- Excellence Academy. (2016) *Indiana University*, http://iuhoosiers.com/sports/2015/6/25/GEN_0625153134.aspx, accessed 9 September.
- Eyre-Walker A and Stoletzki N (2013) The assessment of science: The relative merits of post-publication review, the impact factor, and the number of citations. Edited by Jonathan A. Eisen. *PLoS Biology*; **11** (10): e1001675.
- Fanelli D (2009) How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS One*; **4** (5): e5738.
- Fanelli D (2011) Negative results are disappearing from most disciplines and countries. *Scientometrics*; **90** (3): 891–904.
- Fang FC, Bowen A and Casadevall A (2016) NIH peer review percentile scores are poorly predictive of grant productivity. *eLife*; **5** (February). doi:10.7554/eLife.13323.
- Fang FC, Steen RG and Casadevall A (2012) Misconduct accounts for the majority of retracted scientific publications. *Proceedings of the National Academy of Sciences of the United States of America*; **109** (42): 17028–17033.
- Fitzpatrick K (2011) *Planned Obsolescence*. New York University Press: New York.
- Fleck L (1979) *Genesis and Development of a Scientific Fact*; Bradley F and Trenn TJ (trans). Trenn TJ and Merton RK (eds). University of Chicago Press: Chicago, IL.
- Fortin J-M and Currie DJ (2013) Big science vs. Little science: How scientific impact scales with funding. *PLoS One*; **8** (6): e65263.
- Funtowicz SO and Ravetz JR (1990) *Uncertainty and Quality in Science for Policy*. Theory and Decision Library A-Springer: The Netherlands.
- Funtowicz SO and Ravetz JR (2003) Post-Normal Science. In: *International Society for Ecological Economics Internet Encyclopaedia of Ecological Economics*. <http://isecoeco.org/pdf/pstnormsc.pdf>.
- Gans JS and Shepherd GB (1994) How are the mighty fallen: Rejected classic articles by leading economists. *The Journal of Economic Perspectives: A Journal of the American Economic Association*; **8** (1): 165.
- Goldacre B (2011) I Foresee That Nobody Will Do Anything about This Problem. *Bad Science*. 23 April, <http://www.badscience.net/2011/04/i-foresee-that-nobody-will-do-something-about-this-problem/>.
- Gonon F, Kongsman J-P, Cohen D and Boraud T (2012) Why most biomedical findings echoed by newspapers turn out to be false: The case of attention deficit hyperactivity disorder. *PLoS One*; **7** (9): e44275.
- Goodrich DW (1945) An analysis of manuscripts received by the editors of the *American Sociological Review* from May 1, 1944 to September 1, 1945. *American Sociological Review*; **10** (6): 716–725.
- Gordon R and Poulin BJ (2009) Cost of the NSERC science grant peer review system exceeds the cost of giving every qualified researcher a baseline grant. *Accountability in Research*; **16** (1): 13–40.
- Grieneisen ML and Zhang M (2012) A comprehensive survey of retracted articles from the scholarly literature. *PLoS One*; **7** (10): e44118.
- Groppe DM (2015) Combating the scientific decline effect with confidence (intervals) *BioRxiv*. doi:10.1101/034074.
- Guedj D (2009) Nicholas Bourbaki, collective mathematician: An interview with Claude Chevalley. *The Mathematical Intelligencer*; **7** (2): 18–22.
- Hassell MP and May RM (1974) Aggregation of predators and insect parasites and its effect on stability. *The Journal of Animal Ecology*; **43** (2): 567–594.
- Health Research Council of New Zealand (2016) Explorer Grants. *Health Research Council*, <http://www.hrc.govt.nz/funding-opportunities/researcher-initiated-proposals/explorer-grants>, accessed 19 February.
- Herbert DL, Barnett AG, Clarke P and Graves N (2013) On the time spent preparing grant proposals: An observational study of Australian researchers. *BMJ Open*; **3** (5): e002800.
- Herrnstein Smith B (1988) *Contingencies of Value: Alternative Perspectives for Critical Theory*. Harvard University Press: Cambridge, MA.
- Hill H and Pitt J (2014) Failure to replicate: A sign of scientific misconduct? *Publications*; **2** (3): 71–82.
- Hoover WG, Moran B, Holian BL, Posch HA and Bestiale S (1988) Computer simulation of nonequilibrium processes. In: Schmidt SC and Homes NC (eds) *Shock Waves in Condensed Matter 1987*; North-Holland. pp 191–194.
- Hoover WG, Posch HA and Bestiale S (1987) Dense-fluid Lyapunov spectra via constrained molecular dynamics. *The Journal of Chemical Physics*; **87** (11): 6665–6670.
- Ioannidis JPA, Boyack KW, Small H, Sorensen AA and Klavans R (2014) Bibliometrics: Is your most cited work your best? *Nature*; **514** (7524): 561–562.
- Kennedy D (2004) The old file-drawer problem. *Science*; **305** (5683): 451.
- Kilkenny C, Browne WJ, Cuthill IC, Emerson M and Altman DG (2010) Improving bioscience research reporting: The ARRIVE guidelines for reporting animal research. *PLoS Biology*; **8** (6): e1000412.
- King M, West JD, Jacquet J, Correll S and Bergstrom CT (2014) Gender Composition of Scholarly Publications. *Eigenfactor*, <http://www.eigenfactor.org/gender/self-citation/>, accessed 7 January.
- King MM, Bergstrom CT, Correll SJ, Jacquet J and West JD (2016) Men set their own cites high: Gender and self-citation across fields and over time. arXiv [physics.soc-ph], <http://arxiv.org/abs/1607.00376>.
- Kohn A (1999) *Punished by Rewards: The Trouble with Gold Stars, Incentive Plans, A's, Praise, and Other Bribes*. Houghton Mifflin: Boston, MA.
- Kuhn TS ((1962) 2012) *The Structure of Scientific Revolutions*, Fourth edition, University of Chicago Press: Chicago, IL.
- Labbé C (2010) I like Antkare: One of the great stars in the scientific firmament. *ISSI Newsletter*; **6** (2): 48–52.
- Lamont M (2009) *How Professors Think: Inside the Curious World of Academic Judgment*. Harvard University Press: Cambridge, MA.
- Latour B (1987) *Science in Action: How to Follow Scientists and Engineers Through Society*. Harvard University Press: Cambridge, MA.

- Latour B and Woolgar S (1986) *Laboratory Life: The Construction of Scientific Facts*. Princeton University Press: Princeton, NJ.
- Lawrence PA (2007) The mismeasurement of science. *Current Biology*; **17** (15): R583–R585.
- Lehrer J (2010) The Truth Wears Off. *New Yorker* 13 December, <http://www.newyorker.com/magazine/2010/12/13/the-truth-wears-off>.
- Lerner J and Wulf J (2006) Innovation and Incentives: Evidence from Corporate R&D. W11944. National Bureau of Economic Research, <http://www.nber.org/papers/w11944.pdf>.
- Lindner MD and Nakamura RK (2015) Examining the predictive validity of NIH peer review scores. *PLoS One*; **10** (6): e0126938.
- Lord RG, de Vader CL and Alliger GM (1986) A meta-analysis of the relation between personality traits and leadership perceptions: An application of validity generalization procedures. *The Journal of Applied Psychology*; **71** (3): 402.
- Lundh A, Barbateskovic M, Hróbjartsson A and Gøtzsche PC (2010) Conflicts of interest at medical journals: The influence of industry-supported randomised trials on journal impact factors and revenue—Cohort study. *PLoS Medicine*; **7** (10): e1000354.
- MacCallum CJ (2011) Why ONE is more than 5. *PLoS Biology*; **9** (12): e1001235.
- Marcovitch H (2010) Editors, publishers, impact factors, and reprint income. *PLoS Medicine*; **7** (10): e1000355.
- Matzinger P and Mirkwood G (1978) In a fully H-2 incompatible chimera, T cells of donor origin can respond to minor histocompatibility antigens in association with either donor or host H-2 type. *The Journal of Experimental Medicine*; **148** (1): 84–92.
- Meng W (2016) Peer Review: Is NIH Rewarding Talent? *Science Transparency*, <https://sciencetransactions.wordpress.com/2016/01/10/peer-review-is-nih-rewarding-talent/>, accessed 10 January.
- Merton RK (1968) The Matthew Effect in science. *Science*; **159** (3810): 56–63.
- Merton RK (1988) The Matthew Effect in science, II: Cumulative advantage and the symbolism of intellectual property. *Isis: An International Review Devoted to the History of Science and Its Cultural Influences*; **79** (4): 606–623.
- Miller S (2015) *Excellence and Equity: The National Endowment for the Humanities*. University Press of Kentucky: Lexington, KY.
- Moran B, Hoover WG and Bestiale S (2016) Diffusion in a periodic Lorentz gas. *Journal of Statistical Physics*; **48** (3–4): 709–726.
- Morgan D (2016) Lessons Learned, and How the Landscape Has Already Changed. Lecture presented at the Open Access @ UNT/Library Publishing Forum, University of North Texas, 19 May, <https://openaccess.unt.edu/symposium/2016/live-streaming-oa-uc-press-lessons-learned-and-how-landscape-has-already-changed>.
- Mrs Kinpaisby (2008) Taking stock of participatory geographies: envisioning the communitarity. *Transactions of the Institute of British Geographers*; **33** (3): 292–299.
- Munafò MR, Stothart G and Flint J (2009) Bias in genetic association studies and impact factor. *Molecular Psychiatry*; **14** (2): 119–120.
- Networks of Centres of Excellence of Canada, Communications. (2015) Networks of Centres of Excellence, http://www.nce-rce.gc.ca/index_eng.asp, accessed 13 April.
- Nicholson JM and Ioannidis JPA (2012) Research grants: Conform and be funded. *Nature*; **492** (7427): 34–36.
- Nielsen MA (2012) *Reinventing Discovery: The New Era of Networked Science*. Princeton University Press: Princeton, NJ.
- Nosek BA, Spies JR and Motyl M (2012) Scientific Utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*; **7** (6): 615–631.
- Nowotny H (2014) Excellence Attracts Excellence—and What about the Rest? Reflections on Excellence and Inclusion. Lecture presented at the EMBO–EMBL Anniversary Science and Policy Meeting, Heidelberg, July, http://www.helganowotny.eu/downloads/helga_nowotny_b160.pdf.
- O'Connor P and O'Hagan C (2015) Excellence in university academic staff evaluation: A problematic reality? *Studies in Higher Education*; **41** (11): 1943–1957.
- O'Donnell DP (2015) Could We Design Comparative Metrics That Would Favour the Humanities? *Daniel Paul O'Donnell*, <http://people.uleth.ca/~daniel.odonnell/Teaching/could-we-design-comparative-metrics-that-would-favour-the-humanities>, accessed 29 March.
- OECD (2014) Chapter 6. The German Excellence Initiative. In: *Promoting Research Excellence*. OECD Publishing, pp 145–163.
- Office of Excellence and Multicultural Student Success (2016) *University of Toledo*. <http://www.utoledo.edu/success/excel/>, accessed 9 September.
- Open Science Collaboration. (2015) Estimating the reproducibility of psychological science. *Science*; **349** (6251): aac4716.
- Pagano M (2006) American Idol and NIH grant review. *Cell*; **126** (4): 637–638.
- Peters DP and Ceci SJ (1982) Peer-review practices of psychological journals: The fate of published articles, submitted again. *The Behavioral and Brain Sciences*; **5** (2): 187–195.
- Potts J, Hartley J, Montgomery L, Neylon C and Rennie E (2016) A Journal is a club: A new economic model for scholarly publishing. *Social Science Research Network* (April). doi:10.2139/ssrn.2763975.
- Psych Filedrawer (2016) The Filedrawer Problem. *PsychFileDrawer.org*, <http://www.psychfiledrawer.org/TheFiledrawerProblem.php>, accessed 19 February.
- Public Library of Science. (2016) Who we are. *PLoS*, <https://www.plos.org/who-we-are>, accessed 12 May.
- Ravetz JR (1973) *Scientific Knowledge and Its Social Problems*. Penguin Books: London, UK.
- Readings B (1996) *The University in Ruins*. Harvard University Press: Cambridge, MA.
- Rehman J (2013) Cancer Research in Crisis: Are the Drugs We Count on Based on Bad Science? *Salon*, 1 September, http://www.salon.com/2013/09/01/is_cancer_research_facing_a_crisis/.
- Resnik DB and Dinse GE (2013) Scientific retractions and corrections related to misconduct findings. *Journal of Medical Ethics*; **39** (1): 46–50.
- Resnik DB, Wager E and Kissling GE (2015) Retraction policies of top scientific journals ranked by impact factor. *Journal of the Medical Library Association*; **103** (3): 136–139.
- Roderick GK and Gillespie RG (1998) Speciation and phylogeography of Hawaiian terrestrial arthropods. *Molecular Ecology*; **7** (4): 519–531.
- Roorda S (2009) The real cost of the NSERC peer review is less than 5% of a proposed baseline grant. *Accountability in Research*; **16** (4): 229–231.
- Rothstein HR (2014) Publication bias. In: *Wiley StatsRef: Statistics Reference Online*. John Wiley & Sons.
- Schooler J (2011) Unpublished results hide the decline effect. *Nature*; **470** (7335): 437.
- Schulz KF, Altman DG, Moher D and CONSORT Group (2010) CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *BMJ*; **340** (March): c332.
- Secord JA (2015) *Visions of Science: Books and Readers at the Dawn of the Victorian Age*. University of Chicago Press: Chicago, IL.
- Şekericioğlu ÇH (2013) Citation opportunity cost of the high impact factor obsession. *Current Biology*; **23** (17): R701–R702.
- Sewitz S (2014) The Excellence Agenda Is a Trojan Horse for Austerity. *Research*, http://www.researchresearch.com/index.php?option=com_news&template=rr_2col&view=article&articleid=1346207, accessed 3 September.
- Shirky C (2010) *Cognitive Surplus: Creativity and Generosity in a Connected Age*. Penguin Press: New York.
- Siler K, Lee K and Bero L (2015) Measuring the effectiveness of scientific gatekeeping. *Proceedings of the National Academy of Sciences*; **112** (2): 360–365.
- Simons DJ, Holcombe AO and Spellman BA (2014) An introduction to registered replication reports at *Perspectives on Psychological Science*. *Perspectives on Psychological Science*; **9** (5): 552–555.
- Smith R (2006) Research misconduct: The poisoning of the well. *Journal of the Royal Society of Medicine*; **99** (5): 232–237.
- Solomon DJ (2014) A survey of authors publishing in four megajournals. *PeerJ*; **2** (April): e365.
- Steen RG (2011) Retractions in the scientific literature: Is the incidence of research fraud increasing? *Journal of Medical Ethics*; **37** (4): 249–253.
- Sterling TD (1959) Publication decisions and their possible effects on inferences drawn from tests of significance—Or vice versa. *Journal of the American Statistical Association*; **54** (285): 30–34.
- Stilgoe J (2014) Against Excellence. *The Guardian* 19 December, <https://www.theguardian.com/science/political-science/2014/dec/19/against-excellence>.
- Stirling A (2007a) 'Opening up' and 'Closing down': Power, participation, and pluralism in the social appraisal of technology. *Science, Technology & Human Values*; **33** (2): 262–294.
- Stirling A (2007b) A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society, Interface / the Royal Society*; **4** (15): 707–719.
- Sustaining Excellence Awards. (2016) *Wellcome Trust*, <http://www.wellcome.ac.uk/Funding/Public-engagement/Funding-schemes/Sustaining-Excellence-Awards/index.htm>, accessed 19 May.
- Tartamela V (2014) The True Story of Stronzo Bestiale (and Other Scientific Jokes). *Parolacce*, <http://www.parolacce.org/2014/10/05/the-true-story-of-stronzo-bestiale/>, 5 October.
- The Economist*. (2015) The World Is Going to University, <http://www.economist.com/news/leaders/21647285-more-and-more-money-being-spent-higher-education-too-little-known-about-whether-it>.
- The University of Tokyo Global Center of Excellence. (2016) Global Center of Excellence for Mechanical Systems Innovation. *The University of Tokyo Global COE*, <http://www.u-tokyo.ac.jp/coe/english/list/category2/base7/summary.html>, accessed 12 May.
- Tian X and Wang TY (2011) Tolerance for failure and corporate innovation. *The Review of Financial Studies* (December). doi:10.1093/rfs/hhr130.
- Tijdink JK, Verbeke R and Smulders YM (2014) Publication pressure and scientific misconduct in medical scientists. *Journal of Empirical Research on Human Research Ethics*; **9** (5): 64–71.

- U15 Group of Canadian Research Universities/Regroupement des universités de recherche du Canada. (2016) Our Impact. *U15*, <http://u15.ca/our-impact>, accessed 18 May.
- University Grants Commission. (2016) Universities (UPE). *University Grants Commission*, [http://www.ugc.ac.in/page/Universities-\(UPE\).aspx](http://www.ugc.ac.in/page/Universities-(UPE).aspx), accessed 19 February.
- University of Arizona Commission on the Status of Women. (2015) *Avoiding Gender Bias in Reference Writing*. University of Arizona: Tucson, AZ, http://www.csw.arizona.edu/sites/default/files/csw_2015-10-20_lorbias_pdf_0.pdf.
- USC Center for Excellence in Teaching. (2016) <http://cet.usc.edu/>, accessed 12 May.
- Usher A (2016) Massification Causes Stratification. *Higher Education Strategy Associates*, <http://higherstrategy.com/massification-causes-stratification/>, accessed 5 May.
- Vessuri H, Guedon J-C, Cetto and Mara A (2014) Excellence or quality? Impact of the current competition regime on science and scientific publishing in Latin America and its implications for development. *Current Sociology*; **62** (5): 647–665.
- Wang J, Veugelers R and Stephan PE (2016) Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Social Science Research Network*; (January). doi:10.2139/ssrn.2710572.
- Weigley S and Hess AEM (2013) Universities Getting the Most Government Money. *247wallst.com*, <http://247wallst.com/special-report/2013/04/25/universities-getting-the-most-government-money/>, accessed 25 April.
- Weller AC (2001) *Editorial Peer Review: Its Strengths and Weaknesses*. Information Today: Medford NJ.
- Wilson A (2011) Failing to Replicate Bem's Ability to Get Published in a Major Journal. *Notes from Two Scientific Psychologists*, <http://psychsciencenotes.blogspot.ca/2011/05/failing-to-replicate-bems-ability-to.html>, accessed 7 May.
- Wittgenstein L (2001) *Philosophical Investigations: The German Text, with a Revised English Translation*. Blackwell: Oxford, UK.
- Yong Ed. (2012a) A Failed Replication Draws a Scathing Personal Attack from a Psychology Professor. *Not Exactly Rocket Science*, <http://blogs.discovermagazine.com/notrocketscience/2012/03/10/failed-replication-bargh-psychology-study-doyen/#.VsZpH0Leezc>, accessed 10 March.
- Yong E (2012b) Replication studies: Bad copy. *Nature*; **485** (7398): 298–300.
- Young NS, Ioannidis JPA and Al-Ubaydli O (2008) Why current publication practices may distort science. *PLoS Medicine*; **5** (10). doi:10.1371/journal.pmed.0050201.
- Young SS and Bang H (2004) The file-drawer problem, revisited. *Science*; **306** (5699): 1133–1134.
- Zankl H (2003) *Fälscher, Schwindler, Scharlatane*. Erlebnis Wissenschaft. Wiley-VCH Verlag: Weinheim.

Data availability

Data sharing is not applicable as no datasets were analysed or generated during this study.

Author Contributions

The corresponding author is cn. Author contributions, described using the CASRAI CRediT typology (<http://casrai.org/credit>), are as follows: conceptualization: me, sm, cn, dod, dp; methodology: me, sm, cn, dod, dp; investigation: me, sm, cn, dod, dp; resources: me, sm, cn, dod, dp; writing – original draft preparation: me, cn, dod; writing – review and editing: me, sm, cn, dod, dp; funding acquisition: sm.

Acknowledgements

In keeping with our argument, and following an extensive tradition of subverting traditional scarce markers of prestige, the authors have adopted a redistributive approach to the order of their names in the byline. As an international collaboration of uniformly nice people (cf. Moran *et al.*, 2016; Hoover *et al.*, 1987; see Tartamella, 2014 for an explanation), lacking access to a croquet field (cf. Hassell and May, 1974), writing as individuals rather than an academic version of the Borg (see Guedj, 2009), and not identifying any excellent pun (cf. Alpher *et al.*, 1948; Lord *et al.*, 1986) or “disarmingly quaint *nom de guerre*” (cf. Mrs Kinpaisby, 2008, 298 [thanks to Oli Duke-Williams for this reference]) to be made from the ordering of our names, we elected to assign index numbers to our surnames and randomize these using an online tool. For the avoidance of doubt, while several of the authors have pets, none of them are included as authors (cf. Matzinger and Mirkwood, 1978); none of us are approaching a tenure decision (cf. Roderick and Gillespie, 1998); and none of us are fictional entities who generate their papers algorithmically using SciGen (see Labbé, 2010 for the contrasting case of “Ike Antkare,” who nevertheless greatly outranked all the authors of this paper on several formal measures of excellence before being outed). This article arose from a meeting at the Triangle Scholarly Communications Institute funded by the Andrew W. Mellon Foundation. The authors wish to thank Ben Johnson, and commenters on Hacker News for their criticisms and comments on an earlier version of this manuscript.

Additional information

Competing interests: Moore, Neylon, and Pattinson are all previous employees of PLOS. Eve and O'Donnell declare no competing interests.

Reprints and permission information is available at http://www.palgrave-journals.com/pal/authors/rights_and_permissions.html

How to cite this article: Moore S *et al.* (2017) “Excellence R Us”: university research and the fetishisation of excellence. *Palgrave Communications*. 3:16105 doi: 10.1057/palcomms.2016.105.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>