Peer review for science funding: a review

Valentina Tartari, Copenhagen Business School, DK, vt.si@cbs.dk
Christos Kolympiris, Warwick Business School, UK, christos.kolympiris@wbs.ac.uk

Abstract

Peer review is widely considered the gold standard for evaluating research and indeed it is the most widely means to allocate funds among competing proposals in the sciences. In this chapter we review what we know about the peer review system. We discuss evidence suggesting where the system delivers desired outcomes such as funding the most meriting proposals and where it appears to be prone to shortcomings including biases originating from the gender, status and ethnicity of proposers. We then review alternatives to the traditional peer review such as funding people rather than projects and discuss how these alternatives have performed. We conclude the chapter with suggestions on research avenues that could improve our understanding of what could work in order to improve the allocation of research funds.

Introduction

The role of institutions in translating scientific knowledge into welfare-enhancing innovation has long been acknowledged by scholars (Dasgupta and David, 1994; Mokyr, 2002). One of these key institutions is the grant system. From the patronage allowing Galileo's astronomical observations to the early governmental grants sponsoring Babbage's difference engine and the investigator-initiated, peer-reviewed scientific grants that emerged around WWII, the system of science funding has evolved and is recognized as one of the main institutions supporting the development of knowledge and, consequently, societal progress.

One of the cornerstones of the modern system of fund allocation is evaluation through peer-review. Peer review is widely considered the gold standard for evaluating research and, in broad strokes, appears effective at selecting for quality, novelty and likely impact (Packalen and Bhattacharya, 2020; Park et al., 2015; Siler et al., 2015). Assessment by external, disinterested parties is critical for preserving scientific integrity, assessing value and interest to the field, and performing cost-benefit analysis contributions. Given its efficiency and the widespread trust it enjoys, the Royal Society (2007) defined peer-review as "the only effective way of properly assessing the quality of research proposals". Indeed, as Wessely (1998, p. 301) put it "…the peer review of grant proposals may be more relevant to the health of science than publication practices."

However, since becoming the method of choice to assess the quality of science and the subsequent allocation of resources, questions have been raised about grant peer-review. Is peer-review the best system to select the best people and the best ideas? Can evaluators be *really* objective when they are evaluating their peers? In particular, as the scientific system expands and need for reviews is constantly increasing, the task of reviewers becomes increasingly difficult and time-consuming. As a result, the system may end up overburdened, making discerning quality a challenging task. This in turns undermines the promise of peer review to be able to weed out lower quality projects and award resources to the most meriting.

For example, some scholars have revealed biases in peer review; using data from the Netherlands Organization of Scientific Research (NWO) grants, Bol et al. (2018) document the presence of the Matthew effect, whereby success breeds success, and Van Der Lee et al. (2015) "reveal gender bias favoring male applicants over female applicants". Witteman et al. (2019) and Kaatz et al. (2016) also report gender bias in funding from the Canadian Institutes of Health Research and the National Institutes of Health (NIH) in the US respectively. Also within the context of the NIH, Ginther et al. (2011) show evidence consistent with ethnicity bias, even though the approach of the proposed project and the significance[1] score are the main predictors of funding success at the NIH (Eblen et al., 2016) despite low agreement among NIH reviewers when scoring the same proposal (Pier et al., 2018). Finally, on a different potential source of bias, Materia et al. (2015) using data on agricultural research projects funded by the Emilia Romagna regional government in Italy find that, net of a proposal's scientific merit, the composition of the reviewing team may also affect the chances of funding. The effect size of these biases on the probability of funding success can stretch to the point where "the outcome of the grant review depended more on the reviewer to whom the grant was assigned than the research proposed in the grant" (Pier et al., 2018).

Additional costs of the grant peer review may include the time and effort allocated to it, which can even exceed the potential benefits. For example, using contest models, Gross and Bergstrom (2019) estimate "that the effort researchers waste in writing proposals may be comparable to the total scientific value of the research that the funding supports, especially when only a few proposals can be funded." And Herbert et al. (2013) found that for the case of Australian scientists submitting proposals for the largest funding scheme of 2012 "the equivalent of some four centuries of effort returned no immediate benefit to researchers and wasted valuable research time." Similarly, submitting funding proposals induces stress, entails high opportunity costs and comes at the expense of attending to family matters (Herbert et al., 2014). Finally, the grant peer review system may also favor more conservative approach at the expense of more daring, risk taking proposals (Franzoni et al., 2021). To this end, Ayoubi et al. (2021) leverage applications to a leading Swiss research funding program to conclude that scientists with a history of novel research are less likely to receive funding despite the fact that they are more likely to apply. Along the same lines, Lane et al. (2021) conduct field experiments on funding allocations within a leading US research institution to find that when expert reviewers share information with each other they tend to make more conservative funding allocation decisions.

[1] The NIH criteria define significance as the answer to the following questions: "Does the project address an important problem or critical barrier to progress in the field? Is there a strong scientific premise for the project? If the aims of the project are achieved, how will scientific knowledge, technical capability, and/or clinical practice be improved? How will successful completion of the aims change the concepts, methods, technologies, treatments, services, or preventative interventions that drive this field?" (from NOT-OD-09.025, retrieved 04/25/2022)

The findings above on biases and other potential shortcomings of the grant review system are not conclusive on the efficiency and fairness of peer review primarily because they are conducted among relatively small samples compared to the scope of the whole grant system. But, they do invite additional scrutiny of the existing system because if they hold on a larger scale they can affect outcomes and decrease participation and representation opportunities in scientific discovery (Demarest et al., 2014). Indeed, a handful of alternatives have been suggested with Linton (2016), for example, arguing that funding priority should go to projects with the greatest disagreement among evaluators (rather than consensus on merit and impact). [2] Overall, we lack systematic evidence on the effectiveness of *alternative forms of proposal evaluation*. Are there feasible process innovations in evaluation that will yield larger long-term scientific impact of proposals, that will increase participation of underrepresented groups, that will reduce potential biases and increase confidence in the meritocracy of the results, and that will reduce costs?

In the rest of the chapter we first review what we know about the grant peer review and its limitations. Then, we present the alternatives to and modifications of the standard grant peer review model. We conclude with a reflection of the importance of designing and evaluating new forms of peer review in funding, and recommendations for the future.

Standard Peer Review and its Limitations

Peer review, in place since at least the late 1800s[3], has become the dominant design in the evaluation and dissemination of science. Its extensive use for distributing research grants coincided with the expansion of government funding of research in the USA and Europe after the Second World War. It has since become the main route for distributing money to scientific research: in 2021 the National Institutes of Health (NIH) distributed more than 80% of their US$39 billion budget through peer review[4], while in 2017 the National Science Foundation (NSF) assigned more than 90% of its funding for research based on peer review[5].

While each funding body may present some idiosyncrasies in their review process, traditional peer review of funding proposal usually involves a first stage where experts evaluate applications individually. Then, experts meet as a group to reach consensus on which applications should be funded. The experts who participate in the group (i.e., panelists) may have served as reviewers

[2] Alternatives have also been suggested in the context of paper peer review such as imposing deadlines, apply double blindness, provide payments to reviewers and make referee reports public (Blank, 1991; Bravo et al., 2019; Chetty et al., 2014; Polka et al., 2018). While the underlying principles of paper and grant peer reviews are the same, important differences between them such as that only the latter has an element of forecasting future outcomes suggest that the findings of those studies are difficult to extrapolate to grant peer reviews.

[3] While expert assessment was common practice as early as in the 17th century when scientific discoveries were discussed by scholars at meetings of the Royal Society in the UK, peer review in the modern sense began with scientific publishing. In particular, the Royal Society's Philosophical Transactions was the first scientific journal using peer review to select manuscripts for publication.

[4] https://www.nih.gov/about-nih/what-we-do/budget

[5] https://www.nsf.gov/nsb/publications/2018/nsb201915.pdf

themselves or may use reviews by other experts as input to decision making. Reviews may be collected only in writing, or only using in-person meetings (or a mixture of the two). Some funders, such as the European Research Council (Veugelers, 2017) require an initial summary of the proposed project, and in a second phase evaluate detailed applications, while others evaluate directly an extended application. Some funders may rely on a more stable group of evaluators (which may be renewed after a certain number of years), while others may use more extensively ad hoc evaluators.

The peer review system, as applied to grants and beyond, is based on a specific belief: because the underlying scientific discovery process is unquestionably improved through expert evaluation and feedback, everyone involved in the scientific process should gain from it and thus has incentives to participate, invest efficient effort levels, and accurately report their unvarnished expert opinions. Ideally, such system should engender a virtuous cycle. Authors are motivated to improve their work in response to scrutinized feedback; reviewers are given the opportunity to read the latest developments that can improve their own work in the field; and decision-makers are able to make sharper decisions based on the consultative advice of reviewers, resulting in subpar works being weeded out while promising avenues are explored. Indeed, there is evidence suggesting that peer review can potentially weed out work of lower quality and identify *ex-ante* higher quality research as measured by publications and citations (Li, 2017; Park et al., 2015; Siler et al., 2015). It is a noble vision of collaborative science creating public knowledge benefits for the common good even as competition in the market for ideas determines the distribution of the private component of such gains (who gets funded, published, cited, promoted, etc.).

There are several problems with this rosy picture. Stephan (2012) lists some of them: a) excessive time spent for reviewing, b) discouragement of risk taking and pursuing novel lines of research, c) low tolerance for failure, d) reduced chances of success for younger scholars and e) likely mismatch between increases in grant funding and productivity. Similarly, variation in success rates across years driven by differences in the availability of funds, can also influence decision makers' preferences shifting attention away from scientific merit as the prime funding criterion (Stephan, 2013). Indeed, we can categorize the main challenges of the traditional system of proposal evaluation based on peer-review across three main axes: reliability, biases (including risk aversion), and cost.

Reliability

The first challenge about peer review relates to its effectiveness; and how to measure it. Ideally, one would compare the scores of reviewers to the actual impact of funded projects, thereby estimating their validity. It is rather evident that such an exercise would encounter several problems in its implementation. First, information typically exists only on funded projects: this means that it would be exceedingly difficult to evaluate whether the system is effective at funding the best proposals, only the extent to which funding the chosen projects produced a benefit. Second, it is inherently challenging to find reliable indicators for measuring the impact of science, especially basic science that is far from commercial applications. Indeed, several studies have tried to tackle this issue of "evaluating the evaluation" using mainly publication metrics: the number of papers produced, the numbers of papers published in high-impact journals and the number of citations of those papers. Some studies found a positive, though somewhat weak, correlation between reviewers' scores and bibliometric measures of funded projects (Li and Agha, 2015; Lindner et al., 2016), but others found a negative association (Fang et al., 2016). Similar comparisons based on career progressions of successful and (marginally) unsuccessful applicants have also failed to find significant differences as long as those who failed are able to source funds from elsewhere (Jacob and Lefgren, 2011; Klaus and del Alamo, 2018; Wang et al., 2019). From the frequent complaints of Nobel Prize winners about the difficulties they have experienced in getting their award-winning work past grant review committees, to the account of the decades long failure to secure funding for the mRNA research that has brought us a Covid19 vaccine in record time in 2021, there is plenty of anecdotal evidence suggesting that review panels do a suboptimal job at predicting scientific success. Third, it is difficult to find a reliable way to account for the uncertainty at the time the opinion is provided. This is particularly salient for grant peer review, as it essentially aims at forecasting future outcomes, compared to editorial peer review, which focuses on judging a completed work.

An alternative to evaluate the validity of the scores of a peer review process is to check their consistency, namely the extent to which different panel members agree among themselves about the merit of the projects they evaluate (Mutz et al., 2012; Sattler et al., 2015). A study conducted in Australia in 2009 by Graves et al. (2011), revealed a "high degree of randomness" in the scoring process, raising important concerns given the importance of funding decisions to individuals' careers. In particular, the problem of inconsistency seems particularly acute in the "middle-group" of proposals (often referred to as the grey zone), which seems to be subjected to some semi-random process (Crossley, 2015). Indeed, it has been highlighted that grant peer review is insufficiently precise to provide reliable rank ordering of applications (Kaplan et al., 2008). This problem is particularly severe in multidisciplinary schemes, where the grey zone might consist of applications that are so different that cannot be compared and selected against each other (Feller, 2006).

Biases

Several studies demonstrate systematic biases in the standard peer-review design. Highly novel projects are routinely penalized by reviewers, largely because of bounded-rational evaluations and difficulties in the reviewers' appreciation of the work (Ayoubi et al., 2021; Boudreau et al., 2016; Lane et al., 2021). Although many funding agencies emphasize that their aim is to fund innovative research, peer review systems of evaluation might instead be rather conservative, and favor well-established views rather than radically new ones (Guthrie et al., 2018; Guthrie et al., 2019; Nicholson and Ioannidis, 2012). Even expert reviewers, who can better appreciate novelty in the context of the existing literature, suffer from such self-serving biases. Boudreau et al. (2016) conducted an experiment in which they randomized the assignment of evaluators to proposals in a university-wide grant proposal competition and found that "evaluators gave systematically lower scores to research proposals that were closer to their own areas of expertise" (p. 2766), suggesting a tradeoff between expertise and a resistance to competitive encroachment.

In a study of National Institutes of Health (NIH) reviews, Li (2017) found that evaluators are better informed about the quality of projects but at the same time more biased in their assessments when the subject is close to their own area of expertise. Besides the desire to maintain professional stature and funding, such reviewers may be (deliberately or subconsciously) slanting their reviews (whether upward or downward) to express their personal preferences for the direction that scientific inquiry in their field should take. While these expert volunteer reviewers' efforts are highly valued and their advocacy tolerated (or, sometimes, encouraged), such biases mean that reviews cannot be taken at face value without an appreciation from the context of their source, even in the traditional external/anonymous format designed to produce disinterested reviews.

Biases may also materialize at the level of the evaluating committee that aggregates individual reviews (Materia et al., 2015). Such biases often stem from characteristics of the proposer and lead to review outcomes, which do not depend entirely on stated program goals such as intellectual merit or broader impact. For instance, in the context of proposals submitted to the NIH, Asian-heritage PIs are 4% less likely, and black or African-American PIs 13% less likely, to receive NIH investigator-initiated research funding (Ginther et al., 2011). At the same time, proposers from states represented by House Appropriations Committee members receive 5.9–10.3% more research funds than those at unrepresented institutions (Hegde, 2009)[6]. As well, Van Der Lee et al. (2015) and Witteman et al. (2019) use large scale datasets from the Netherlands and Canada respectively to reveal systematic gender bias favoring male over females candidates in terms of their odds of securing grants. Proposers do not control their ethnic heritage, gender or, at least in the short run, their institution's location or political representation, yet their success rates are shifted by each, even for proposals of similar merit. Literature has also shown the presence of bias in the evaluation of grant proposals that are submitted by teams: in their analysis of EPSRC grants, Banal-Estañol et al. (2019) show that diverse teams (in terms of in knowledge and skills, education, and/or ability) are not only penalized but are also biased against as they are significantly less likely to obtain funding, but are generally more likely to be successful. As highlighted above, funding agencies may be biased against diversity because applications of diverse teams are more difficult to evaluate or because they are perceived as riskier or less achievable. This may be exacerbated by the dominant structure of review panels, in which only a few reviewers examine each proposal in detail (especially at the initial stages), reducing the possibilities of having a large breadth in terms of expertise from the reviewers side (Gluckman, 2012).

Social capital and centrality in a professional network can introduce biases as well: Feinberg and Price (2004) analyzed funded and unfunded proposals in the economics program at the NSF to conclude that, net of ability, being a National Bureau of Economic Research associate (their measure of social capital) increases the probability of a research proposal being funded. While the general wish to advance those similar (but not too similar) to us may be a well-established and unavoidable part of human nature, biased reviews hamper the efficient and equitable process of allocating resources for scientific research. This form of cronyism imposes a disproportionate disadvantage on

[6] Although congressional appropriators do not directly earmark federal funds, the paper shows they support allocations for those research fields that are most likely to benefit performers in their constituencies, thereby interfering with merit-driven allocations.

parts of the scientific system that already suffer from under-representation, such as women and ethnic minorities, or that are at the periphery of the system, such as researchers in smaller and less well-endowed institutions.

Here it is important to note that evidence on gender, ethnicity and other biases is rather exploratory as the body of work has not reached a critical mass. However, given the importance of securing funding in the scientific profession, it is crucial to be aware of the possibility of the existence of those biases, and how they may affect specific groups of researchers.

Costs

The problems related to traditional peer review and proposal evaluations could potentially be justifiable if the costs of the system were relatively low. However, that does not seem to be the case always, neither in terms of the costs of administering the evaluations themselves, nor in terms of the costs associated to preparing the proposals.

On the evaluation side, a common problem is the difficulty in finding qualified volunteer reviewers. Even when reviewers can be identified and recruited, the burden often falls disproportionately on a small minority whose (opportunity) costs of acting as reviewers are unlikely to balance any benefits given the magnitude of such costs (van den Besselaar, 2012). For example, in biomedicine, the top 5% of contributing reviewers provided 30% of the 63.4 million hours devoted to peer review in 2015 (Kovanis et al., 2016). Also, in 2015, the NSF called upon 16,255 scientists to evaluate 51,588 proposals, with the total time spent by reviewers estimated to be on average 360 person-years, and each reviewer spending about 3.9 hours per review (not counting the time spent participating in panels)[7]. NSF's own Merit Review Working Group has identified the shortage of qualified reviewers and the challenge of integrating ever-increasing review loads into their professional activities as a perennial and growing problem (NSF, 2017). A simple process-balance analysis indicates that, because each proposal requires a minimum of three reviews, submitting Principal Investigators must (in the long run) volunteer three reviews for every proposal they submit. The incentives for the PIs' institutions to create demand for reviews are at odds with the NSF's mission to maintain high reviewing standards: some institutions encourage (or require) multiple proposals per year, but NSF discourages programs from calling upon reviewers multiple times per year, making a balanced process inherently impossible (and permanently short of qualified reviewers) in the traditional setup. More to it, while the NIH seeks at least 4 reviewers per proposal, Kaplan et al. (2008) develop an analytical model suggesting that a significantly larger number of evaluators is needed to improve the precision of decision making. The NIH indicates that supported Principal investigators are particularly suitable as referees and invites them to serve on panels; however, it also states that "this expectation for service is entirely voluntary and an inability to serve has no impact on an investigator's ability to compete for grant support"[8].

[7] https://www.nsf.gov/nsb/publications/2016/nsb201641.pdf
[8] https://grants.nih.gov/grants/peer/becoming_peer_reviewer.htm

On the applicants' side, there is ample evidence that applying for grants is a costly endeavor even though data from a Swiss funding program indicates that submitting applications *per se* boost scholarly productivity (Ayoubi et al., 2019). Link et al. (2008) show that researchers at R1-universities in the United States spend on average more than four hours a week writing project applications. A similar study conducted in Australia for the National Health and Medical Research Council indicate that the time investment in 2009 was as much as 180 years of research time to fund 620 projects, and by 2013, the costs had gone up to more than 500 years of research time – equivalent to € 41 million in salary (Herbert et al., 2013). In the same Australian study, the authors also tried to detect a correlation between extra time spent on a proposal and its likelihood of receiving funding. Given the power of their study, the conclusion was that, on average, 10 extra days spent on a proposal were likely to increase the likelihood of success by 2.8% at most. Along the same lines, the analytical model of Gross and Bergstrom (2019) demonstrates that when only a small subset of applications is funded, the effort exerted for writing proposals can equal the total value of the science produced as the result of grants; the effort can be even higher when scientists submit proposals for reasons not directly linked to supporting the research at hand (e.g., prestige or meeting promotion criteria).

Modification and alternatives to peer review

Given the above challenges to the current system of peer review for the allocation of funding, several alternative proposals have been put forward – ranging from a complete elimination of peer review, to more incremental changes to the system, such as anonymization of proposals. In the following, we will examine the rationale of the main alternatives or modifications to peer review, providing also examples of funding bodies that have implemented such changes. As the literature in this area is still in its early stages, we will also refer, selectively based mostly on fit, to evidence from peer review in other contexts, such as journal publications.

Elimination of peer review

Given the apparent severity of the problems of the peer review system, the most radical proposal entails its entire elimination as a way of distributing research funding. A rather controversial approach would be to distribute the available funding equally to all qualified researchers without selecting proposals (Ioannidis, 2011), effectively reverting fully to a block funding system for university research. Following this proposal, Vaesen and Katzav (2017) have calculated that an equalitarian distribution of research funding in the UK, the US and the Netherlands would provide each scientist enough money for research and travel costs, but would not be enough to support large and expensive research projects.

A system that would preserve the evaluation of proposals, but would eliminate the need for a peer review system, would be one relying completely of expert administrators to select directly proposals for funding, rather seeking advice from external experts. The US Defense Advances Research

Projects Agency (DARPA) currently uses this model, employing around 100 qualified programme managers to distribute about US$3 million every year. The managers are employed for a maximum of 4 years and have direct responsibility for the outcome of the funded projects, enabling a very fast decision project and potentially the ability to fund high risk and unconventional projects. Indeed, Azoulay et al. (2019) advocate this model built on organizational flexibility and significant authority given to programme managers as a promising means of funding breakthrough research. Numerous success stories related to DAPRA support this view. These include Wimmer's de novo synthesis of the polio virus in 2002, and funding early research laying the foundations for the modern internet, GPS and unmanned aerial vehicles (Tollefson, 2021). More recently, DAPRA provided financial support to Moderna in 2013 to pursue messenger RNA–based antibody drugs and vaccines (eventually leading to a successful Covid19 vaccine). Still, it is unclear how this model would extend to other contexts, which may not be able to recruit such a pool of expert programme managers. In 2018, the Wellcome Trust announced the creation of a DARPA-like fund for health and life sciences, the Wellcome Leap Fund. Initially endowed with £250 million, the fund has started operations in 2020 for an initial period of 5 years, with the explicit mission of funding high risk research with transformational potential and on an accelerated timescale. As a testament to the desire of replicating the DARPA model, Wellcome Leap is led by former DARPA director R.E. Dugan and will focus on early-stage, high-risk ideas, coupled with funding at scale.[9]

Partial randomization

In order to address the issue of reliability, especially concerning the "grey zone" of proposals discussed above, several authors have started advocating the use of partial randomization in grant allocation (Fang and Casadevall, 2016; Gross and Bergstrom, 2019; Roumbanis, 2019). Partial randomization (or "modified lottery") is a mechanism that complements peer review for allocating research funding, and it is applied only to a subset of already selected applications; hence, the "partial" aspect. First, peer review is used to identify applications that are worthy of funding (i.e., that are above a pre-defined cut-off of quality) and those that are not; then randomization is applied to select among the worthy applications. In variations of this practice, the randomization process is applied to the middle part of the distribution of the proposals, while those above a certain quality threshold are funded for sure (and those below a bottom threshold are excluded for sure from the competition).

Partial randomization is not yet part of mainstream practice, and therefore there are only few examples available, which makes it difficult to determine precisely its actual effectiveness (and fairness). Notably, the first two funding agencies to introduce randomization in their funding procedures were both located in New Zealand: the Health Research Council of New Zealand for its Explorers Grants in 2013, followed by the Science for Technological Innovation National Science Challenge in 2014. In Europe, the first attempt at was made by a private funder, the German Volkswagen Foundation, with the Experiment! Initiative from 2017 to 2020, followed by two public funding agencies, the Swiss National Science Foundation (initially with its Postdoc Mobility fellowship scheme and then

[9] See Wilkinson (2010) for the Trust's early plans to fund people over ideas

extending it to the other funding programs), and the Austrian Science Fund (with its 1000 ideas grant programme). So far, most attempts at partial randomization have been applied to relatively smaller schemes, providing seed funding for high-risk and potentially transformative early stage ideas, which are the ones more at risk of not getting funding through traditional peer review mechanisms. As an increasing number of funders experiment in this direction, more evidence should become available on the actual realization of the benefits promised by randomization, such as reduced bias and increased diversity, higher rewarding of unconventional ideas, and lower costs in the evaluation process. At the same time, it would be important to address a major concern about partial randomization, namely that it may seem as contradicting the traditional decision-making procedures in science based on merit. Funding agencies and grants administrators may be worried that politicians and society at large would react negatively to a lottery, interpreting it as a lack of will to do a thorough assessment (Barnett, 2016). Researchers may see it as going against the legitimacy of the scientific system itself, and at odds with long-ingrained ideas about the role of meritocracy in evaluation decisions (Liu et al., 2020; Roumbanis, 2019). At the same time, a lottery system may have some beneficial effects on scientists' persistence in applying for funding, as a rejection from chance may be less emotionally charged than a rejection by a group of your peers. Early qualitative evidence seem to suggest that both funders and younger generations of scientists are open to the idea of partial lotteries, but more research in this area is needed to understand how such a funding design would impact the actual behavior of scientists (Bendiscioli, 2019; Bendiscioli and Garfinkel, 2021).

Funding people rather than projects

Most peer review models base decisions on the assessed merits of the proposed project. This is for good reasons as for example such models can help permeate the Matthew effect in science where early success breeds later success irrespective of quality (Azoulay et al., 2014; Bol et al., 2018; Drivas and Kremmydas, 2020). However, also motivated by concerns of public accountability, these models can be rather conservative in promoting less novel works thus limiting the potential for more creative, breakthrough outcomes (Azoulay et al., 2011). Indeed, groundbreaking outcomes are more likely to come about under conditions of heightened freedom allowing for the pursuit of untraditional methods and techniques (Manso, 2011). With such background, funding, carefully selected, people rather that projects can potentially yield above par outcomes as long as these people are provided not only with enough resources but also, and perhaps more importantly, with adequate support including lax reviews, long time horizons to complete their research and high-quality feedback. While these are undoubtedly desirable outcomes, it is important to consider possible unintended consequences of a shift towards funding people rather than projects. In their study of the Canadian Institutes of Health Research grant programs between 2011 and 2016, Witteman et al. (2019) find that shifting the focus of the evaluation from the project to the investigator may be detrimental for women. The authors found that when reviewers primarily assessed an applicant's proposed science, no statistically significant differences existed between percentages of success for male and female principal investigators, while when they explicitly assessed the principal investigator as a scientist, the gender gap was significantly larger. This may be the result of individual bias (namely the reviewers' subjective evaluations of principal investigators reflecting conscious or unconscious gender bias) or systemic bias in terms of grant program design (for example review criteria that unfairly favors male principal investigators because of cumulative advantage). As well, funding people rather than projects, may exacerbate the concentration of funding to a handful of top performers even at the expense of early career researchers who had not had enough time to demonstrate their potential and skills (Peifer, 2017). In any case, such unintended effect needs to be taken into consideration when designing grant evaluation mechanism focusing on the researcher rather than the project.

A few funders offer grants focusing on the principal investigator rather than the proposal. The Howard Hughes Medical Institute (HHMI), a non-profit medical research organization, offers such grants through its investigator program which funds people not projects and this manifests in many ways including the continuation of funding when the investigators' early approaches do not work. Contrasting HHMI winners to similar scientists who were awarded National Institutes of Health (NIH) grants, which fund projects and are not particularly tolerant of early failure, Azoulay et al. (2011) reveal that HHMI awardees produce higher impact and more novel research. Another example is the The MacArthur Fellows Programme in the USA, where evaluators focus only on applicants' past performance, rather than judging the validity of a proposed project. This method is argued to reduce conservatism, as scientists wanting to change fields or with unconventional ideas can be supported on the basis of their past performance.

An open question with these sorts of funding schemes is their scalability. What works on a smaller scale might be too challenging to roll out broadly. For example, requiring thorough reviews and detailed feedback, as provided by HHMI, on federal funds would almost certainly burden even further qualified reviewers putting additional pressure to the system. Or, being more tolerable of failure on a large scale would likely raise concerns of public accountability and optimal use of taxpayers' money. Concerns of this kind, suggest that scalability of funding schemes resembling the HHMI awards are pertinent especially as the scientific enterprise has grown considerably over time (by a factor of twelve in the past fifty years in the U.S. (Stephan, 2012)).

Rivalrous peer reviews

The standard external/disinterested reviewer setup is subject to three (and largely separable) issues: a *motivation* problem, a *selection* problem, and a *reporting* problem. The *motivation* problem arises because qualified reviewers' agreement to review, even conditional upon their contributions being requested by the funding agency cannot be taken for granted. Self-interested, qualified reviewers may decline to participate in the peer review process to economize on effort, even as they endorse others' participation. Northcraft and Tenbrunsel (2011) conceptualize peer review as a *volunteer's dilemma* to shed light on the question of causation of bias in peer reviews. Reviewers weigh costs and benefits and decide whether to volunteer for the benefit of the whole group (some of which accrues to them as members) or freeride for their own benefit. Although freeriding improves everyone's outcome no matter what peers do, if everyone decides to free ride, the whole system falls apart, which is a suboptimal outcome for everyone. The attempt to provide the public good, despite widespread (and Nobel-winning, cf. (Ostrom, 1990)) research into incentive-compatible institutional mechanisms, frequently fails. The *selection* problem arises because even without the motivation problem, those who *agree* to review may not be the *most qualified* to do so (compared purely to a first-best, idealized assignment in which reviewer incentives did not distort their participation choices). The problem is that the most desirable reviewers typically have the highest opportunity costs in terms of publications, funding, or other activities forgone. Some reviewers may choose to participate for reasons other than the selfless promotion of science: to eliminate competition, to misappropriate ideas, or to favor (or stymie) certain applicants over others for reason unrelated to merit. Finally, there is the problem of *reporting*. Even reviewers without a direct financial stake in the rating of a given proposal—addressed through conflict-of-interest rules—may have preferences, beliefs, and expectations, conscious or unconscious, which affect review scores. Moreover, there is no reason to believe that upward or downward biases are independent and identically distributed, such that the aggregating shaded reviews does not alleviate the problem.

With these challenges as a conceptual background, Merrifield and Saari (2009) have proposed a rivalrous-review design that can simultaneously address the motivation problem, the selection problem, and the reporting problem while (a) maintaining a balance between reviews demanded and reviews supplied and (b) satisfying basic conflict-of-interest criteria. The basic characteristic of rivalrous reviewing is the requirement that submitters must review competing submissions, which eliminates the need to seek outside reviewers (who may be disinterested or less qualified to review). The fact that providing impartial, high quality reviews helps the reviewer's own project get funded can alleviate biases, as reviewers operate in an environment that provides nonmonetary rewards for honest, diligent, complete, and impartial reviews. Such a setup, designed so that providing honest and high-quality reviews is incentivized, can offer several advantages in the efficient and equitable production of scientific knowledge and in balancing the supply of and demand for qualified reviewers, even in traditional review formats. For instance, if reviewers' features (e.g., their gender and age) predict which reviewers will provide 'maverick' rankings which are accurate yet meaningfully different from the remaining rankings, a rules-based reviewer-classification scheme or a reviewer-recruitment and assignment system could be devised based on these features. Such a decision-support system could improve the efficiency and effectiveness of the merit review process, even within a traditional external review structure that does not provide explicit incentives.

This model has been tested before at the National Science Foundation in 2013 at the NSF Sensors and Sensing Systems (SSS) program for its October 2013 ("O13") round. In that funding round, each proposing PI was assigned to review 7 competing proposals. Reviewers received extra points on their own proposal if their assessment of the competing proposals generally matched the assessment of the same proposals by other reviewers (an intended bonus for "good" reviewing, defined here as agreeing with the consensus—a standard of inter-rater reliability). The rivalrous reviews design was disclosed *ex ante* to prospective submitters and applicants could thus voluntarily self-select into that round or defer the application to the next round in February 2014. The O13 round attracted roughly 40% more proposals than the equivalent round of the year before and each review had, on average, 40% more words—an indicator of more thorough reviewing (Mervis, 2014). Along the same lines, artificial intelligence analytical models reveal desired properties of such rivalrous reviews designs (Aziz et al., 2016; Kurokawa et al., 2015; Naghizadeh and Liu, 2013). Several telescopes assign at least part of their time through this system, such as the Gemini Observatory and European Southern Observatory (Andersen, 2020).

Anonymization

Blinding the identity of authors of proposals may provide a remedy against the most obvious reviewers' biases, such as those related to gender, age, ethnicity, and institutional affiliation (Lee et al., 2013). On the other hand, critics of the anonymous review process argue that retaining real anonymity may be infeasible, and that research proposals need to be evaluated in the light of the full research portfolio of the scientist submitting the proposal. As the concerns on the effects of biases on the evaluation of grant proposals are similar to those in the publication process, we can use some of the evidence from studies of double-blind peer review (as opposed to single-blind peer review, which is still the norm in many disciplines) in scientific journals. Evidence on the efficacy of double-blind peer review in reducing bias in practice is mixed. While some studies find that papers authored by famous authors and/or authors from prestigious institutions were rated higher and more likely to be recommended for acceptance in single-blind settings (Okike et al., 2016; Sun et al., 2021; Tomkins et al., 2017), others find that the effects may be inconsistent (Fisher et al., 1994) or non-linear, with authors from mid-tier institutions benefiting from single-blind review, while acceptance rates for authors from top and bottom rank institutions being unaffected by the type of peer review employed (Blank, 1991). Turning to gender bias, one of the most well-known contributions looking at "blinding" the evaluators to the gender of the evaluated, is the paper by Goldin and Rouse (2000) studying the impact of having musicians auditioning for orchestra playing behind a screen for a jury that cannot see them. The conclusion of this study highlight how blind auditions increased the chance of women to get hired: while it is frequently cited in discussions about gender discrimination, results are statistically weak and have been criticized heavily since. Similar results are however found in experimental studies: Brooks et al. (2014) found that both professional investors and nonprofessional evaluators preferred pitches presented by male entrepreneurs compared with pitches made by female entrepreneurs, even when the content of the pitch was the same; while McIntyre et al. (1980) found evidence of preferential treatment of males over females in preselection decisions based on CVs. A recent paper by Kolev et al. (2020) starts from the observation that despite blinded review, female applicants to the Gates Foundation receive significantly lower scores, which cannot be explained by reviewer characteristics, proposal topics, or *ex-ante* measures of applicant quality. The authors find that the gender score gap is no longer significant after controlling for text-based measures of proposals' titles and descriptions, suggesting that differing communication styles are a key driver of the gender score gap, which may render useless double-blinding solutions.

An example of anonymization of grant proposals to encourage the submission of unconventional ideas is a funding program launched in 2018 in Denmark by the Villum and Velux foundations, called "The Villum Experiment"[10]. As stated in the call, the program is "*created for those research projects out of the ordinary that challenge the norm and have the potential to change fundamentally the way*

[10] https://veluxfoundations.dk/en/technical-and-scientific-research/villum-experiment

we approach important topics". The rationale behind the anonymity of the proposals is to decouple the researcher's profile from the idea itself, allowing for taking greater risks and exploring new avenues for research. As stated on the Foundation 's website: "*The applicants are anonymous to their reviewers to sharpen the focus on the research idea and to let the researchers think freely in relation to their past merits"*. As it is intended to fund fundamentally new ideas, there is no evidence yet on its impact, but the Foundation has given a mandate to a public research organization to evaluate the program by 2024[11].

Conclusion

Scientific funding represents a multifaceted system in which a variety of actors operate, and studying it requires taking into account the complexity of the multilevel and dynamic relationships at play (Gläser and Velarde, 2018). In particular, three main categories of actors are of primary relevance: funding agencies, reviewers, and applicants. Funding agencies are 'intermediary organizations' (Braun, 1993; Braun, 1998) that regulate and influence researchers' behaviors and decouple them from direct political pressures. Funding agencies' nature, mission, and objectives direct researchers' efforts towards specific areas of research. For example, the shift towards providing solutions to societal problems and so-called 'grand challenges' has become particularly pronounced in the direct criteria for funding allocation (Gross et al., 1999), arguably diverting attention from 'blue-sky' research. Additionally, funding agencies contribute to establishing researchers' (and possibly research itself) legitimacy by designing and redesigning the standards by which research is evaluated. Different forms of assessment may indeed favor different types of applicants and applications. However, even the best-designed evaluation system depends on the availability and quality of reviewers. While peer review is widely considered the gold standard for evaluating research (Siler et al., 2015), it is also known to be afflicted by a variety of biases affecting evaluation outcomes and decreasing participation and representation opportunities in scientific discovery (Demarest et al., 2014). Finally, it is important to remember that researchers retain a special position in the scientific system because of their role in choosing the content and methods of their research. Decisions that researchers make in their everyday activities and about their research goals and objectives shape the way in which research is conducted and, inevitably, the knowledge that is produced (Knorr-Cetina, 1983). At the same time, the way in which evaluation is designed shapes the incentives of investigators, in terms of who decides to apply and with which proposals. The complexity of the system means that particular attention needs to be paid to the linkages between these actors, and clearly the mechanisms of evaluation of proposals and allocation of funds are central to the whole system.

As competitive funding become a condition *sine qua non* for participating in the scientific enterprise, its characteristics are bound to create and reinforce inequalities in the scientific system. Since the seminal works of Lotka (1926) and Merton (1968), scholars of science have been interested in the inequalities that characterize scientific fields, from levels of research funding to the availability of

[11] https://veluxfoundations.dk/en/content/significance-anonymous-application

equipment to graduate students and the recognition they receive (Stephan, 2012). Who gets to participate in the scientific enterprise (i.e. who gets funding) shapes the type and content of research conducted and, as governance arrangements change in science, it is crucial to understand if and how these changes will exacerbate or reduce inequalities in science.

It is therefore important to establish a new research agenda that will allow us to identify the relationship between specific elements of the review system for funding, the quality of the research outcome, and the participation of different groups of individuals in the scientific enterprise. Such an agenda would have relevant practical implications, as with science policy research in general, incremental improvements in decision making can have outsized practical results, both economic and scientific. For example, if alternative forms of peer review are able to provide more accurate and less noisy peer evaluations of the intellectual merit of proposals, funding will be more accurately devoted to supporting the meritorious rather than allocated to the lucky or, at least in some areas, the privileged. Even if privilege or bias is not the primary driver of variation in funding success, with the rest being pure randomness, more accurate targeting is still a desirable goal. Furthermore, if these alternative forms of review can reduce or even eliminate structural biases in the peer review process, access to funding by historically underrepresented populations, including women and minority scientists, is likely to expand. In order to do so, a close relationship between science and innovation scholars and funding bodies is necessary to establish new evaluation mechanisms and rigorous methods for testing their efficiency and fairness.

Besides evaluating alternative forms of peer review and tweaks of the current system, further testing its present state is a fruitful way forward because it can pinpoint to its strengths and weaknesses, thus allowing for more targeted interventions when necessary. For example, given significant differences among and within disciplines (Rahmandad and Vakili, 2019; Sauermann and Stephan, 2013) inconsistencies in the accessibility of data (and the type of data shared – e.g., content of reviews and proposals) across funding agencies limits our understanding of where the current system delivers desired outcomes and where it does not. Such differences in team size, publication strategies and the like may condition the effectiveness of the peer review system. For example, what works say in medicine or the life sciences may not necessarily work in physics or in economics. But, while the NIH, funding medicine, through strict protocols that safeguard confidentially, has provided access to both funded and unfunded proposals at a large scale (e.g., Jacob and Lefgren, 2011; Myers, 2020; Wang et al., 2019) the NSF, funding physics and economics, has not. This is in contrast with the fact that the NSF has been innovative in the way it funds science in part by having launched a secondment program since 1970 where academics, called rotators, on loan from their universities, run peer reviews and make funding decision on submitted proposals (Hoenen and Kolympiris, 2020; Kolympiris et al., 2019).

References

Andersen, M., 2020. Easing the burden of peer review. Nature Astronomy 4, 646-647.

Ayoubi, C., Pezzoni, M., Visentin, F., 2019. The important thing is not to win, it is to take part: What if scientists benefit from participating in research grant competitions? Research Policy 48, 84-97.

Ayoubi, C., Pezzoni, M., Visentin, F., 2021. Does it pay to do novel science? The selectivity patterns in science funding. Science and Public Policy 48, 635-648.

Aziz, H., Lev, O., Mattei, N., Rosenschein, J., Walsh, T., 2016. Strategyproof peer selection: Mechanisms, analyses, and experiments, Proceedings of the AAAI Conference on Artificial Intelligence.

Azoulay, P., Fuchs, E., Goldstein, A.P., Kearney, M., 2019. Funding breakthrough research: promises and challenges of the “ARPA Model”. Innovation policy and the economy 19, 69-96.

Azoulay, P., Graff Zivin, J.S., Manso, G., 2011. Incentives and creativity: Evidence from the academic life sciences. RAND Journal of Economics 42, 527-554.

Azoulay, P., Stuart, T., Wang, Y., 2014. Matthew: Effect or fable? Management Science 60, 92-109.

Banal-Estañol, A., Macho-Stadler, I., Pérez-Castrillo, D., 2019. Evaluation in research funding agencies: Are structurally diverse teams biased against? Research Policy 48, 1823-1840.

Barnett, A.G., 2016. Funding by lottery: political problems and research opportunities. MBio 7, e01369-01316.

Bendiscioli, S., 2019. The troubles with peer review for allocating research funding: Funders need to experiment with versions of peer review and decision-making. EMBO reports 20, e49472.

Bendiscioli, S., Garfinkel, M., 2021. Dealing with the limits of peer review with innovative approaches to allocating research funding. EMBO. https://www. embo. org/documents/science_policy/peer_review_report. pdf.

Blank, R.M., 1991. The effects of double-blind versus single-blind reviewing: Experimental evidence from the American Economic Review. American economic review, 1041-1067.

Bol, T., De Vaan, M., Van De Rijt, A., 2018. The Matthew effect in science funding. Proceedings of the National Academy of Sciences of the United States of America 115, 4887-4890.

Boudreau, K.J., Guinan, E.C., Lakhani, K.R., Riedl, C., 2016. Looking across and looking beyond the knowledge frontier: Intellectual distance, novelty, and resource allocation in science. Management Science 62, 2765-2783.

Braun, D., 1993. Who governs intermediary agencies? Principal-agent relations in research policy-making. Journal of Public Policy 13, 135-162.

Braun, D., 1998. The role of funding agencies in the cognitive development of science. Research Policy 27, 807-821.

Bravo, G., Grimaldo, F., López-Iñesta, E., Mehmani, B., Squazzoni, F., 2019. The effect of publishing peer review reports on referee behavior in five scholarly journals. Nature communications 10, 322.

Brooks, A.W., Huang, L., Kearney, S.W., Murray, F.E., 2014. Investors prefer entrepreneurial ventures pitched by attractive men. Proceedings of the National Academy of Sciences of the United States of America 111, 4427-4431.

Chetty, R., Saez, E., Sándor, L., 2014. What policies increase prosocial behavior? An experiment with referees at the Journal of Public Economics. Journal of Economic Perspectives 28, 169-188.

Crossley, M., 2015. The ins and outs of research grant funding committees, The Conversation.

Dasgupta, P., David, P., 1994. Toward a new economics of science. Research Policy 23, 487-521.

Demarest, B., Freeman, G., Sugimoto, C.R., 2014. The reviewer in the mirror: examining gendered and ethnicized notions of reciprocity in peer review. Scientometrics 101, 717-735.

Drivas, K., Kremmydas, D., 2020. The Matthew effect of a journal's ranking. Research Policy 49, 103951.

Eblen, M.K., Wagner, R.M., RoyChowdhury, D., Patel, K.C., Pearson, K., 2016. How criterion scores predict the overall impact score and funding outcomes for national institutes of health peer-reviewed applications. PLoS ONE 11.

Fang, F.C., Bowen, A., Casadevall, A., 2016. NIH peer review percentile scores are poorly predictive of grant productivity. Elife 5, e13323.

Fang, F.C., Casadevall, A., 2016. Research funding: The case for a modified lottery. Am Soc Microbiol, pp. e00422-00416.

Feinberg, R.M., Price, G.N., 2004. The funding of economics research: does social capital matter for success at the national science foundation? Review of Economics and Statistics 86, 245-252.

Feller, I., 2006. Multiple actors, multiple settings, multiple criteria: issues in assessing interdisciplinary research. Research Evaluation 15, 5-15.

Fisher, M., Friedman, S.B., Strauss, B., 1994. The effects of blinding on acceptance of research papers by peer review. Jama 272, 143-146.

Franzoni, C., Stephan, P., Veugelers, R., 2021. Funding risky research. National Bureau of Economic Research.

Ginther, D.K., Schaffer, W.T., Schnell, J., Masimore, B., Liu, F., Haak, L.L., Kington, R., 2011. Race, ethnicity, and NIH research awards. Science 333, 1015-1019.

Gläser, J., Velarde, K.S., 2018. Changing funding arrangements and the production of scientific knowledge: Introduction to the special issue. Springer, pp. 1-10.

Gluckman, P.D., 2012. Which science to fund: time to review peer review? Office of the Prime Minister's Science Advisory Committee.

Goldin, C., Rouse, C., 2000. Orchestrating impartiality: The impact of" blind" auditions on female musicians. American economic review 90, 715-741.

Graves, N., Barnett, A.G., Clarke, P., 2011. Funding grant proposals for scientific research: retrospective analysis of scores by members of grant review panel. Bmj 343.

Gross, C.P., Anderson, G.F., Powe, N.R., 1999. The relation between funding by the National Institutes of Health and the burden of disease. New England Journal of Medicine 340, 1881-1887.

Gross, K., Bergstrom, C.T., 2019. Contest models highlight inherent inefficiencies of scientific funding competitions. PLoS biology 17, e3000065.

Guthrie, S., Ghiga, I., Wooding, S., 2018. What do we know about grant peer review in the health sciences? F1000Research 6.

Guthrie, S., Rincon, D.R., McInroy, G., Ioppolo, B., Gunashekar, S., 2019. Measuring bias, burden and conservatism in research funding processes. F1000Research 8, 851.

Hegde, D., 2009. Political influence behind the veil of peer review: an analysis of public biomedical research funding in the United States. The Journal of Law and Economics 52, 665-690.

Herbert, D.L., Barnett, A.G., Graves, N., 2013. Australia's grant system wastes time. Nature 495, 314-314.

Herbert, D.L., Coveney, J., Clarke, P., Graves, N., Barnett, A.G., 2014. The impact of funding deadlines on personal workloads, stress and family relationships: a qualitative study of Australian researchers. BMJ open 4, e004462.

Hoenen, S., Kolympiris, C., 2020. The value of insiders as mentors: Evidence from the effects of NSF rotators on early career scientists. Review of Economics and Statistics 102, 852-866.

Ioannidis, J.P., 2011. Fund people not projects. Nature 477, 529-531.

Jacob, B.A., Lefgren, L., 2011. The impact of research grant funding on scientific productivity. Journal of public economics 95, 1168-1177.

Kaatz, A., Lee, Y.G., Potvien, A., Magua, W., Filut, A., Bhattacharya, A., Leatherberry, R., Zhu, X., Carnes, M., 2016. Analysis of national institutes of health r01 application critiques, impact, and criteria scores: Does the sex of the principal investigator make a difference'. Academic Medicine 91, 1080-1088.

Kaplan, D., Lacetera, N., Kaplan, C., 2008. Sample size and precision in NIH peer review. PLoS ONE 3, e2761.

Klaus, B., del Alamo, D., 2018. Talent Identification at the limits of Peer Review: an analysis of the EMBO Postdoctoral Fellowships Selection Process. bioRxiv, 481655.

Knorr-Cetina, K.D., 1983. The ethnographic study of scientific work: Towards a constructivist interpretation of science.

Kolev, J., Fuentes-Medel, Y., Murray, F., 2020. Gender Differences in Scientific Communication and Their Impact on Grant Funding Decisions, AEA Papers and Proceedings, pp. 245-249.

Kolympiris, C., Hoenen, S., Klein, P., 2019. Learning by seconding : evidence from NSF rotators. Organization Science 30, 528-551.

Kovanis, M., Porcher, R., Ravaud, P., Trinquart, L., 2016. The global burden of journal peer review in the biomedical literature: Strong imbalance in the collective enterprise. PLoS ONE 11, e0166387.
Kurokawa, D., Lev, O., Morgenstern, J., Procaccia, A.D., 2015. Impartial peer review, Twenty-Fourth International Joint Conference on Artificial Intelligence.
Lane, J.N., Teplitskiy, M., Gray, G., Ranu, H., Menietti, M., Guinan, E.C., Lakhani, K.R., 2021. Conservatism Gets Funded? A Field Experiment on the Role of Negative Information in Novel Project Evaluation. Management Science.
Lee, C.J., Sugimoto, C.R., Zhang, G., Cronin, B., 2013. Bias in peer review. Journal of the American Society for Information Science and Technology 64, 2-17.
Li, D., 2017. Expertise versus Bias in evaluation: Evidence from the NIH. American Economic Journal: Applied Economics 9, 60-92.
Li, D., Agha, L., 2015. Big names or big ideas: Do peer-review panels select the best science proposals? Science 348, 434-438.
Lindner, M.D., Vancea, A., Chen, M.-C., Chacko, G., 2016. NIH peer review: scored review criteria and overall impact. American Journal of Evaluation 37, 238-249.
Link, A.N., Swann, C.A., Bozeman, B., 2008. A time allocation study of university faculty. Economics of education review 27, 363-374.
Linton, J.D., 2016. Improving the Peer review process: Capturing more information and enabling high-risk/high-return research. Research Policy 45, 1936-1938.
Liu, M., Choy, V., Clarke, P., Barnett, A., Blakely, T., Pomeroy, L., 2020. The acceptability of using a lottery to allocate research funding: a survey of applicants. Research Integrity and Peer Review 5, 1-7.
Lotka, A.J., 1926. The frequency distribution of scientific productivity. Journal of the Washington academy of sciences 16, 317-323.
Manso, G., 2011. Motivating Innovation. Journal of Finance 66, 1823-1860.
Materia, V.C., Pascucci, S., Kolympiris, C., 2015. Understanding the selection processes of public research projects in agriculture: The role of scientific merit. Food Policy 56, 87-99.
McIntyre, S., Moberg, D.J., Posner, B.Z., 1980. Preferential treatment in preselection decisions according to sex and race. Academy of Management Journal 23, 738-749.
Merrifield, M.R., Saari, D.G., 2009. Telescope time without tears: a distributed approach to peer review. Astronomy & Geophysics 50, 4.16-14.20.
Merton, R., 1968. Social theory and social structure. Simon and Schuster.
Mervis, J., 2014. A radical change in peer review: Each NSF applicant evaluated seven competing proposals. Science 345, 248-249.
Mokyr, J., 2002. Innovation in an historical perspective: tales of technology and evolution, Technological innovation and economic performance. Princeton University Press, pp. 23-46.
Mutz, R., Bornmann, L., Daniel, H.-D., 2012. Heterogeneity of inter-rater reliabilities of grant peer reviews and its determinants: a general estimating equations approach. PLoS ONE 7, e48509.
Myers, K., 2020. The elasticity of science. American Economic Journal: Applied Economics 12, 103-134.
Naghizadeh, P., Liu, M., 2013. Incentives, quality, and risks: A look into the NSF proposal review pilot. arXiv preprint arXiv:1307.6528.
Nicholson, J.M., Ioannidis, J.P., 2012. Conform and be funded. Nature 492, 34-36.
Northcraft, G.B., Tenbrunsel, A.E., 2011. Effective matrices, decision frames, and cooperation in volunteer dilemmas: A theoretical perspective on academic peer review. Organization Science 22, 1277-1285.
NSF, 2017. Report to the National Science Board on the National Science Foundation's Merit Review Process Fiscal Year 2016.
Okike, K., Hug, K.T., Kocher, M.S., Leopold, S.S., 2016. Single-blind vs double-blind peer review in the setting of author prestige. Jama 316, 1315-1316.
Ostrom, E., 1990. Governing the commons: the evolution of institutions of collective action. Cambridge University Press, Cambridge.
Packalen, M., Bhattacharya, J., 2020. NIH funding and the pursuit of edge science. Proceedings of the National Academy of Sciences 117, 12011-12016.

Park, H., Lee, J., Kim, B.C., 2015. Project selection in NIH: A natural experiment from ARRA. Research Policy 44, 1145-1159.
Peifer, M., 2017. The argument for diversifying the NIH grant portfolio. Molecular biology of the cell 28, 2935-2940.
Pier, E.L., Brauer, M., Filut, A., Kaatz, A., Raclaw, J., Nathan, M.J., Ford, C.E., Carnes, M., 2018. Low agreement among reviewers evaluating the same NIH grant applications. Proceedings of the National Academy of Sciences of the United States of America 115, 2952-2957.
Polka, J.K., Kiley, R., Konforti, B., Stern, B., Vale, R.D., 2018. Publish peer reviews. Nature 560, 545 - 547.
Rahmandad, H., Vakili, K., 2019. Explaining heterogeneity in the organization of scientific work. Organization Science 30, 1125-1145.
Roumbanis, L., 2019. Peer review or lottery? A critical analysis of two different forms of decision-making mechanisms for allocation of research grants. Science, Technology, & Human Values 44, 994-1019.
Sattler, D.N., McKnight, P.E., Naney, L., Mathis, R., 2015. Grant peer review: improving inter-rater reliability with training. PLoS ONE 10, e0130450.
Sauermann, H., Stephan, P., 2013. Conflicting logics? A multidimensional view of industrial and academic science. Organization Science 24, 889-909.
Siler, K., Lee, K., Bero, L., 2015. Measuring the effectiveness of scientific gatekeeping. Proceedings of the National Academy of Sciences of the United States of America 112, 360-365.
Stephan, P., 2012. How economics shapes science. Harvard University Press.
Stephan, P., 2013. The Endless Frontier: Reaping What Bush Sowed?, in: Jaffe, A., Jones, B. (Eds.), The Changing Frontier: Rethinking Science and Innovation Policy. National Bureau of Economic Research.
Sun, M., Danfa, J.B., Teplitskiy, M., 2021. Does double-blind peer-review reduce bias? Evidence from a top computer science conference. arXiv preprint arXiv:2101.02701.
The_Royal_Society, 2007. Response to the RCUK consultation on the Efficiency and Effectiveness of Peer Review.
Tollefson, J., 2021. The rise of'ARPA-everything'and what it means for science. Nature.
Tomkins, A., Zhang, M., Heavlin, W.D., 2017. Reviewer bias in single-versus double-blind peer review. Proceedings of the National Academy of Sciences of the United States of America 114, 12708-12713.
Vaesen, K., Katzav, J., 2017. How much would each researcher receive if competitive government research funding were distributed equally among researchers? PLoS ONE 12, e0183967.
van den Besselaar, P., 2012. Selection committee membership: Service or self-service. Journal of Informetrics 6, 580-585.
Van Der Lee, R., Ellemers, N., Fiske, S.T., 2015. Gender contributes to personal research funding success in The Netherlands. Proceedings of the National Academy of Sciences of the United States of America 112, 12349-12353.
Veugelers, R., 2017. Supporting the next generation of biomedical researchers in Europe: The impact from ERC funding on career development for young biomedical researchers.
Wang, Y., Jones, B.F., Wang, D., 2019. Early-career setback and future career impact. Nature communications 10, 1-10.
Wessely, S., 1998. Peer review of grant applications: what do we know? The Lancet 352, 301-305.
Wilkinson, E., 2010. Wellcome Trust to fund people not projects. The Lancet 375, 185-186.
Witteman, H.O., Hendricks, M., Straus, S., Tannenbaum, C., 2019. Are gender gaps due to evaluations of the applicant or the science? A natural experiment at a national funding agency. The Lancet 393, 531-540.