

CONTRIBUTED PAPER

Peer Review, Innovation, and Predicting the Future of Science: The Scope of Lotteries in Science Funding Policy

Jamie Shaw

Leibniz University Hannover, Hannover, Germany
Email: jshaw222@uwo.ca

(Received 19 January 2023; revised 25 January 2023; accepted 31 January 2023; first published online 17 February 2023)

Abstract

Recent science funding policy scholars and practitioners have advocated for the use of lotteries, or elements of random chance, as supplementations of traditional peer review for evaluating grant applications. One of the primary motivations for lotteries is their purported openness to innovative research. The purpose of this article is to argue that current proponents of funding science by lottery overestimate the viability of peer review and thus unduly restrict the scope of lotteries in science funding practice. I further show how this analysis suggests a different way of introducing lotteries into science funding policy.

1. Introduction

Science funding policy is going through a state of crisis. The paradigm of allocating grants through peer review, which has dominated since the end of World War II, has encountered several criticisms that it seems increasingly unable to address. As a result, many are losing confidence in peer review and, thereby, proposing novel methods for allocating funds. One criticism is that traditional peer review discriminates against marginalized groups, thus perpetuating extant funding gaps (Lee 2015; Guthrie et al. 2018). Another criticism is that peer review is incredibly expensive as the opportunity costs of writing and reviewing grants is enormous and unsustainable as science labor markets grow (Aczel et al. 2021). Finally, as we focus on in this article, many argue that peer review fosters conservatism by disincentivizing innovation (Brezis 2007; Stanford 2019). As a result, novel funding allocation mechanisms have been proposed and implemented that supplement funding via peer review.

One of the most hotly debated funding allocation methods involves the introduction of *lotteries* into funding decisions (Boyle 1998; Gillies 2014; Fang and Casadevall 2016; Avin 2019; Osterloh and Frey 2020). Lotteries have been increasingly common in practice and have been implemented in New Zealand, Canada, Switzerland, Germany, the United

Kingdom, and Austria. Despite the title, lotteries are not a full-blown alternative to traditional peer review; peer review is still used in funding-by-lottery (Shaw 2023). After initial screening of grants to ensure their relevance and institutional viability, the most common proposal and practice is of *partial lotteries* where traditional peer review winnows down postscreened applicant pools and lotteries are used for the remainder. The Volkswagen Foundation, for example, used peer review to select 15 to 20 grants (out of 300–50 grants) to be funded and entered the rest into a lottery (Roumbanis 2019, 1012–13). Moreover, lotteries are exclusively used in calls for *innovative* research. For example, the Health Research Council of New Zealand calls for “ideas considered to be transformative, innovative, exploratory or unconventional, and have potential for major impact” (HRC 2014). This situation is justified through a popular, if not universally held, interpretation of the current evidence on the reliability of peer review.

In this article, I argue for a novel way of understanding the policy implications of current studies of peer review. My interpretation suggests that *nonpartial* lotteries should be practiced in distinct epistemic contexts. Specifically, I suggest that we move away from partial lotteries in the context of I call “luxury science” rather than innovative science (Shaw 2022).

2. Peer review and innovation

It is widely acknowledged that innovation, or deviating from conventionally accepted scientific norms, constitutes an essential part of scientific inquiry (Carrier 2021). Innovation is often crucial for opening up novel avenues for research and challenging widely held beliefs. While the value of innovation is widely accepted, many argue that traditional science funding policy discourages innovative research, thus leading to a more conservative landscape of scientific pursuits (Stanford 2019). More acutely, it has been argued that peer review is *unreliable* for judging the merits of innovative research proposals thus making it more difficult for them to accrue funds.

This argument stems from *interrater reliability* studies (IRR studies) on peer review. The goal of IRR studies is to discern the rate at which reviewers will agree when judging the same object. In the current context, IRR studies seek to show the rate at which reviewers will agree that a project is worthy of funds (in binary cases), or reviewers will agree about a numerical scoring of a grant. To illustrate, consider a simple case. Imagine we have two reviewers who review 19 grants. 10 of these grants are accepted by both reviewers, 2 are rejected by both reviewers, 4 are rejected by reviewer #1 but accepted by reviewer #2, and 3 are rejected by reviewer #2 but accepted by reviewer #1. The first step is to calculate the observed agreement (OA), which is the sum of the grants that both reviewers agree should be accepted or rejected (in this case, $OA = 10 + 2 = 12/19 = 0.63$). Because most grant evaluations are a numerical score rather than a binary decision, OA is usually measured through standard deviation analyses of grant scores (omitted here for the sake of simplicity). Next, we calculate the agreement by chance (AC), defined as the possibility that two reviewers will agree regardless of the properties of what is being reviewed. This is calculated by the following formula:

$$AC = \frac{1}{N^2} \sum_K n_{ki}$$

Where N is the total number of reviews, and k is chance that a reviewer, i , will place a review in a particular category. In this case, the possibility of an agreement by chance is 0.58. Finally, we calculate the Kappa coefficient (κ) through the formula:

$$\kappa = \frac{(OA - AC)}{1 - AC}$$

In this case, $\kappa = 0.12$. This value stands for the reliability of reviewers #1 and #2 agreeing about whether an individual proposal should be funded, though it does not tell us whether the reasoning of the reviewers is the same or even consistent.

IRR studies can be made more complex in several ways to uncover what causes κ -values to fluctuate. One way is to change the object that is being judged (e.g., manuscripts for publication, CVs, grant proposals) or constructing more fine-grained categories (e.g., grant proposals in the biological sciences, grant proposals with smaller amounts of text, interdisciplinary grant proposals). Similarly, one can change the reviewers to be, for example, reviewers from a particular discipline or social group (e.g., female reviewers, reviewers from top-tier universities, reviewers from biomechanics). While some IRR studies have been conducted experimentally, the majority of them extract data from peer review practices.

Hundreds of IRR studies have been conducted providing a bewildering range of findings (Guthrie et al. 2018). The relevance of these studies for science funding policy follows what can be called the *conventional interpretation* of IRR studies:

Conventional Interpretation: Peer review should be employed in cases in which κ is sufficiently high and not used when κ is sufficiently low.

The conventional interpretation has never, to my knowledge, been stated outright. However, it is essential for filling in the gap between discovered κ -values and practical recommendations. Of course, the thresholds of what counts as “sufficiently” high or low are vague and vary depending on the context. The conventional interpretation is an intuitive stance to take. If one thinks that the judgment of well-chosen peers is the best gatekeeper there is, then it seems sensible to think that insofar as peer review is *reliable*, then it is the best method of evaluating grant proposals. The conventional interpretation, when applied to lotteries, suggests that the results from IRR studies determine the limits of peer review and the appropriate scope of lotteries. Funding policy scholars are hesitant to open the floodgates *tout court* because this will likely also open doors to research that genuinely lacks rigor. The hope, then, is to be more selective of the roles peer review plays when reviewing innovative research. This will allow to allow funding bodies to have the best of both worlds: They can provide more opportunities for innovative projects while not wasting funds on fruitless research. If peer review can reliably determine what research falls into which category, then funding bodies can achieve both goals. The pertinent question, on this approach, becomes: *When is peer review reliable?*

One extremely well-confirmed finding is that peer review is unreliable when research proposals are *innovative*—that is, research that deviates from accepted scientific norms in at least one important respect (Guthrie et al. 2018). These studies have been conducted on large sample sizes in various epistemic and institutional contexts around the world. There are several contributing factors that explain this.

The first is that novelty is typically lowly valued when compared to other virtues. Regression analyses of 32,546 applications to the National Institute of Health showed that the contribution of “novelty” toward grant scores was only 1.4 points, in contrast to 6.7 points for “significance” (Lee 2015, 1275). Therefore, innovative aspects of proposals do little for increasing a grants chance of success. Second, the innovative components of research proposals typically lead to a perception that the research lacks rigor. This is understandable because innovation often involves deviating from conventional norms (Luukkonen 2012). Moreover, as reviewers are typically pressured to *reject* grant proposals rather than *accept*, any semblance of a problem or obstacle provides reasons to rank innovative research lower in the pool of applicants. Because innovative research, by its very nature, depends on less assured grounds, it has an especially difficult task in convincing reviewers that the promises will be fulfilled.

This, however, is not the end of the story. Further studies suggest that κ -values fluctuate depending on the *level of quality* of the grant application, though the findings are mixed. Some claim that κ -values are *high* when determining which grants are of the highest (relative) quality (Li and Agha 2015). Others claim that κ -values are high only when determining which grants are of the lowest (relative) quality (Gillies 2014). Still others claim κ -values are only low for the middle-level proposals (Osterloh and Frey 2020).¹ The final view is the most popular amongst scientists as indicated in a recent poll by Phillips (2021, 9) which showed that 82% of surveyed scientists believed that peer review was only unreliable for middle-ground grants. Regardless, whichever result is accepted justifies the ways in which peer review winnows down the applicant pool—by eliminating the good, the bad, or the good and the bad and using lotteries for the rest.

3. Evaluating the conventional interpretation

The conventional interpretation does not conflict with the aspirations of traditional peer review; that is, to provide rational justifications for funding allocation decisions. Rather, it concedes that these aspirations cannot be realized in special contexts. The conviction that peer review provides the best possible gatekeeper for funding bodies remains intact, but IRR studies show its lamentable limitations. It is my contention that the conventional interpretation does not tell a full story and, as a result, we have not broken with tradition as much as we should.

IRR studies are most commonly conducted on peer review for *publication* and are often cited to contain the same implications for funding. For one example, Ismail et al.’s (2009) meta-study cites several papers providing κ -values for *journal* peer review in support of claims concerning the use of peer review in funding policy. Of course, reviewers for journals are looking for different virtues than funding bodies, but it seems to be assumed that reviewers are equally reliable or unreliable for identifying merit in different contexts. When reviewing a publication, peer reviewers might assess whether the statistical analysis is sound, if there is unconsidered or misinterpreted evidence, or a mistake in a proof. That is, publication review often

¹ It is unclear what these studies show in the context of evaluating innovative research in particular. It is *assumed* that IRR scores will be more or less equal when reviewing, say, poor proposals in innovative and conservative contexts. For the sake of this article, I will grant this premise.

focuses on whether the manuscript forwards knowledge that might be *accepted*. When reviewing a grant proposal, the primary virtue that is being assessed is its *pursuitworthiness*—that is, the possibility that the research proposed will eventually aid in the development of scientific knowledge. Philosophers have already done a great deal to make it clear that these are distinct epistemic attitudes one may take (Whitt 1990; Patton 2012; Barseghyan 2015). The question that confronts us now is what implications this has for the reliability of peer review.

Like judgments of acceptance, pursuitworthiness assessments require information from a “cognitive horizon” or a collection of accepted theories, norms, methods, and so on (Šešelja and Straßer 2014). For any given review, a judgment of pursuitworthiness assumes that the parts of the cognitive horizon that are employed in that judgment will remain significantly unchanged. This is a distinct assumption judgments of peer review in science funding policy must make. If they were to change (in a way relevant to its use for evaluating pursuitworthiness), then the grounds of the pursuitworthiness assessment would evaporate. Consider Derek Turner’s claim that we would never know the color of the dinosaurs. Ground-breaking research after Turner’s initial prediction on the fossilization process allowing for the preservation of information on dinosaur melanosomes transformed the cognitive horizon Turner’s prediction was grounded in. This, ultimately, was the reason why his prediction turned out false (Turner 2016). If we assume that any part of a cognitive horizon, no matter how entrenched, may change, then our pursuitworthiness assessments can always become undermined by future epistemic changes. Moreover, even in cases in which the cognitive horizon appears to be at odds with the research project and we had some assurance that it would not change, the reviewer must believe that the project will not develop to *eventually* become consistent with the cognitive horizon. Because of this, reviewers must *predict the future of science*—they must be confident that those parts of the cognitive horizon that ground their pursuitworthiness assessments will not change—to judge a research project as genuinely pursuitworthy or not.

Philosophers have typically thought that predicting the future of science is unpredictable. Wesley Salmon, for example, implies that we can never know the absolute confirmation of a hypothesis because we cannot predict the future of science (Salmon 1990). Similar claims were put forth by Vannevar Bush, Feyerabend, Polanyi, Popper, and Toulmin. A more refined analysis was offered by Derek Turner (2016, 54–56), who has taken a closer look at the assumptions behind predicting future science and has come to more nuanced categorizations of different kinds of predictions about future science. Turner’s appraisal of these kinds of predictions is largely grounded in intuitive assessments of which ones are more likely to be reliable. But we do not need to rest content with this analysis. Another kind of study on peer review, called *predictive validity studies* (PVSs), has been conducted on peer reviewer predictions on future performance. PVSs have the advantage of assessing scientists’ judgments of pursuitworthiness while incorporating a wide variety of “external” factors affect the growth of science (e.g., legal restrictions, practical limitations) rather than just scientific considerations (as in Turner’s analysis). While the current PVSs are not fine-grained enough to test the various claims Turner puts forth, there remain lessons to be learned. Before this, it is worth describing how PVSs work.

PVSs aim to retrospectively correlate citation scores with reviewer scores of grant proposals. Because there is a selection bias for funded research, as citation scores will be higher for projects that received funding, the correlations are subject to random variation analysis (see van den Besselaar and Sandström 2015, 830ff for the details). Given that citation practices vary wildly across disciplines (e.g., cognitive scientists cite frequently while mathematical physicists are stingier), the citation scores are field normalized (Bornmann 2011). While it is unclear to what extent citations provide a proxy for “success,” they have some advantages and limitations. The obvious, practical advantage is that they are extremely tractable. Other methods of studying success may require large-scale questionnaires that are not only onerous but have their own limitations (Coughland et al. 2009). Another advantage is they allow for a liberal sense of “success.” “Success” does not necessarily mean “true” (or “empirically adequate,” “robust,” or some cognate concept)—it means simply that it has garnered attention. Even critical attention can be a marker of success because it shows the ideas were stimulatory. A limitation is that bibliometric evaluations cannot directly account for other successes such as uptake in policy, popular culture, legal proceedings, curriculum, or concrete practices (Haustein and Larivière 2015). Moreover, some citations are not indicators of “success” but nepotism, reviewer pressure, or gathering “allies” (see Latour 1987). Finally, many publications contain *potentially* valuable findings but are not highly cited due to low visibility, which is often correlated with how researchers are situated in scientific communities (Marx and Bornmann 2015).

Despite these limitations, PVSs can provide a useful indicator of scientists’ predictive powers while admitting that more work needs to be done. A few dozen PVSs have been conducted that show little variability of the predictive abilities of scientists across disciplines or cultures. More acutely, though, a pattern has been discovered that shows consistent, replicated results in which the predictive validity of scientist’s judgments *5 years or under* is fairly high and *after 5 years* becomes extremely low (Shaw 2022, 107). These trends hold (roughly) equally for predictions about what *will* succeed and what *won’t* succeed. In other words, scientists of various stripes are typically good at predicting future success and failure under 5 years and extremely poor when evaluating the research’s prospects after the 5-year mark. While future research will surely change these conclusions, they will not undermine the argument I will now make for the methodological primacy of PVSs.

4. Implications for science funding policy

Why should we value a local reliable consensus for science funding policy if those consensus are unreliable at identifying research that will be impactful? If it turns out that peer reviewers will reliably give grants similar scores, *but* those scores do not reflect future success, then it seems as if there is no reason to give practical weight to IRR studies. The purpose of funding a grant is to support the science that it will eventually produce. If we cannot tell whether the project, no matter how high or lowly it is perceived by reviewers, will be successful it seems as if there is no reason to subject the research to peer review. Given this, even if IRR studies show that κ -values are consistently high in particular contexts, those scores are irrelevant if they do not reliably track future success or failure.

This does not show that IRR studies are entirely irrelevant for peer review in funding policy. A blanket condemnation of their value is too rash of a conclusion to draw. As mentioned, PVS studies show that the predictive validity scores are high insofar as the retrospective period is 5 years or under. As a result, in cases in which we aim to predict success in short term, IRR studies become essential as a complementary tool for evaluating peer review. It would be necessary, but not sufficient, for reviewer scores to be predictively valid. These reviewer scores must also be *reliable*, and this should be discovered through IRR studies. Therefore, according to the current evidence from PVSs, the conventional interpretation of IRR studies for discerning the appropriate scope of peer review within lotteries remains sound in short term cases (≤ 5 years). Simply put: We do not have the appropriate evidentiary grounds for using peer review even in its more minimal function in partial lotteries for assessing long-term research potential.

One may accept this conclusion and infer that partial lotteries should be employed in calls for grants that are ≤ 5 years—most scientific grants—but not for longitudinal grants. Turner, for example, argues that the “claim that we probably won’t get the evidence we need anytime soon would be enough to motivate a decision to work on something else” (Turner 2016, 66). This inference, straightforward as it seems, does not hold. There is a larger question lurking in the background: Under what conditions should we appraise research proposals for their short-term success? The inference just mentioned makes sense if we hold a piecemeal approach where scientific progress is made in small, incremental steps. If we held this view, what is going to be successful in the long term may be unpredictable, but each step will be predictable *in situ*. This conception of progress, though, runs afoul of the conditions that are necessary for innovative research to thrive. Innovative research rarely provides fruits in the short term. When one deviates from accepted knowledge, it often raises a host of new questions and challenges that need to be addressed. While the time frame of innovative discoveries is not well studied, it seems intuitive that innovative research will frequently *fail* according to short-term metrics. If we were to engage in a piecemeal approach, we would not leave adequate room for innovative projects that were one of the main motivations behind lotteries in the first place.

Rather, the reason that we should engage in short-term appraisals is because we *need* particular results by a deadline. In other words, the science must be *urgent*. Research may be urgent for all sorts of reasons—moral (e.g., uptake in policy contexts, designing new technologies, or medical treatments that should happen sooner rather than later), practical (e.g., scientists’ need to publish their findings for promotion), and so on. While I do not have the space to elaborate on the notion of urgency here, or how urgency is to be determined (see Shaw 2022), we can rest content with the following conclusion: IRR studies should be conventionally interpreted in cases of urgent science insofar as the grant proposals purporting to address the urgent need fall within the bounds of predictively validity.

Not all science is urgent science. In many cases, scientists are relatively unconstrained in terms of deliverables. Given the value that we place on innovative research, we should allow for space for scientists to not be held directly accountable in terms of research outputs. As has been stated by proponents of the “Slow Science” movement:

We do need time to think. We do need time to digest. We do need time to misunderstand each other, especially when fostering lost dialogue between humanities and natural sciences. We cannot continuously tell you what our science means; what it will be good for; because we simply don't know yet. Science needs time. (Slow Science Academy 2010)

We can call research conducted in this context *luxury science* in which scientists have the luxury of pursuing research without firm deadlines. This freedom allows for scientists to pursue research that does not have any obvious, forthcoming practical value or even research that may not bear fruits for decades. As anyone familiar with the history of science will know, science needs such luxuries to grow.

This has important implications for the practice of funding science by lottery. As mentioned, lotteries are currently practiced in the context of *innovative* research and are *partial* lotteries. The foregoing suggests a different practice with a distinct scope of funding by lottery and a transformation of the kinds of lotteries to be implemented. Specifically, it seems sensible to employ *nonpartial* lotteries in the context of *luxury science*. Within luxury science, our current knowledge of predicting future science suggests that local consensus are unreliable so peer review cannot be used even if just to winnow applicant pools. Within urgent science, we may employ partial lotteries though it's possible that other funding allocation methods may be also necessary here.

While luxury science and innovative research overlap, they are not entirely coextensive. Scientists may engage in much more conventional, run-of-the-mill research during luxury science. For example, scientists who are trying to find evidence for the existence of proton decay use fairly standard methods and are not “innovative” in any obvious sense. This constitutes perfectly respectable research within luxury science. Moreover, some innovative research is needed urgently, and we might be willing to take the risks that the research will not pay off when we want it to pay off. As such, the category of luxury science picks out distinct research proposals from innovative science suggesting a different epistemic context in which lotteries should be used.

This further suggests that we cannot use peer review, even in the restricted sense allotted to it in partial lotteries, for luxury research as reviewer judgments will be no better than random chance. Because peer review leads to a more conservative research landscape, lotteries seem preferable, *ceteris paribus*. This points us toward, rather, utilizing nonpartial lotteries that use little peer review for research evaluation beyond initial screening processes. While this suggestion raises its own questions and concerns, the foregoing at least pushes us strongly in this direction and suggest different calls for experimentation in the use of lotteries (e.g., Horbach et al. 2022).

5. Concluding remarks

This article's suggestions raise several questions about PVSs and nonpartial lotteries. For the former, given their increased importance, there is greater reason to conduct more fine-grained studies on different kinds of predictions scientists make when evaluating grants, and more research needs to be done to better understand the strengths and limitations of bibliometric analyses and alternative means of

measuring success. For the latter, opening up more room for lotteries will certainly reinvigorate worries about increased funding of pointless, fundamentally flawed, or ethically problematic research. Additionally, evaluating research promise is one of many possible functions that peer review could perform and so there is a need for more fine-grained research on the different ways peer review might be utilized. Much more research needs to be done to better understand how nonpartial lotteries may work before we can confidently assert that they deserve to be practiced in luxury science. Regardless of how this research turns out, I hope to have taken some initial steps that may lead to drastic changes in science funding policy practice.

Acknowledgments. Iterations of this paper were presented at the Leibniz Universität Colloquium series and the Philosophy of Science Association meeting in Pittsburgh (2022). I appreciate the feedback from both audiences.

References

- Aczel, Balazs, Barnabas Szaszi, and Alex O. Holcombe. 2021. "A Billion-Dollar Donation: Estimating the Cost of Researchers' Time Spent on Peer Review." *Research Integrity and Peer Review* 6 (1):1–8.
- Avin, Shahar. 2019. "Mavericks and Lotteries." *Studies in History and Philosophy of Science Part A* 76:13–23.
- Barseghyan, Hakob. 2015. *The Laws of Scientific Change*. New York: Springer.
- Bornmann, Lutz. 2011. "Scientific Peer Review." *Annual Review of Information Science and Technology* 45 (1): 197–245.
- Boyle, Conall. 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 (2): 291–321.
- Brezis, Elise. 2007. "Focal Randomisation: An Optimal Mechanism for the Evaluation of R&D Projects." *Science and Public Policy* 34 (10):691–98.
- Carrier, Martin. 2021. "How to Conceive of Science for the Benefit of Society: Prospects of Responsible Research and Innovation." *Synthese* 198 (19):4749–68.
- Coughlan, Michael, Patricia Cronin, and Frances Ryan. 2009. "Survey Research: Process and Limitations." *International Journal of Therapy and Rehabilitation* 16 (1):9–15.
- Fang, Ferric, and Arturo Casadevall. 2016. "Research Funding: The Case for a Modified Lottery." *mBio* 7 (2):1–7.
- Gillies, Donald. 2014. "Selecting Applications for Funding: Why Random Choice Is Better Than Peer Review." *RT. A Journal on Research Policy and Evaluation* 2 (1):1–14.
- Guthrie, Susan, Ioana Ghiga, and Steven Wooding. 2018. "What Do We Know about Grant Peer Review in the Health Sciences?" *F1000Research*, 6.
- Haustein, Stefanie, and Vincent Larivière. 2015. "The Use of Bibliometrics for Assessing Research: Possibilities, Limitations and Adverse Effects." In *Incentives and Performance*, edited by Isabell Welpe, Jutta Wollersheim, Stefanie Ringelhan, and Margit Osterloh, 121–39. Cham, Switzerland: Springer.
- Horbach, Serge, Joeri K. Tjink, and Lex M. Bouter. 2022. "Partial Lottery Can Make Grant Allocation More Fair, More Efficient and More Diverse." *Science and Public Policy* 49 (4): 1–3.
- HRC. 2014. "HRC Explorer Grant Application Guidelines (EX215)." Technical Report, RAND Europe.
- Ismail, Sharif, Alice Farrands, and Steven Wooding. 2009. *Evaluating Grant Peer Review in the Health Sciences—A Review of the Literature*. Santa Monica, CA: RAND Corporation.
- Latour, Bruno. 1987. *Science in Action*. Cambridge, MA: Harvard University Press.
- Lee, Carole. 2015. "Commensuration Bias in Peer Review." *Philosophy of Science* 82 (5):1272–83.
- Li, Danielle, and Leila Agha. 2015. "Big Names or Big Ideas: Do Peer-Review Panels Select the Best Science Proposals?" *Science* 348 (6233):434–38.
- Luukkonen, Terttu. 2012. "Conservatism and Risk-Taking in Peer Review: Emerging ERC Practices." *Research Evaluation* 21 (1):48–60.
- Marx, Werner, and Lutz Bornmann. 2015. "On the Causes of Subject-Specific Citation Rates in Web of Science." *Scientometrics* 102 (2):1823–27.

- Osterloh, Margit, and Bruno S. Frey. 2020. "How to Avoid Borrowed Plumes in Academia." *Research Policy* 49 (1):103831.
- Patton, Lydia. 2012. "Experiment and Theory Building." *Synthese* 184 (3):235–46.
- Philipps, Axel. 2021. "Science Rules! A Qualitative Study of Scientists' Approaches to Grant Lottery." *Research Evaluation* 30:102–11.
- Roumbanis, Lambros. 2019. "Peer Review or Lottery? A Critical Analysis of Two Different Forms of Decision-Making Mechanisms for Allocation of Research Grants." *Science, Technology, and Human Values* 44 (6):994–1019.
- Salmon, Wesley. 1990. "The Appraisal of Theories: Kuhn Meets Bayes." *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 2:325–32.
- Šešelja, Dunja, and Christian Straßer. 2014. "Epistemic Justification in the Context of Pursuit: A Coherentist Approach." *Synthese* 191 (13):3111–41.
- Shaw, Jamie. 2023. "Peer Review in Funding-by-Lottery: A Systematic Overview and Expansion." *Research Evaluation* 32 (1):86–100.
- Shaw, Jamie. 2022. "On the Very Idea of Pursuitworthiness." *Studies in the History and Philosophy of Science* 19:103–12.
- Slow Science Academy. 2010. "Slow Science Manifesto." *Slow-Science.org*. Accessed March 22, 2022, <http://slow-science.org/f>
- Stanford, P. Kyle. 2019. "Unconceived Alternatives and Conservatism in Science: The Impact of Professionalization, Peer-Review, and Big Science." *Synthese* 196 (10):3915–32.
- Turner, Derek. 2016. "A Second Look at the Colors of the Dinosaurs." *Studies in History and Philosophy of Science Part A* 55:60–68.
- van den Besselaar, Peter, and Ulf Sandström. 2015. "Early Career Grants, Performance, and Careers: A Study on Predictive Validity of Grant Decisions." *Journal of Informetrics* 9 (4):826–38.
- Whitt, Laurie Anne. 1990. "Theory Pursuit: Between Discovery and Acceptance." *Proceedings of the Biennial Meeting of the PSA* 1 (1):467–83.

Cite this article: Shaw, Jamie. 2023. "Peer Review, Innovation, and Predicting the Future of Science: The Scope of Lotteries in Science Funding Policy." *Philosophy of Science* 90 (5):1297–1306. <https://doi.org/10.1017/psa.2023.35>