The modified lottery: Formalizing the intrinsic randomness of research funding

Steven De Peuter, PhDa\* and Stijn Conix, PhDb

a Faculty of Psychology and Educational Sciences, KU Leuven, Leuven, Belgium

 ORCiD [0000-0003-4137-4431](https://orcid.org/0000-0003-4137-4431)

b *Centre for Logic and Philosophy of Science, Institute of Philosophy, KU Leuven, Leuven, Belgium*

 ORCiD [0000-0002-1487-0213](https://orcid.org/0000-0002-1487-0213)

\*Address for correspondence:

Steven De Peuter

University of Leuven

Methods, Individual and Cultural Differences, Affect and Social Behavior (MICAS)

Tiensestraat 102 box 3727

3000 Leuven, Belgium

[steven.depeuter@kuleuven.be](mailto:steven.depeuter@kuleuven.be)

[stijn.conix@kuleuven.be](mailto:stijn.conix@kuleuven.be)

ACKNOWLEDGEMENTS

Steven De Peuter thanks Prof. Dr. G. Storms for his comments on an early draft of the paper and for relentlessly propagating scientific integrity. The “Denkgroep Optimalisering Onderzoeksmiddelen” of the University of Leuven provided the inspiration for the paper and their report was most insightful.

The authors wish to thank S. Avin for his constructive, helpful, and challenging review of a previous version of the paper. His suggestions were most helpful and substantially improved the quality of the paper.

DECLARATION OF INTEREST STATEMENT

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. Steven De Peuter has no conflicts of interest to declare. Stijn Conix has no conflicts of interest to declare.

The Version of Record of this manuscript has been published and is available in Accountability in Research (ISSN 1545-5815) <http://www.tandfonline.com/> <https://doi.org/10.1080/08989621.2021.1927727>.

**ABSTRACT**

Competition for research funds has, in the recent decade, become hypercompetitive. Commonly, to determine which proposals receive funding, a system of peer review is used, which is broadly accepted, easily understood, and broadly trusted among researchers. It is often considered the best system in use, but it suffers from important shortcomings and adaptations to overcome these shortcomings have small and often short-lived effects. Hence, the preference for peer review does not mean it *necessarily* outperforms all other systems. In fact, it is time for an open discussion about alternative allocation mechanisms. Random allocation of research funding may be a viable alternative to the current peer review system. In particular the “organized randomness” of a *modified lottery* is interesting, combining the benefits of randomization with some of the most valuable aspects of peer review. Still, many questions remain and this is certainly not a plea to allocate all research funds using lotteries without further research. But we need to be prepared to consider alternatives, *even though they are not perfect*, and modified lotteries should be part of the solution.

Keywords: grant peer review; lottery; randomness; research funding; research integrity

**Introduction**

Competition for research funds has, in the recent decade, become hypercompetitive. An increasing number of researchers apply for funding, but budgets have remained more or less constant. Success rates are falling (van Noorden 2010): in the US, National Institutes of Health (NIH) success rates are around 20% and for some institutions closer to 10% (Fang and Casadevall 2016a).

Currently, the method most often used to select proposals is peer review: scientists with relevant expertise review, judge, and score proposals; their scores are used to rank the proposals and a cut-off (often based on the available funding) determines which proposals are funded. This system is widely accepted, easily understood, and broadly trusted among researchers (Philipps 2020): more than three out of four researchers agreed that peer review is the best way to allocate funding to the strongest proposals and the best research (Hayes and Hardcastle 2019). However, to say the least, it is far from perfect (Barnett 2016; Bendiscioli 2019; Chubin 1994) and it is time for an open discussion about alternative allocation mechanisms (Guthrie, Ghiga, and Wooding 2018).

In the following section, the most prominent problems of grant peer review are discussed. The second part of the paper discusses the modified lottery as a valuable alternative to peer review. For an extensive criticism of the grant peer review system, we refer the reader to the RAND report by Guthrie, Ghiga and Wooding (2018). In short: peer review is expensive, it is unreliable, it is biased, and it incentivizes breaches of research integrity.

# Why the Current Peer Review System is Problematic

## Peer review is expensive

With the current success rates, the majority of the time invested in drafting, writing, submitting, and re-writing applications for research funding goes to waste (Bendiscioli 2019; Hayes and Hardcastle 2019; Herbert et al. 2013; NIH 2008; Roumbanis 2019) . Surveying 285 scientists who together submitted 632 proposals to the Australian National Health and Medical Research Council (NHMRC) in 2012, Herbert and colleagues (2013) estimated that the 3727 proposals submitted that year took 550 working years to write. That amounts to AU$66 million annual salary cost, or 14% of the total NHMRC budget of AU$458 million – and only 21% of the applications were funded. These futile efforts, mainly at the expense of actually *doing* research, bring frustration and often even discouragement (Martin 2000). It has been argued that writing proposals assists scientists to reflect on their past, current, and future research; and that peer review is part of the organized scepticism that is under threat in the publish-or-perish climate (Reinhart and Schendzielorz 2020, but see Roumbanis [2020] for a strong counter-argument). However, using a mathematical model, Gross and Bergstrom (2019) show that as budgets are tighter and success rates decline, the value of the science researchers forbear by writing funding applications easily exceeds the funding budget.

Importantly, apart from applicants literally thousands of reviewers are also devoting substantial time to read, review, judge, discuss, and select the applications (Adam 2019; Bendiscioli 2019; Gillies 2014; Martin 2000). Surveying chief investigators who led NHMRC proposals, Graves, Barnett, and Clarke (2011) estimate that 85% of the cost of the 2009 project grants scheme was incurred by applicants, but 9% was due to peer review (and 5% of the costs were associated with administering the scheme). As Gillies (2014, p.7) puts it: if an average researcher spends 10% of her time on peer review, this increases the “true cost of research” by 11%. That more than 40% of reviewers declines an invitation to act as a reviewer can probably also be interpreted as a sign that reviewers are overburdened ([Publons](https://publons.com/community/gspr) 2019).

Peer review may have further indirect costs. Most importantly, it seems to lead to a distribution of research funding that is highly imbalanced. The result is that some researchers get more funding than they can effectively use, many get too little to do proper research, and most researchers are far removed from the funding “sweet spot” at which funders get most value for each dollar they spend (Mongeon et al. 2016). For example, Wahls (2018) shows that while the productivity of NIH researchers (measured in publications and impact factors) per dollar is highest around $400k per year, a select few get far more and many get considerably less.

## Peer review is unreliable

### Peer review lacks precision “behind the comma”

Peer reviewers typically assign one global and/or several sub scores to proposals. However, due to the imbalance between the number of proposals and budgets, proposals have to be differentiated based on decimal scores and raw reviewer scores don’t allow using average scores with such sensitivity (Chubin 1994; Fang and Casadevall 2016b). Using differences in average scores as small as 0.01 on the NHS one-to-five scoring scale would require more than 38,000 reviewers to achieve reliability (Kaplan, Lacetera, and Kaplan 2008). Even reliably differentiating proposals at the level of *one* decimal would require each proposal to be scored by 384 reviewers (Kaplan, Lacetera, and Kaplan 2008). Similarly, Mayo et al. (2006) found that at least 10 reviewers would have to evaluate each of the 32 proposals in their study to reach sufficient consistency.

Furthermore, there is a more fundamental objection to averaging reviewers’ scores: it assumes the scores not just represent an *ordinal* scale – “sufficient” is worse than “good”, “good” is worse than “excellent” – but an *interval* scale – i.e., there are similar distances between the scores (Sattler et al. 2015). However, considering that scoring criteria are often interpreted at reviewers’ own discretion (despite training, cf. next section), it is very unlikely that reviewer scores represent an interval scale.

### Peer review scores are ill defined and inconsistent

Typically, reviewers score several criteria, such as the quality of the proposal, its predicted scientific impact, societal relevance, and whether applicants have the expertise to successfully perform the proposed research. Increasingly, funders provide detailed instructions on how to interpret these criteria, but this does not avoid substantial subjectivity – not just in what each criterium entails, but also in how to weigh the different criteria in their final quotation (Avin 2019b; Luukkonen 2012; Pier et al. 2018). Reviewers differ in how they take strengths and weaknesses into account; and weaknesses influence final scores more strongly and more often than strengths (Pier et al. 2018). Experience does not necessarily attenuate this bias: whereas inexperienced reviewers may have insufficient knowledge and understanding of both criteria and scores, experienced reviewers tend to overestimate their understanding and deviate deliberately from scoring criteria (Sattler et al. 2015).

Avin (2019a) adds that although scientific peers may admittedly be best placed to estimate the scientific merit of proposals, they cannot reliably do so because what they need to estimate is in the future (see also Brezis [2007], Gillies [2014], and Mallapaty [2018]). What reviewers *can* do, is base their prediction on their experience with previous, similar projects and publication records, but this is problematic due to prediction error and regression to the mean.

Furthermore, and contrary to publication peer review, grant peer reviewers often have relevant expertise but are most often not *the* field experts. However, in particular for outstanding proposals, scores from field experts are more meaningful than an average score from groups of less specialist reviewers (Boudreau et al. 2016). This is further exacerbated by reviewer fatigue, which often makes it hard to find reviewers with relevant expertise.

As a result, there is limited agreement between reviewer scores (Graves, Barnett, and Clarke 2011), to the extent that Pier et al. (2018) have argued that the peer review becomes completely random above a certain quality threshold. For example, more than three quarters of the variance in scientific merit scores can be attributed to a combination of the specific reviewer, the interaction between reviewer and application, and random noise (Gallo, Sullivan, and Glisson 2016). Consequently, receiving favorable reviewer scores involves *luck*. Consider the following illustrations: variability in reviewer scores could shift about one third of the funding decisions by the NHMRC from “fund” to “not fund” (or vice versa; Graves, Barnett, and Clarke 2011); and statistically correcting for the uncertainty included in NIH peer review scores could push up to 25% of the proposals from “not funded” to “successful” (and vice versa; Johnson 2008). Especially when proposals are evaluated by a small number of reviewers, the choice of reviewers can “make or break” success: drawing pairs out of the scores of 11 reviewers who scored 32 applications at the Canadian McGill University Health Center Research Institute, Mayo and colleagues (2006) showed that even top rated proposals could easily fail to meet the funding cut off. As such, the evidence is firmly against Reinhart and Schendzielorz’s (2020) defense of *legitimacy* as one of peer review’s strongest assets. Peer review may rely on “the result of critical deliberation according to scientific criteria” (p. S27), but if its results are unreliable, its legitimacy becomes null (Roumbanis 2020).

There is some evidence that peer review is reasonably reliable when it comes to the projects with the best reviewer ratings. Thus, peer review may be a good method to identify excellent projects (Fang & Casadevall 2016b; Li & Agha 2015). In that case, peer reviewers’ added value to the funding process lies in their ability to identify the strongest applications (Li & Agha 2015). However, the existing evidence is rather weak and inconclusive (Guthrie, Ghiga, and Wooding 2018) and further research with more robust designs is needed to draw strong conclusions about the cost-effectiveness of peer review.

### Peer review is biased

Peer review is biased (Fang and Casadevall 2016b; Martin 2000), even despite efforts to prevent it from being so. Although probably only a minority of peer reviewers is deliberately or consciously biased, peer review is nevertheless favoring certain types of research and researchers.

*The usual suspects.* Although formal studies are scarce and the available evidence is mixed (Guthrie et al. 2019), peer review is supposedly disadvantaging researchers based on race and ethnicity (Ginther et al. 2011); gender – even in blinded review because males tend to use more “vague” language (Else 2019) which increases chances of success (Kolev, Fuentes-Medel, and Murray 2019; Pohlhaus et al. 2011; but see Sato et al. [2020] for an interesting discussion of gender bias in grant review); seniority; and the institute of the applicant (Brezis 2007; Daniels 2015; Pier et al. 2018). Cronyism, or rewarding grants to collaborators or researchers from a reviewer’s institution (or scientific field) is prevalent (Guthrie, Ghiga, and Wooding 2018; Jang et al. 2017; Mom, Sandström, and van den Besselaar 2018). Similarly, an applicant’s publication record, citations and grants, and reputation generate bias (Guthrie et al. 2019; for an in-depth discussion of these metrics as “perverse incentives”, see e.g. Nosek, Spies, and Motyl [2012]).

Lack of reliability and bias can together create a growing divide between initially equally skilled and talented researchers who have to compete for the same scarce funding. The variability in, and inconsistency of, reviewer scores may put the proposal of one of those researchers just above the funding line and the other below, whereas *in reality*, their proposals are equally strong. The simple fact of getting this (first) funding may give the funded researcher the aura of being “more successful” or “worth funding”, increasing their future chances of being funded (the so-called Matthew effect; Bol, de Vaan, and van de Rijt 2018). The less fortunate researcher will lack this aura and be obliged to keep investing time in new funding applications, taking time away from actually doing research – and building the necessary track record to prove their scientific merit. This can easily result in a downward spiral (Martin 2000) 1.

*Innovative research is disadvantaged.* Innovative research – “blue sky” research; “edge science” – is strongly disadvantaged by peer review (Adam 2019; Brezis 2007; Gillies 2014; Hayes and Hardcastle 2019; Li and Agha 2015; Packalen and Bhattacharya 2020) and peer review promotes low-risk research that fits within existing paradigms (Fang 2011; Nicholson and Ioannidis 2012). Packalen and Bhattacharya (2020) screened the MEDLINE database for biomedical research papers that mention NIH funding and found that there is a 7 to 10 year-gap between the emergence of new ideas and those ideas receiving maximal NIH funding.

This may come as no surprise: as it treads new ground, both the specific outcomes of novel research and the probability that these outcomes will be obtained are more difficult to predict, possibly leading to worse scores for “expected outcomes” and for “risk” (Brezis 2007). Furthermore, reviewers’ definition of “excellent” may be at odds with “paradigm-changing”: the former implies a degree of methodological rigor and clarity of purpose that the latter is inherently lacking (Brezis 2007; Gallo et al. 2018; Luukkonen 2012). In particular when prevailing paradigms are challenged, reviewers’ resistance can be fierce, as is illustrated by the anecdotal stories of Nobel prize winners and their struggles to get their research published (cf. “John Snow’s Grant Application”, Rothman 2016; Gillies 2014). As we already mentioned, even a single aberrant reviewer score can push a proposal below the funding line (Kaplan, Lacetera, and Kaplan 2008) – especially when reviewer panels need to reach consensus.

Moreover, scoring applicants’ experience for innovative science can create a double handicap: innovative research is novel by definition, hence relevant experience is often lacking. Reviewers could then take the track record of applicants into account – their previous successes may suggest they know what they are doing – but this obviously disadvantages courageous early career scientists.

*Interdisciplinary research is disadvantaged.* Peer review often underestimates the scientific merit of interdisciplinary research (Pluchino et al. 2019). Bromham, Dinnhage, and Hua (2016) used the number of Field of Research codes from more than 18.000 applications to the Australian Research Council’s Discovery Programme as a proxy of interdisciplinarity and report a negative correlation between their interdisciplinarity measure and the chance to get funded.

Furthermore, reviewers tend to be wary to review interdisciplinary proposals because they don’t feel qualified to evaluate the research if they lack expertise in one or more fields of the application. Unfortunately, this results in proposals being reviewed by scientists with possibly even less relevant expertise (Hayes and Hardcastle 2019).

In addition, review criteria and reviewer scores are inconsistent between disciplinary fields, which renders interdisciplinary applications’ scores difficult to combine into a global score (Luukkonen 2012; Roumbanis 2020).

## Peer review incentivizes breaches of research integrity

Because funding distributed by peer review is often assumed to be entirely merit-based, successful grant applications have come to serve as an indicator of academic success. As such, they play an important role in hiring, promotion, and even further grant applications. In combination with low success rates, this generates a strong form of competition between researchers that is – to say the least – not conducive to research integrity..

For example, it is relatively common to submit highly similar or identical proposals in multiple competitions, so-called ‘double-dipping’, to increase the chances of success or even get multiple grants for the same research (Garner et al. 2013). Applicants are incentivized to inflate their publication list and boost their curricula in other ways to increase their chances (Dinov 2020; Guthrie et al. 2019). Often, not all authors of an application are properly acknowledged, mostly because senior researchers are under high pressure to apply for as many grants as possible, and thus have to rely on junior staff to write applications for them. Indeed, it is not uncommon for labs to hire someone specifically to prepare grant applications. Furthermore, while empirical research on this is missing, it seems likely that grant applications are often dishonest about their importance and potential, either by over-promising results or by exaggerating the expected impact (Serrano Velarde 2018).

Whereas the practices listed above are clear violations of generally accepted rules of research integrity, there are also a range of behaviors connected to peer review that, even though generally accepted, are in clear tension with research integrity. Most notable here is “grantsmanship”: the skill of writing attractive projects (Rasey 1999), which has little or no relation to a researcher’s ability to do sound research. Grantsmanship distracts from the content of a proposal – the very thing reviewers should be evaluating – and adds noise to an already noisy system by exploiting factors that reviewers cannot help to be influenced by (Dinov 2020). The paradoxical importance of grantsmanship is reflected in the administrative resources universities devote to it, in the manuals that are offered2 – and in commercial parties that advertise substantial higher success rates than what is common. The role of grantsmanship is particularly problematic because science funding is a zero-sum game: Superior grantsmanship in one proposal is likely to push other well-deserving candidates below the funding-threshold on the basis of criteria that should not be part of the evaluation.

Other common breaches of research integrity in the context of peer review relate to dishonesty and accountability. As already mentioned, reviewers are asked to predict the success of research proposals – something that is known to be nearly impossible (Mallapaty 2018) – and even to do this with a degree of precision that enables them to choose between many excellent proposals. In addition, reviewers in panels typically have to read thousands of pages of applications, often outside their direct area of expertise and with little or no compensation (Herbert et al 2013). Thus, they are incentivized (or even forced) to skim through applications rather than read them thoroughly. The result is that reviewers have to make decisions they cannot justify, and report them with more confidence than they (can) have. Given the importance of these decisions for the applicants – their careers depend on it – this lack of justification poses unquestionable ethical problems. There are similar problems of accountability on the applicants’ side: applicants are forced by funding agencies to make precise predictions in the shape of timelines, milestones, deliverables, outcomes, and workplans, even if they know that such predictions are very difficult or impossible to make. These ill-justified predictions are then used to arbitrate between proposals, again feeding into problems of accountability on the reviewers’ side.

Finally, it is also worth pointing out that there is a broad ethical dimension to many of the other problems of peer review that we discussed in previous sections. That is because society values, supports, and funds science ultimately for the impact it has on society. If peer review is epistemically suboptimal – because it is costly, discourages innovation, or fails to fund the best research – this inevitably has an ethical opportunity cost as well, namely, the impact that science could have had on society but did not because societal funds were wasted.

# Improving Peer Review

A list of changes and modifications have been suggested and tested to deal with the aforementioned problems of peer review (for an overview, cf. NIH 2008), with varying degrees of success.

## Solving inconsistencies

First of all, reviewer training in rating and weighing proposals’ various aspects has been suggested to increase inter-rater reliability (Sattler et al. 2015). Similarly, training has been suggested to better detect flaws and weaknesses in applications (Schroter et al. 2004). However, despite reviewers’ enthusiasm to receive training, its effects are limited (Pier et al. 2018) and short-lived (Schroter et al. 2004). One might assume that reviewer *experience* may improve peer review. Indeed, experienced reviewers can extract more information from an application (Brezis 2007), but this also increases the chances that they discover weaknesses and possible flaws they subsequently tend to weigh heavier than possible strengths (Boudreau et al. 2016).

Having each proposal reviewed by more reviewers would increase the reliability of the average scores and rankings, but as discussed before this would substantially increase the necessary resources and achieving the precision to reliably rank projects requires an unrealistically large pool of reviewers (Kaplan, Lacetera, and Kaplan 2008; Mayo et al. 2006).

To avoid a single bad score ruining an applicant’s chances, Daniels (2015) suggested to remove the lowest scores from review panels. This would also even the odds for more daring proposals that cannot convince every single reviewer. Relatedly, Kaplan, Lacetera, and Kaplan (2008) and Linton (2016) suggested to use differences in reviewer scores to detect innovative or high-risk/high-return projects. However, to date only a single study tested this claim: Barnett, Glisson, and Gallo (2018) calculated the relative citation ratio of peer-reviewed publications associated with 227 applications funded by the American Institute of Biological Sciences and investigated whether it was associated with greater reviewer disagreement (standard deviation and range or scores). Surprisingly, they found no association. Given that this is the first study and it only included projects that received funding (11% of 2063 applications), the authors indicate that more research is needed to draw firm conclusions.

Holliday and Robotin (2010) even suggest to implement a reverse strategy: instead of selecting the best applications, they suggest to remove the weakest projects, resulting in the “survival” of only the very best projects after several rounds.

An interesting, although for the time being theoretical, proposal from Bedessem (2020) is to let reviewers select the projects they want to see funded out of a pool of applications and to fund applications that are unanimously selected. Bedessem also suggests expanding reviewer panels to include the target audience: e.g. patients, politicians, farmers,… to make scores on “societal relevance” more tangible (see also Fleurence et al. [2014], Holbrook and Frodeman [2011], and Martin [2000]). However, this would obviously increase the cost of the review process.

## Reducing the cost/burden

Shorter applications may alleviate some of the burden of proposal writing, but applicants may invest just as much time in the shorter applications (Barnett et al. 2015) – trying to “just get it right” (Shepherd et al. 2018). A two-step review process, inviting full proposals only after screening brief initial applications is used by e.g. the European Research Council (ERC) and Cancer Research UK (CRUK). In a study evaluating UK’s National Institutes for Health Research’s Research for Patient Benefit (RfPB) Programme, a two-step process was shown to be 30% cheaper and more efficient. Moreover, applicants who were not selected received the decision sooner in the process – although the final decisions about the full proposals were delayed in time (Morgan et al. 2020).

Online reviewer panels instead of on-site meetings reduce the burden on reviewers, increase efficiency, and save substantial time and travel costs (Shepherd et al. 2018). Considering that panel discussions do not improve the reliability of grant evaluations (Fogelholm et al. 2012), dropping them may lead to even larger savings.3

# Introducing Organized Randomness

In sum, grant peer review suffers from important shortcomings and suggested solutions have small, if any, and often short-lived effects (Fang and Casadevall 2009) – as Roumbanis (2020) puts it: “some of the dilemmas inherent in peer review are almost impossible to neutralize” (p. S130). Hence, although peer review may be perceived as the best system in use, this does not mean it necessarily outperforms all alternatives. In fact, it is time for an open discussion about alternative allocation mechanisms and we could start by admitting that there is little evidence that the current system is the best (Guthrie, Ghiga, and Wooding 2018).

Already in 1998, Greenberg proposed to formalize the “chance” aspect that is inherent to peer review by allocating research funding *randomly*, i.e. using a lottery. He proposed to investigate the validity of a lottery in allocating research funding by slicing off “some respectable percentage of the research funds—say, 15–20% over 5 years—and set them aside. These funds would be awarded by lottery to applicant scientists whose qualifications and projects have been certified as respectable, ratings easily determined at a small fraction of the cost of peer review” (p. 686).

Brezis (2007), starting from the assumption that reviewers are limited in the amount of information they can distill from grant applications, especially innovative or groundbreaking applications, suggested something she termed *focal randomization*. In her view, projects that are ranked at the top by all referees should be accepted, projects unanimously ranked at the bottom should be rejected, whereas projects for which reviewers do not agree should be selected by a lottery – because the latter may be truly innovative projects with a possible high return.

Recently, based on their research of the productivity and return of NIH funded projects (e.g. Fang, Bowen, and Casadevall 2016; Fang and Casadevall 2009), Fang and Casadevall (2016a) have plead for NIH to implement this kind of *modified lottery*. It would include an initial *triage* phase to exclude researchers with insufficient competence to perform the research (Gillies 2014; Greenberg 1998) – which could be “fairly routine matter” (Gillies 2014, p.6) – and to remove proposals with obvious shortcomings (Fang and Casadevall 2016b; Liu et al. 2020). Subsequently, a rudimental review of proposals allows outstanding proposals to bypass the lottery – and proposals unanimously ranked at the bottom to be rejected (Brezis 2007). In the middle tier of proposals that are difficult to discriminate (Graves, Barnett, and Clarke 2011) or for which there is disagreement between reviewers (Brezis 2007) resources are assigned randomly. If, however, budgets are so tight that there are more outstanding proposals than the payline allows, Fang, Bowen, and Casadevall (2016) suggest to organize a lottery in the top tier nevertheless.

Boyle (1998) proposed to use a *graduated lottery*: after a rudimental ranking of proposals, for example 50% of the proposals in the lowest quartile of admissible proposals is randomly selected. These selected proposals are added to the second-lowest quartile and again 50% of the proposals from this new pool is selected. These are added to the second-highest quartile, etc. This way, proposals with a higher ranking would have an increasing chance of being funded without relying heavily on review scores.

Gross and Bergstrom (2019) suggest this could be doneusing *shortened* and/or *simplified* proposals: it reduces the writing effort (Herbert et al. 2013), whereas review panels would still have the opportunity to reward excellence, which obviously resonates with what Brezis (2007) proposed.

## Real-life case studies and researchers’ attitudes towards random funding

Together with his proposal to experiment with random allocation of research funding, Greenberg (1998, p. 686) proposed: “After a few years, let’s look back and evaluate the science that came out of this system”, something Gillies (2014) more recently subscribed to, suggesting a lottery should be experimented with to evaluate the research that comes out of it.

However, for the time being, only a few funders have implemented randomness: the Health Research Council of New Zealand with its Explorer Grant funds6 and the New Zealand’s government’s Science for Technological Innovation National Science Challenge (SfTI); the Swiss National Science Foundation (SNSF) postdoctoral fellowships for early-career scientists (Adam 2019); and the Volkswagen Foundation’s Experiment! Grants7. It is the explicit aim of the latter to compare the lottery to the more traditional selection process that runs in parallel, (finally) answering to Greenberg’s (1998) call to test random allocation of funding empirically. In this respect, the study to further evaluate the return on investment of different funding mechanisms of Barnett et al. (2015) is important and necessary.

A recent survey study showed that New Zealand’s researchers generally evaluate the random allocation of Explorer Grant funds positively, although predominantly funded researchers were enthusiastic: 78% thought random allocation of funding is acceptable, compared with 44% for declined applicants (Liu et al. 2020). In general, 63% of those surveyed said that random allocation of funding is acceptable for Explorer Grant funds, but only 40% favored the idea for other grant types (Liu et al. 2020).

Philipps (2020) concluded from his semi-structured interviews with 32 researchers in the field of physical and life sciences – some of which were awarded an Experiment! Grant by lottery – that researchers would be willing to give random grant allocation a try. However, they felt that at least a minimal evaluation mechanism should be used to ascertain the quality control that is perceived as a crucial aspect of peer review – which is fully consistent with the idea of a modified lottery.

Interestingly, Herbert et al. (2013) assessed researchers’ expectations of peer-review reliability with a hypothetical scenario: suppose two different panels of 10 experts would select 20 proposals for funding out of a pool of 100 applications, what is the acceptable difference between the two selections (i.e. how many *different* proposals would be acceptable)? The majority of respondents indicated to be willing to accept two to five different proposals – and the most frequently selected response was five, which corresponds to a 25% disagreement in the funding decision. This seems to imply that researchers are already aware of, and willing to accept, the “chance” aspect that is inherent to peer review.

## Mathematical models

A number of scholars have modelled the scientific and economic return of research funding, comparing peer review with a modified lottery. Brezis (2007) included numerical examples to show how her system of focal randomization creates a larger economic return than that of aggregated reviewer scores (see also Linton [2016]).

Additionally, although the agent-based mathematical model of Avin (2015) inevitably represents highly idealized versions of reality, it showed that funding based on initial triage (separating outstanding from good proposals) and subsequently allocating funding randomly in the latter group *outperforms* peer review. The crucial differences between both procedures are that random allocation is cheaper (Gillies 2014); and more fair because proposals that cannot be statistically differentiated from one another are being treated as equal.

Building on Avin’s model, Harnagel (2019) created a mid-level model between “theoretical” and “empirical” by integrating real-world publication and citation counts from Web of Science from 2006 to 2016 and showed that peer review was *not* the best model to generate significant science. Instead, randomness would benefit scientific creativity.

Gross and Bergstrom (2019) likewise showed that randomly allocating funding to proposals that pass a minimum quality threshold would decrease the process’ costs. Because applicants would cease to overly invest in their attempts to write the “best” application, they would also sacrifice less of their valuable and costly research time. The model showed that a lottery in which 45% of applicants qualify is as efficient as peer review selecting the 45% “best” proposals (independent of the chance of getting funded in the lottery). Furthermore, with the current emphasis on individual achievements for evaluation and promotion some researchers may increasingly apply for funding *even though* *this hampers* their scientific productivity (cf. our discussion of research integrity above), further reducing the efficiency of the peer review system. Although Gross and Bergstrom admit that their model did not (yet) include the possibility to resubmit partial revisions after first rejections, they are resolute: the smaller the success rate, the larger the benefits of a lottery.

With a similar model, Pluchino, Biondo, and Rapisarda (2018) and Pluchino et al. (2019) showed that although a certain amount of talent and skill is *necessary*, it is not a *sufficient* condition for success. Contrary to common belief, it also takes luck to be successful5. In fact, moderately talented individuals have a better chance to “make it”: because there are more moderately talented researchers than super talented, the former’s chances to be lucky are higher. These findings, importantly, also call into question the meritocratic model and the metrics that are used to evaluate applicants’ track record (Osterloh and Frey 2020) – and drive the Matthew effect.

## The case for a modified lottery

Considering these findings, a modified lottery may be a viable alternative to the current peer review system to allocate research funding. Various reasons stand out.

First, a modified lottery could save applicants, reviewers and funders considerable time they would otherwise devote to write, re-write, read, evaluate, and select proposals (Herbert et al. 2013; Gillies 2014; Greenberg 1998). A system that uses simplified, brief applications can relieve some of the hyper competition (Roumbanis 2019); knowing the selection is based on a lottery can also to some extent restore the balance between the time spent writing grant proposals and doing research (Gillies 2014; Gross and Bergstrom 2019; Herbert et al. 2013). The time saved could be devoted to actual science, the money saved could be used to fund research: It would sharply increase the process’ cost-efficiency.

Evidently, modified lotteries also require substantial resources to write, read, evaluate, and rank the applications and whether they will actually reduce the time applicants actually put in their applications has to be investigated (Philipps 2020). One way to minimize the costs of a modified lottery is reduce the triage stage to a formal check that only removes applications that do not meet a list of easily checkable criteria (Gillies 2014). This reduces costs on the funders' side, and removes the incentive for applicants to put a lot of time in their short applications. The downside of this solution is that there is less quality control and excellence is not rewarded. However, given the potential gains in efficiency and thin evidence that reviewers can actually identify the best proposals (Fang et al. 2016), this may well be a viable and cost-effective solution. In general, any modified lottery has to find a balance between the cost of peer review and the ability to reward excellence. Should future evidence indicate that (only) peer review can select excellent research, more peer review-based quality control could be included; inversely, the role of peer review in lotteries could be reduced if future research suggests that peer review does not effectively identify excellent research after all.

Second, modified lotteries would reduce bias. As Stone (2009) argued, because random allocation avoids decisions being based on reason(s), it side-lines *bad* reasons (i.e., bias). Thus, a modified lottery could overcome the conservative bias (Gillies 2014) and drive innovation in science by encouraging novelty (Brezis 2007; Avin 2019b). Similarly, a modified lottery would probably reduce other biases of the peer review system as well (Fang and Casadevall 2016a). Because of this, a lottery would make the process of research funding more divers, fair and impartial and might even make science itself, as an institution, fairer (Gildenhuys 2020, Roumbanis 2019). Of course, because lotteries also preclude *good* (i.e., relevant) reasons from guiding funding decisions, their use must be carefully considered.

A modified lottery further allows to “tweak” some of the selection criteria – levelling the playing field for minority groups; or counteracting reviewer bias without imposing explicit quota (Adam 2019). Furthermore, specific tracks could be created for specific groups. For example, the NIH has funding streams for early career scientists4 (Daniels 2015). Importantly, by defining those selection criteria, they become explicit, instead of being intertwined with others and obscured in (sub)scores. The most important advantage of this approach is, no doubt, that relevant aspects can be accounted for without pretending to have accurate estimates.

Third, a modified lottery would create a more realistic image of the merit of achieving funding (Gross and Bergstrom 2019) and prevent others from experiencing a rejection as a personal failure (Osterloh and Frey 2020). It would create a “substantiated meritocracy” and reduce competition – which would promote scientific integrity and counteract the Matthew effect.

Finally, a lottery could increase the visibility of the shrinking success rates and the vast amount of potentially productive and innovative research that is currently *not* able to secure funding (Fang and Casadevall 2016a).

Importantly, as Greenberg (1998) already stated: funding by lottery doesn’t necessarily need to *outperform* peer review in terms of productivity of the research it selects. Because of its supposed benefits, it should get a fair chance even if it *does not perform worse*. If it does get a chance, implementing it comes with a range of operational questions that are outside the scope of this paper. Should eligible but unsuccessful applications be automatically added to the next round (and for how many rounds; Barnett et al. 2014; Fang and Casadevall 2016a), or should they be resubmitted, immediately or after a waiting period? Should the number of proposals per applicant be limited (Fang and Casadevall 2016a)? Should there be a cooling-off period after several unsuccessful applications (Graves, Barnett, and Clarke 2011; Barnett et al. 2014)?

The list of unanswered questions is long (but Sloman [2014] gives a detailed example of how time, received funding and researcher output can be used to determine the lottery’s parameters). However, these same questions have to be answered when setting up grant peer review. It seems it is only fair to conclude that, regardless of the practical details, a modified lottery is more fair, closer to the reality of the allocation process, and potentially cheaper than peer review.

# Conclusion

This may be the time to admit that the current grant allocation system may have been performant when it was installed, but it doesn’t fit the current highly competitive research arena. We need to be prepared to consider alternatives, *even though they are not perfect* (Guthrie, Ghiga, and Wooding 2018).

Introducing “organized randomness” into the allocation of research funds could be a way to overcome some of the problems of the current grant peer review system. Importantly, it would not mean we do not want to thoroughly evaluate the research we are financing or give into despair about the current system (Reinhart and Schendzielorz 2020). Instead, it would demonstrate we are committed to admitting and tackling the shortcomings and limits of the current system (Barnett 2016; Gillies 2014).

Organized randomness could come with reduced costs, and even if the initial triage phase and subsequent lotteries (and the administration of resubmissions etc.) would not reduce the costs and funder resources, it could overcome some of the bias that is inherent to peer review (cf. Bedessem 2020, but see also Barnett et al. [2017]; Bollen et al. 2014).

Still, many questions remain and this paper is not intended as a plea to allocate all research funds using lotteries without further research (Gillies 2014). The trade-off between the cost of peer review and its potential to select excellent research requires further investigation. Some research fields can only thrive with long-term funding, for example when large or expensive equipment is necessary (Avin 2019; Stone 2009; see also Roumbanis [2019]). Needless to say also modified lotteries have loopholes and may be subject to bias. But many of the shortcomings of modified lotteries are part of the peer review system as well, so it wouldn’t be fair to use the argument to dismiss only modified lotteries – which would represent precisely the bias against novelty (or resistance against paradigm shift) formal randomness could help overcome (Avin 2019b; Gillies 2014). Modified lotteries are part of the solution, not *the* solution (Roumbanis 2019). As Greenberg (1998, p. 686) put it: “A lottery might turn out to be the worst possible system, except for all others.”

NOTES

1. Some researchers seem to benefit from an initial setback in their career, working even harder and becoming more successful than researchers who received funding early in their career, but that is a small minority (Wang, Jones, and Wang 2019).

2. Even the Office of Research Integrity offers advice on its website: <https://ori.hhs.gov/education/products/wsu/writing_gra.html>

3. Other alternative systems of funding research have been proposed, such as egalitarian distribution of funding (Ioannidis 2011), setting up a system in which funding is distributed “bottom-up” by researchers (Bollen et al. 2014), or financing researchers instead of projects (Ioannidis 2011). The latter has been implemented by the Howard Hughes Medical Institute (HHMI – <https://www.hhmi.org/programs/biomedical-research/investigator-program>) awarding great freedom to experiment, being failure-tolerant, and rewarding long-term success. As a result, HHMI funded researchers produce more high-impact publications than a group of similarly accomplished researchers receiving NIH funding (Azoulay, Zivin, and Manso 2011), although they also produce a larger number of “flops”.

4. New Innovator Awards (DP2) for exceptionally creative early stage investigators, the Early Independence Awards (DP5) for young scientists to pursue independent research immediately after a terminal degree, and the Pathway to Independence Awards (K99/R00), which combine a mentored research phase with later, independent research support.

5. Janosov et al. (2020) use a similar model to show that the role of luck in science is no different from other “creative” domains such as music, literature, and film.

6. <http://www.hrc.govt.nz/funding-opportunities/researcher-initiated-proposal/explorer-grants>

7. <https://www.volkswagenstiftung.de/en/funding/our-funding-portfolio-at-a-glance/experiment>

References

Adam, D. 2019. “Science Funders Gamble on Grant Lotteries.” *Nature* 575: 574-75. <https://doi.org/10.1038/d41586-019-03572-7>

Avin, S. 2015. “Breaking the grant cycle: on the rational allocation of public resources to scientific research projects” (Doctoral thesis). <https://doi.org/10.17863/CAM.16172>

Avin, S. 2019a. “Centralized Funding and Epistemic Exploration.” *The British Journal for the Philosophy of Science* 70: 629-56. <https://doi.org/10.1093/bjps/axx059>

Avin, S. 2019b. “Mavericks and lotteries.” *Studies in History and Philosophy of Science* 76: 13-23. <https://doi.org/10.1016/j.shpsa.2018.11.006>

Azoulay, P., J. Zivin, and G. Manso. 2011. “Incentives and creativity: Evidence from the academic life sciences.” *The RAND Journal of Economics* 42: 527-54. <https://www.jstor.org/stable/23046811>

Barnett, A. G. 2016. “Funding by lottery: political problems and research opportunities.” *mBio* 7: e01369-16. <https://doi.org/10.1128/mBio.01369-16>

Barnett, A. G., P. Clarke, C. Vaquette, and N. Graves. 2017. “Using Democracy to Award Research Funding: An Observational Study.” *Research Integrity and Peer-Review* 2: 1-9. <https://doi.org/10.1186/s41073-017-0040-0>

Barnett, A. G., S. R. Glisson, and S. Gallo. 2018. “Do funding applications where peer reviewers disagree have higher citations? A cross-sectional study.” [version 2; peer review: 2 approved] *F1000Research* 7: 1030. <https://doi.org/10.12688/f1000research.15479.2>

Barnett, A. G., N. Graves, P. Clarke, and T. Blakely. 2015. “What is the impact of research funding on research productivity?” [Working paper] Available from: <https://eprints.qut.edu.au/83127/>

Barnett, A. G., D. L. Herbert, M. Campbell, N. Daly, J. A. Roberts, A. Mudge, and N. Graves. 2015. “Streamlined research funding using short proposals and accelerated peer review: an observational study.” *BMC Health Services Research* 15:55. <https://doi.org/10.1186/s12913-015-0721-7>

Barnett, A., D. Herbert, P. Clarke, and N. Graves. 2014. “The research lottery: the pressures on the Australian grant system.” *AQ: Australian Quarterly* 85: 4-9.

Bedessem, B. 2020. “Should we fund research randomly? An epistemological criticism of the lottery model as an alternative to peer review for the funding of science.” *Research Evaluation* 29: 150-7. <https://doi.org/10.1093/reseval/rvz034>

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. <https://doi.org/10.15252/embr.201949472>

Bol, T., M. de Vaan, and A. van de Rijt. 2018. “The Matthew effect in science funding.” *Proceedings of the National Academy of Sciences* 115: 4887-90. <https://doi.org/10.1073/pnas.1719557115>

Bollen, J., D. Crandall, D. Junk, Y. Ding, and K. Börner. 2014. “From Funding Agencies to Scientific Agency.” *EMBO Reports* 15: 131-3. <https://doi.org/10.1002/embr.201338068>

Boyle, C. 1998. “Organizations selecting people: how the process could be made fairer by the appropriate use of lotteries.” *Journal of the Royal Statistical Society: Series D (The Statistician)* 47: 291-321. <http://www.jstor.org/stable/2988669>

Brezis, E. S. 2007. “Focal randomisation: an optimal mechanism for the evaluation of R&D projects.” *Science and Public Policy* 34: 691-8. <https://doi.org/10.3152/030234207X265394>

Bromham, L., R. Dinnhage, and X. Hua. 2016. “Interdisciplinary research has consistently lower funding success.” *Nature* 534: 684-7. <https://doi.org/10.1038/nature18315>

Boudreau, K. J., E. C. Guinan, K. R. Lakhani, and C. Riedl. 2016. “Looking across and looking beyond the knowledge frontier: Intellectual distance, novelty, and resource allocation in science.” *Management Science* 62: 2765-83. <https://doi.org/10.1287/mnsc.2015.2285>

Chubin, D. E. 1994. “Grants peer review in theory and practice.” *Evaluation Review* 18: 20-30. <https://doi.org/10.1177/0193841X9401800103>

Daniels, R. J. 2015. “A generation at risk: young investigators and the future of the biomedical workforce.” *Proceedings of the National Academy of Sciences* 112: 313-8. <https://doi.org/10.1073/pnas.1418761112>

Dinov, I. D. 2020. “Flipping the grant application review process.” *Studies in Higher Education* 45: 1737-45. <https://doi.org/10.1080/03075079.2019.1628201>

Else, H. 2019. “Male Researchers' 'vague' Language More Likely to Win Grants.” *Nature News*. <https://doi.org/10.1038/d41586-019-01402-4>

Fang, H. 2011. “Peer review and over-competitive research funding fostering mainstream opinion to monopoly.” *Scientometrics* 87: 293–301. <https://doi.org/10.1007/s11192-010-0323-4>

Fang, F. C., A. Bowen, and A. Casadevall. 2016. “NIH peer review percentile scores are poorly predictive of grant productivity.” *eLife* 5: e13323. <https://doi.org/10.7554/eLife.13323>

Fang, F. C., and A. Casadevall. 2009. “NIH peer review reform--change we need, or lipstick on a pig?” *Infection and immunity* 77: 929-32. <https://doi.org/10.1128/IAI.01567-08>

Fang, F. C., and A. Casadevall. 2016a. “Research funding: the case for a modified lottery.” *mBio* 7: e00422-16. <https://doi.org/10.1128/mBio.00422-16>

Fang, F. C., and A. Casadevall. 2016b. “Grant funding: Playing the odds.” *Science* 352: 158. <https://doi.org/10.1126/science.352.6282.158-a>

Fleurence, R. L., L. P. Forsythe, M. Lauer, J. Rotter, J. P. A. Ioannidis, A. Beal, L. Frank, and J. V. Selby. 2014. “Engaging Patients and Stakeholders in Research Proposal Review: The Patient-Centered Outcomes Research Institute.” *Research and Reporting Methods* 161: 122-30. <https://doi.org/10.7326/M13-2412>

Fogelholm, M., S. Leppinen, A. Auvinen, J. Raitanen, A. Nuutinen, and K. Väänänen. 2012. “Panel Discussion Does Not Improve Reliability of Peer Review for Medical Research Grant Proposals.” *Journal of Clinical Epidemiology* 65: 47–52. <https://doi.org/10.1016/j.jclinepi.2011.05.001>

Gallo, S. A., J. H. Sullivan, and S. R. Glisson. 2016. “The Influence of Peer Reviewer Expertise on the Evaluation of Research Funding Applications.” *PLOS One* 11: e0165147. <https://doi.org/10.1371/journal.pone.0165147>

Gallo, S., L. Thompson, K. Schmaling, and S. Glisson. 2018. “Risk evaluation in peer review of grant applications.” *Environment Systems and Decisions* 38: 216-29. <https://doi.org/10.1007/s10669-018-9677-6>

Garner, H. R., L. J. McIver, and M. B. Waitzkin. 2013. “Same Work, Twice the Money?” *Nature* 493: 599–601. <https://doi.org/10.1038/493599a>

Gildenhuys, P. 2020. “Lotteries make science fairer.” *Journal of Responsible Innovation* 7 (sup2): S30-S43. <https://doi.org/10.1080/23299460.2020.1812485>

Gillies, D. 2014. “Selecting applications for funding: why random choice is better than peer review.” *RT. A Journal on Research Policy and Evaluation* 2. <https://doi.org/10.13130/2282-5398/3834>

Ginther, D. K., W. T. Schaffer, J. Schnell, B. Masimore, F. Liu, L. L. Haak, and R. Kington. 2011. “Race, Ethnicity, and NIH Research Awards.” *Science* 333: 1015-9. <https://doi.org/10.1126/science.1196783>

Graves, N., A. Barnett, and P. Clarke. 2011. “Funding grant proposals for scientific research: retrospective analysis of scores by members of grant review panel.” *British Medical Journal* 343: d4797. <https://doi.org/10.1136/bmj.d4797>

Greenberg, D. S. 1998. “Chance and grants.” *The Lancet* 351: 686. <https://doi.org/10.1016/S0140-6736(05)78485-3>

Gross, K., and C. T. Bergstrom. 2019. “Contest models highlight inherent inefficiencies of scientific funding competitions.” *PLOS Biology* 17: e3000065. <https://doi.org/10.1371/journal.pbio.3000065>

Guthrie, S., I. Ghiga, and S. Wooding. 2018. “What do we know about grant peer review in the health sciences? An updated review of the literature and six case studies.” *RAND report*, RAND Corporation. <https://www.rand.org/pubs/research_reports/RR1822.html>

Guthrie, S., D. Rodriguez Rincon, G. McInroy, B. Ioppolo, and S. Gunashekar. 2019. “Measuring bias, burden and conservatism in research funding processes” [version 1; peer review: 1 approved, 1 approved with reservations]. *F1000Research* 8: 851. <https://doi.org/10.12688/f1000research.19156.1>

Harnagel, A. 2019. “A mid-level approach to modeling scientific communities.” *Studies in History and Philosophy of Science Part A* 76: 49-59. <https://doi.org/10.1016/j.shpsa.2018.12.010>

Hayes, M., and J. Hardcastle. 2019. “Grant review in focus. Global state of peer review series.” *Publons report*. Downloaded from <https://publons.com/community/gspr/grant-review>

Herbert, D. L., A. G. Barnett, P. Clarke, and N. Graves. 2013. “On the time spent preparing grant proposals: an observational study of Australian researchers.” *BMJ Open*3**:** e002800. <https://doi.org/10.1136/bmjopen-2013-002800>

Holbrook, J. B., and R. Frodeman. 2011. “Peer review and the *ex ante* assessment of societal impacts.” *Research Evaluation* 20:239-46. <https://doi.org/10.3152/095820211X12941371876788>

Holliday, C., and M. Robotin. 2010. “The Delphi process: a solution for reviewing novel grant applications.” *International Journal of General Medicine* 3: 225-30. <https://doi.org/10.2147/IJGM.S11117>

Ioannidis, J. P. A. 2011. “More time for research: fund people not projects.” *Nature* 477: 529-31. <https://doi.org/10.1038/477529a>

Jang, D., S. Doh, G.-M. Kang, and D.-S. Han. 2017. “Impact of Alumni Connections on Peer Review Ratings and Selection Success Rate in National Research.” *Science, Technology, & Human Values* 42: 116-43. <https://doi.org/10.1177/0162243916665466>

Janosov, M., F. Battiston, and R. Sinatra. 2020. “Success and luck in creative careers.” *EPJ Data Science* 9: 9. <https://doi.org/10.1140/epjds/s13688-020-00227-w>

Johnson, V. E. 2008. “Statistical analysis of the National Institutes of Health peer review system.” *Proceedings of the National Academy of Sciences* 105: 11076-80. <https://doi.org/10.1073/pnas.0804538105>

Kaplan, D., N. Lacetera, and C. Kaplan. 2008. “Sample Size and Precision in NIH Peer Review.” *PLOS One* 3: e2761. <https://doi.org/10.1371/journal.pone.0002761>

Kolev, J., Y. Fuentes-Medel, and F. Murray. 2019. “Is blinded review enough? How gendered outcomes arise even under anonymous evaluation.” *National Bureau of Economic Research Working Paper* No. 25759. <https://www.nber.org/papers/w25759>

Li, D., and L. Agha. 2015. “Research funding. Big names or big ideas: do peer-review panels select the best science proposals?” *Science* 348: 434-8. <https://doi.org/10.1126/science.aaa0185>

Linton, J. D. 2016. “Improving the Peer review process: Capturing more information and enabling high-risk/high-return research.” *Research Policy* 45: 1936-8. <https://doi.org/10.1016/j.respol.2016.07.004>

Liu, M., V. Choy, P. Clarke, A. Barnett, T. Blakely, and L. Pomeroy. 2020. “The acceptability of using a lottery to allocate research funding: a survey of applicants.” *Research Integrity and Peer Review* 5: 3. <https://doi.org/10.1186/s41073-019-0089-z>

Luukkonen, T. 2012. “Conservatism and risk-taking in peer review: Emerging ERC practices.” *Research Evaluation* 21: 48-60. <https://doi.org/10.1093/reseval/rvs001>

Mallapaty, S. 2018. “Predicting Scientific Success.” *Nature* 561: S32. <https://doi.org/10.1038/d41586-018-06627-3>

Martin, B. 2000. “Research grants: problems and options”. *Australian Universities’ Review* 43: 17-22. <https://documents.uow.edu.au/~bmartin/pubs/00aur.html>

Mayo, N. E., J. Brophy, M. S. Goldberg, M. B. Klein, S. Miller, R. W. Platt, and J. Ritchie. 2006. “Peering at Peer Review Revealed High Degree of Chance Associated with Funding of Grant Applications.” *Journal of Clinical Epidemiology* 59: 842–48. <https://doi.org/10.1016/j.jclinepi.2005.12.007>

Mom, C., U. Sandström, and P. van den Besselaar. 2018. “Does cronyism affect grant application success? The role of organizational proximity.” *Proceedings of the 23rd International Conference on Science and Technology Indicators*, 12-14 September 2018, Leiden, The Netherlands. <https://openaccess.leidenuniv.nl/bitstream/handle/1887/65246/STI2018_paper_263.pdf>

Mongeon, P., C. Brodeur, C. Beaudry, and V. Larivière. 2016. “Concentration of research funding leads to decreasing marginal returns.” *Research Evaluation* 25: 396–404. <https://doi.org/10.1093/reseval/rvw007>

Morgan, B., L.-M. Yu, T. Solomon, and S. Ziebland. 2020. “Assessing health research grant applications: A retrospective comparative review of a one-stage versus a two-stage application assessment process.” *PLOS One* 15: e0230118. <https://doi.org/10.1371/journal.pone.0230118>

Nicholson, J. M., and J. P. A. Ioannidis. 2012. “Research grants: Conform and be funded.” *Nature* 492: 34–6. <https://doi.org/10.1038/492034a>

NIH. 2008. “Enhancing peer review: A self-study by the NIH in partnership with the scientific community to strengthen peer review in changing times.” Available from <http://enhancing-peer-review.nih.gov/meetings/EnhancingPeerReviewACD2-21-08.pdf>

Nosek, B. A., J. R. Spies, and M. Motyl. 2012. “Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth Over Publishability.” *Perspectives on Psychological Science* 7: 615–31. <https://doi.org/10.1177/1745691612459058>

Osterloh, M., and B. S. Frey. 2020. “How to avoid borrowed plumes in academia.” *Research Policy* 49: 103831. <https://doi.org/10.1016/j.respol.2019.103831>

Packalen, M., and J. Bhattacharya. 2020. “NIH funding and the pursuit of edge science.” *Proceedings of the National Academy of Sciences* 117: 12011-6. <https://doi.org/10.1073/pnas.1910160117>

Philipps, A. 2020. “Science rules! A qualitative study of scientists’ approaches to grant lottery.” *Research Evaluation.* <https://doi.org/10.1093/reseval/rvaa027>

Pier, E. L., M. Brauer, A. Filut, A. Kaatz, J. Raclaw, M. J. Nathan, C. E. Ford, and M. Carnes. 2018. “Low agreement among reviewers evaluating the same NIH grant applications.” *Proceedings of the National Academy of Sciences* 115: 2952-7. <https://doi.org/10.1073/pnas.1714379115>

Pluchino, A., A. E. Biondo, and A. Rapisarda. 2018. “Talent vs luck: The role of randomness in success and failure.” *Advances in Complex Systems* 21 (03n04): 1850014. <https://doi.org/10.1142/S0219525918500145>

Pluchino, A., G. Burgio, A. Rapisarda, A. E. Biondo, A. Pulvirenti, A. Ferro, and T. Giorgino. 2019. “Exploring the role of interdisciplinarity in physics: Success, talent and luck.” *PLOS One* 14: e0218793. <https://doi.org/10.1371/journal.pone.0218793>

Pohlhaus, J. R., H. Jiang, R. M. Wagner, W. T. Schaffer, and V. W. Pinn. 2011. “Sex differences in application, success, and funding rates for NIH extramural programs.” *Academic Medicine* 86: 759-67. <https://doi.org/10.1097/ACM.0b013e31821836ff>

Publons. 2019. “Global State of peer review”. Report available from <https://publons.com/community/gspr>

Rasey, J. S. 1999. “The Art of Grant Writing.” *Current Biology* 9: R387. <https://doi.org/10.1016/S0960-9822(99)80245-0>

Reinhart, M., and C. Schendzielorz. 2020. “The lottery in Babylon—On the role of chance in scientific success.” *Journal of Responsible Innovation* 7 (sup2): S25-9. <https://doi.org/10.1080/23299460.2020.1806429>

Rothman, K. J. 2016. “John Snow’s Grant Application.” *Epidemiology* 27: 311-3. <https://doi.org/10.1097/EDE.0000000000000453>

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*.* <https://doi.org/10.1177/0162243918822744>

Roumbanis, L. 2020. “Two dogmas of peer-reviewism.” *Journal of Responsible Innovation* 7 (sup2): S129-S133. <https://doi.org/10.1080/23299460.2020.1855806>

Sato, S., P. M. Gygax, J. Randall, and M. Schmid Mast. 2020. “The leaky pipeline in research grant peer review and funding decisions: challenges and future directions.” *Higher Education*. <https://doi.org/10.1007/s10734-020-00626-y>

Sattler, D. N., P. E. McKnight, L. Naney, and R. Mathis. 2015. “Grant Peer Review: Improving Inter-Rater Reliability with Training.” *PLOS One* 10: e0130450. <https://doi.org/10.1371/journal.pone.0130450>

Schroter, S., N. Black, S. Evans, J. Carpenter, F. Godlee, and R. Smith. 2004. “Effects of training on quality of peer review: Randomized control trial.” *British Medical Journal* 328: 673-5. <https://doi.org/10.1136/bmj.38023.700775.AE>

Serrano Velarde, K. 2018. “The Way We Ask for Money… The Emergence and Institutionalization of Grant Writing Practices in Academia.” *Minerva* 56: 85–107. <https://doi.org/10.1007/s11024-018-9346-4>

Shepherd, J., G. K. Frampton, K. Pickett, and J. C. Wyatt. 2018. “Peer review of health research funding proposals: A systematic map and systematic review of innovations for effectiveness and efficiency.” *PLOS One* 13: e0196914. <https://doi.org/10.1371/journal.pone.0196914>

Sloman, A. 2014. “How to Select Research Proposals Less Wastefully: Use a Sensibly Designed, Relatively Inexpensive, Dynamic, Weighted Lottery.” Accessed April 27, 2017. <http://www.cs.bham.ac.uk/research/projects/cogaff/misc/lottery.html>

Stone, P. 2009. “The logic of random selection.” *Political theory* 37: 375-97. <https://doi.org/10.1177%2F0090591709332329>

van Noorden, R., and G. Brumfiel. 2010. “Fixing a grant system in crisis.” *Nature* 464: 474-5. <https://doi.org/10.1038/464474a>

Wahls, W. P. 2018. “The NIH must reduce disparities in funding to maximize its return on investments from taxpayers.” *eLife* 7: e34965. <https://doi.org/10.7554/eLife.34965>

Wang, Y., B. F. Jones, and D. Wang. 2019. “Early-career setback and future career impact.” *Nature Communications* 10: 4331. <https://doi.org/10.1038/s41467-019-12189-3>