A Lottery for the Republic of Science: Chance, Merit, and Fairness in the Process of Awarding Research Grants

Walter D. Valdivia

MERCATUS WORKING PAPER

MERCATUS CENTER
George Mason University
3434 Washington Blvd., 4th Floor, Arlington, Virginia 22201
www.mercatus.org

Walter D. Valdivia. "A Lottery for the Republic of Science: Chance, Merit, and Fairness in the Process of Awarding Research Grants." Mercatus Working Paper, Mercatus Center at George Mason University, Arlington, VA, March 2021.

Abstract

Federal agencies have a duty to seek the maximum social return from their investments in research (efficiency) and to confer research awards solely on the basis of merit (fairness). Consequently, the process of awarding grants relies on peer review to evaluate research proposals for merit. Peer review, however, also injects chance and bias into the process. Agencies could pass by good projects because of reviewer disagreement. Thus, any good project could use a little luck in being assigned to concurring reviewers. Formally internalizing chance may thus improve the system. I build on the Fang-Casadevall modified lottery proposal that consists of first screening projects for quality using peer review and then awarding funding via lottery to projects from that pool of preselected finalists. I argue that a modified lottery will encourage, in a few cases, a break from risk-averse projects, thus promoting the advancement of science. Researchers will also benefit because all preselected finalists will be recognized as equals in their profession, not only the lucky awardees. Agencies' cost savings from streamlining the process could be plugged back into research. Furthermore, I suggest tweaks to the modified lottery to better align it with agencies' missions and to discourage researchers from shirking in the performance of their projects.

JEL code: O32

Keywords: federal research, research grants, science awards, bias in science, chance in science

Author Affiliation and Contact Information

Walter D. Valdivia
Senior Policy Editor
Mercatus Center at George Mason University
wvaldivi@gmu.edu

This paper can be accessed at https://www.mercatus.org/publications/lottery-republic-science.

A Lottery for the Republic of Science

Walter D. Valdivia

1. Introduction

In principle, the process of awarding research grants is organized to reward merit. In practice, factors other than merit play a significant role. The prestige of the research team, its prior history of awards, its institutional affiliation, sociodemographic factors, and sheer luck skew award decisions. Measures to reduce the weight of these other factors may be effective to a certain degree, but chance and bias seem inherent in the peer review system. To address this policy paradox, I propose to formally introduce chance into the grant selection process. This counterintuitive proposal, I argue, could nevertheless enhance the role of merit, reduce the role of systemic bias, increase the productivity of research, and improve the process's overall fairness.

I begin in section 2 by surveying studies on reviewer disagreement during award selection. These studies examine the selection process used by federal research agencies and conclude that disagreement is pervasive to such a degree that a measure of chance is inescapable; being assigned to concurring reviewers is, in many instances, a lucky draw from the hat. Section 3 describes the inherent contradictions of a review process in which reliance on experts is a source of both strength and weakness. For all of these reasons, in section 4 I build upon a reform proposed by Fang and Casadevall (2016) that would formalize the role of chance in the grant review process. In a modified lottery, peer review is used as a filter of quality, but once a pool of high-quality proposals has been selected, grants are awarded by a lottery.

In its basic format, this modified lottery—currently used by Health Research Council of New Zealand and the Swiss National Science Foundation—improves the fairness of the award selection system. I discuss variants of the basic format that would address inherent contradictions and suggest extensions that would give federal agencies greater latitude to align their grant review processes with their missions. Section 5 deals with the foremost collateral problem of this proposal: Would a lottery discourage awardees from working hard at their projects? I argue that the lottery is at least as good as the current process at driving productivity and I suggest further tweaks to the basic format that would introduce incentives for high performance. My concluding remarks summarize the anticipated advantages and risks of the proposed reform.

2. The Matthew Effect and Factors Other Than Merit

The peer review system is instrumental to science—including to publication in scientific journals and the selection of research awards—because it is widely accepted that only experts in the relevant field of inquiry can appreciate the promise of a research project and estimate the likelihood that it will deliver on that promise. In effect, all federal agencies that fund research rely on experts' judgment to evaluate the merit of research proposals. But the peer review system is not error-free and, even when ably managed, it may be swayed by factors other than merit. For instance, reviewers may overrate proposals by famous applicants or underrate proposals as a result of unconscious bias.

Fame, in the world of science, has a compounding effect in the accrual of prestige, rewards, and privileges. Robert Merton (1968) called it the Matthew effect, in reference to the Gospel of Matthew: "For whoever has will be given more, and they will have an abundance. Whoever does not have, even what they have will be taken from them" (25:29, New International Version). This effect was carefully documented by Harriet Zuckerman (1977) in interviews with Nobel laureates, all of whom were self-aware of the inordinate amount of credit they received for their work merely because of their fame, compared with the credit their less-well-known peers received for work of similar quality. Likewise, the work of famous scientists is far more widely read, remembered, and cited than that of less famous scientists.

Soon after the Matthew effect was described, the overseers of research agencies in Congress began to voice their preoccupation with the idea that science might be operating as an old boys' club. If that were the case, talent left out of the club would also be left high and dry. The tide of public support would carry only the famous toward Vannevar Bush's (1945) endless frontier. The concerns of legislators were not merely about inclusion, but about whether a system skewed in this manner could ever yield the maximal public benefit from federal investments in research.

In response to these concerns, the National Science Foundation (NSF) commissioned a study with the National Academy of Sciences to examine the foundation's process of research grant selection for the existence of bias. The study had two phases (Cole et al. 1978, 1981). It is noteworthy that in the first phase, the study found no systemic bias. What stirred significant controversy—and nearly two years of internal review at the academy—was the central conclusion of the second phase: that the system allowed chance to be as much a factor in grant selection as merit. The study put 150 research proposals already funded by NSF through the same review process in a controlled environment at the academy. The study found that the reviewers convened by the academy would have made different funding choices in at least a quarter of the cases reexamined. There are legitimate reasons for disagreement among experts about the sort of project that merits an award, but the fact that disagreement is so prevalent suggests that merit is not always a quality of consensus and that chance is thus not a negligible factor (see also Cole and Simon 1981).

A more recent study also found significant reviewer disagreement in the National Institutes of Health (NIH) grant selection process (Pier et al. 2018). At NIH, each funding unit (institute, center, or program) convenes a panel of field experts and each proposal is assigned to two to five reviewers in the panel, who give the proposal a preliminary score. On the basis of that score, the top half of proposals is selected for full-panel discussion. After long deliberations, the full panel assigns a priority score to each proposal (averaging all panelists' scores). This priority score is used by funding units as their primary guidance when allocating funds. Pier's team examined both the quantitative and the qualitative assessments of reviewers, and not only found little agreement in preliminary scoring but also found that panel discussion deepened reviewers' differences (Pier at al. 2017). The full-panel discussion, designed to drive consensus, in many instances promotes dissensus. The Pier et al. (2017) study also revealed that qualitative judgments expressed as weaknesses or strengths do not consistently translate into low or high scores, respectively. These findings, as those about the NSF grant selection process, strongly suggest that chance plays a significant role in award selection.

During the years between the studies of the NSF and NIH processes, a growing body of work has investigated factors other than merit and chance that influence research award decisions. There is now strong evidence suggesting that systemic bias in the grant making process—particularly with respect to race and ethnicity (Ginther et al. 2011) and gender (Ley and Hamilton 2008)—is prevalent (see also Kaatz et al. 2015; Kotchen et al. 2006).

The foregoing evidence reveals a wicked problem in the peer review system: instrumental devices of its design are at cross-purposes with its main goal. The best tool we have to appraise the merit of research is shot through with opportunities for chance and bias to influence decisions. Addressing implicit and overt bias in science is both an urgent and a daunting task. Dealing with inherent chance is by comparison simpler. To this end, I propose a way to enlist the mischief of Fortuna in a manner that could lead to a Pareto improvement in the system. I turn now to those built-in contradictions in peer review to set the stage for my reform proposal.

3. The Inherent Contradictions of the Peer Review System

Some domains of science, such as professional publications, adopt a double-blind peer review system, in order to reduce the influence of factors external to merit in evaluation. But the grant selection process blinds applicants only; reviewers cannot be blindfolded if they are to judge applicants' qualifications and probabilities of success when performing the proposed projects. For this very reason, prestige is an inescapable factor in grant evaluations, because reviewers are likely to ascribe a high probability of success to established scientists and to deposit relatively less confidence in their less-famous peers, in accordance with the Matthew effect. Likewise, the factor of personal amity and antipathy is hard to discount entirely, and so is the more insidious factor of unconscious bias.

The possibility of a double-blind peer review of grant applications has been considered elsewhere. For instance, Taiwan adopted it (Silver 2019). It is, however, impractical in US federal agencies because the laws and norms of public accountability require prior scrutiny of prospective government contractors' track records. Any agency proposing to implement a double-blind grant review process would invite legal challenge and quite possibly public outrage. I will thus proceed assuming that reviewers must know the names and backgrounds of applicants.

A second aspect of peer review that is both necessary and a dogged problem is reviewer disagreement. We learned from the studies of the NSF and NIH processes that discrepancies among reviewers, even widely divergent judgments, are both unavoidable and pervasive.

Averaging two discrepant scores, a common solution, is unsatisfactory: If the high-score review is closer to the mark, an average score will shortchange the proposal. In turn, if the low-score review is the fair judgment, an average score will prop the proposal above others that may be less inadequate. The fact that expert disagreement is legitimate does not imply that wide discrepancies can both be correct, in the narrow sense of correct with respect to the agency's interest. If the test of correctness is who gets funded, then, ex post facto, only one of two widely divergent judgments can be correct.

An alternative to averaging, and possibly a superior solution, is to commission a tiebreaker review and then discard the outlier. This solution is less common than averaging scores because it increases the cost and time of review. Yet there could be ways to manage this solution economically by requiring reviewer agreement in proportion to the budget of a grant proposal. For instance, maybe two reviewers out of three are sufficient to decide on small grants but many more reviewers should agree in order to secure funding for multimillion-dollar projects: perhaps at least eight yea votes in a twelve-person panel (preserving the two-thirds supermajority rule).

Introducing a tiebreaker or, more generally, forcing agreement among reviewers is better than averaging scores, but it is not entirely risk free. A well-understood side effect of peer review is a conservative tilt in science, because the scrutiny of peers makes researchers risk averse. Many researchers who would otherwise take risks to advance knowledge in leaps instead settle for the safety of gradualism. Chubin and Hackett (1990) document how grant applicants tend to believe that reviewers are put off by unorthodox research questions and methods and consequently forgo risky proposals. Travis and Collins (1991) in turn document "cognitive cronyism" or favoritism toward research proposals from the same epistemic family as the reviewer, and by the same token, skepticism about projects from competing or alien camps. Emphasizing agreement in grant selection could reinforce this effect by attracting more of the same: risk-averse proposals, timid research questions, and time-tested methods.

Any improvement to the grant selection process that keeps peer review at its core must contend with these two characteristics of the system: First, reviewer disagreement is inescapable; therefore, some chance is unavoidable. Second, demanding consensus in the review process will discourage risky projects and intensify gradualism in science. The following proposal seeks to meet these challenges.

4. A Modified Lottery of Grants

Given the importance of grant selection in the advancement of science and in the advancement of researchers' careers, even a gradual improvement of the system should produce ample benefits for science and society. I argue in the rest of this paper that formalizing the role of "chance" in the process without banishing merit should improve upon the current system. This can be accomplished by adopting a modified lottery. I borrow from Fang and Casadevall (2016), who originally proposed this idea for research grant selection, and from the more recent work of Osterloh and Frey (2020), who proposed it for paper publication at peer-reviewed journals. As noted earlier, a modified lottery has already been implemented by the Health Research Council of New Zealand (for entry-level exploratory grants) and more recently by the Swiss National Science Foundation on a limited basis (Adam 2019).

A modified lottery of grants has two stages. In the first stage, peer reviewers identify a pool of high-quality proposals. In the second stage, awardees are selected using a lottery. The funding agency would then publicly announce the list of finalists and the lottery winners.

I now turn to the apparent benefits and risks of this system for researchers, for science, and for the management of public affairs.

4.1 Pros and Cons

The benefits for researchers, in terms of career advancement, could be significant, because federal agencies will put a garland on all finalists' heads, not only on those of the awardees, thus expanding the number of beneficiaries of the agency's coveted recognition. It will be public knowledge that, in terms of merit, all finalists are equals and those who got the money were simply lucky. Universities and professional societies will have reason to give finalists credit for having been selected as finalists in grant contests—in the form of promotion, honors, and other rewards—even when they do not hit the jackpot.

The problem of implicit bias will not be resolved, but perhaps will be somewhat mitigated when peer review is limited to the first stage of the grant review process. The second stage, by design, cannot be polluted by any sort of bias because winners will be picked randomly.

The benefits for science will result from limiting peer review to the first stage of the process. Grant applicants will understand that, should their proposal be judged rigorous enough to deserve a lottery ticket, in the second stage their chances of winning will be the same with a conservative project as with a more risky proposal. Consequently, the proposed reform could rebalance the science portfolio, allowing for a marginal increase in the risk appetite of researchers—which is likely to accelerate the advancement of knowledge. The world of finance offers a useful analogy: in order to increase expected returns, fund managers rebalance their investment portfolios to tolerate a marginal increase in risk.

The benefits in terms of better-managed federal research agencies are also predictable. First, organizing a lottery will be less expensive than organizing lengthy full-panel deliberations. The consequent administrative cost savings can be reallocated, if this is permitted by statute, to increase funds for research grants. Second, at no additional cost to taxpayers, research agencies will create additional value for society when the number of researchers receiving recognition jumps by an order of magnitude, from a few awardees in the current system to all the finalists in a modified lottery. (Recall that there are no discernible differences in quality between the finalists in the proposed system.) Third, the implementation of a modified lottery is feasible. It could follow a gradual path, starting with a pilot test for early-career grants and expanding as agencies improve its design to better align the research they fund with their mission and related stock of human capital.

There are foreseeable costs—for researchers, for science, and for the public administration—of implementing a modified lottery. Talented researchers may be discouraged if they experience persistent bad luck. Below, I suggest a tweak to the basic scheme that would make it improbable for a specific researcher to be an unfunded finalist too many times. A drawback for science is the opportunity cost of forgoing funding for projects of eminent quality and promise. Any cohort of applications should have a few projects of undisputable merit that a full panel of reviewers would readily agree to recommend for funding. Again, a small tweak to the basic format of the lottery could reserve a margin in the budget to guarantee funding for outstanding projects.

As for costs to funding agencies, the most important is related to public accountability. Agencies implementing this funding proposal will have to implement evaluation programs that compare the outcomes of modified lotteries to those of traditional processes. The evaluations themselves are not easy to design and will have to be performed over many years, because the outcomes of research are realized at successive points over a long time horizon.

I submit that the benefits of a modified lottery system are likely to outweigh the costs, particularly when small tweaks to the basic design can mitigate some costs. I now turn to those potential tweaks or variants.

4.2 Building upon the Basic Format

To address the problem of persistent bad luck, the process could be amended to increase the chances of winning the lottery every time the same researcher is selected as a finalist. Consider a simple rule that would double the chances of every finalist who was an unfunded finalist in the previous round. The doubling of chances should be compounding, with every successive failure leading to double the prior chances: the first miss would double the chances the next time, the second miss would quadruple the chances, and so on. With such a rule, bad luck could not last too long.

This scheme should discourage the recycling of proposals and the crowding out of new researchers. To these ends, only new (or substantively new) proposals should count for the doubling of chances; recycled proposals should be accepted but given no additional chances. Also, every new funding cycle would attract a mass of repeated finalists, and for that reason a quota for unfunded finalists might be necessary to allow new names and projects to have a fair winning chance.

To address the "problem" of eminently meritorious research projects, the funding unit could establish an excellence fund in each round separate from the lottery fund. Projects that clearly stand out could be marked in the first round for full-panel review, and if a consensus quickly forms about their merit, they could be withdrawn from the lottery but recommended for support from the excellence fund, thus guaranteeing them funding. It is important to note that public recognition of finalists should not differentiate between projects from the excellence fund and from the lottery fund, to preserve equality among all finalists. Truly outstanding projects that produce outstanding results will eventually be publicly recognized as such.

Further adaptations of this process may be needed to better align research funding with agencies' missions. The most obvious of these improvements is for agencies to allow funding units to pool applicants by type of grant, where types could be different themes or award sizes. Units could allocate their research budget among thematic tracks in accordance with the importance of those tracks to the agency's mission. Likewise, funding units could pool lotteries by grant size, reflecting the level of seniority of the researchers or the maturity of the proposed projects to be funded.

In order to allow applicants to estimate their chances in the lottery, funding units could announce in advance the total amount of funding available for each pool. Consider the following concrete (albeit imaginary) example. A program with an $27 million budget could announce three funding lottery pools:

- pool S of small grants, with a total of $12 million, for grants up to $500,000
- pool M of medium grants, with a total of $9 million, for grants up to $2 million
- pool L of large grants, with a total of $6 million, for grants up to $6 million

Another improvement should result from introducing incentives for researchers to submit the lowest possible budget necessary to execute their projects. A rule to this end could make the project's probability of winning the lottery inversely proportional to the project's budget within its pool. In my imaginary example, this idea could be operationalized by issuing lottery tickets (each ticket has the same probability of winning), and distributing tickets to finalists selected for pool S as follows: five tickets for budgets up to $100,000, four tickets for budgets from $100,001 to $200,000, three tickets for budgets from $200,001 to $300,000, two tickets for budgets from $300,001 to $400,000, and one ticket for budgets from $400,001 to $500,000. In this system, the least expensive proposal in the pool has five times the chances of the most expensive proposal.

Another incentive to drive budgets to their minimum feasible amount would be to require a higher number of agreements among reviewers for each successive level of funding. In my imaginary example, the selection of finalists for pool S could be based on the agreement of two reviewers out of three, for pool M on the agreement of four reviewers out of six, and for pool L on the agreement of eight reviewers out of twelve. In this manner, applicants would internalize the fact that it is far easier to become a finalist in the S pool than it is in the L pool. This is also a way to address the risk-aversion problem with peer review, because less expensive proposals satisfy fewer peers and can therefore take more risks, while larger projects remain safe bets on existing puzzles and tested methods.

The variants entertained above are not nearly exhaustive of all the possibilities, but they do exemplify workable adjustments to the basic lottery scheme that could mitigate some of its anticipated shortcomings. In section 5, I turn to the question of potential effects on productivity from winning a lottery ticket.

5. Will a Lottery Discourage Effort?

Assigning a formal role to chance in grant selection presents a collateral problem: Will lottery winners underperform their research contracts? In a pure lottery, tickets would be given without regard to experience or track record. But in a modified lottery, researchers must first satisfy the quality screening before their chances of winning are divorced from past performance.

Yet assessing research performance is neither simple nor clear-cut. First, performance could be measured in process or in outcomes, and the two are not always directly related. Excellent research work does not necessarily yield excellent outcomes and serendipity often plays an outsize role in discovery. Second, neither process nor outcomes are perfectly legible to the funding agency.

Agencies may contract with expensive auditors to review projects. Though these auditors are scientific experts, they may nevertheless reach a very partial assessment of effort, because it is often hard to discern the level of effort put into performing a project. The costs of such audits seem to far exceed the oversight information gained, and thus they are rare. Less-expensive visitations can clear some of the fog about performance, but they are not designed to be probes like audits. In the end, funding agencies will only gain partial information about researchers' efforts. Performance evaluations, as much as other evaluative processes in science (such as publications and tenure promotions), must rely on the good faith of researchers and evaluators.

Some research outputs signal potential for good outcomes. A well-written report, a handful of publications, and a few invention disclosures (filed with the researcher's university) help funding agencies justify, ex post facto, their research investments. Still, the originality or importance of a publication usually takes time to become evident within its own field of inquiry, and more time to be useful in adjoining fields of application. It is well known that the value of research compounds over time, and research achieves its highest value only after a long maturation period, as suggested by the elongating Noble prize delay (Becattini et al. 2014).

Compared to publications, invention disclosures are closer to useful application. But usefulness still depends on a myriad of actors and circumstances. The university's office of technology transfer must find sufficient commercial potential in the disclosure to file a patent, and once the patent is granted, there will still be significant uncertainty about licensing that patent to a technology developer that could eventually make something that is sold and bought. In sum, publications and invention disclosures are measurable outputs but imperfect signals of what really matters: good outcomes. Still, they are usually the best data available to allow agencies to form an idea, however limited and uncertain, of the performance of the research they sponsor.

Against the backdrop of this information asymmetry—researchers know much more about their performance than the agencies funding them do—the modified lottery model may preserve just enough incentive for good performance, because researchers will be aware that future funding rounds will examine their track record with the agency. The first round should filter out not only researchers cited for bad behavior—thus excluding truants and cheaters—but also researchers whose performance was so visibly mediocre that agencies recorded it on their records. On top of this, the modified lottery could involve an increased probability of winning the lottery for past performance awarded a special mention by the agency. Using a similar mechanism to the one described earlier for increasing the chances of previously unfunded finalists, good performance could be rewarded with greater odds of winning.

In the appendix, I use a simple mathematical model to examine the limits of linking effort to the probability of winning the lottery. Still, agencies should be cautious not to reward past performance so much that they discourage young researchers from entering the contest. Perhaps, then, separating pools by experience, as suggested in the previous section, could be a viable and fair alternative.

It is no small challenge to introduce effective incentives for researchers to perform their contracts to the best of their ability. Let's acknowledge that the mores and traditions of scientific inquiry foster a sound work ethic, and let's recognize that energy and focus in research are often driven by the ambition of priority in discovery. Still, mundane incentives are important, because they reinforce those organizational and social dynamics and could make a difference at the margin of vigor in the research enterprise. For that reason, those involved in designing award selection mechanisms must take into consideration the link between the modified lottery and the productivity of sponsored research.

6. Conclusion

The proposal presented in this paper is for federal agencies to adopt a modified lottery for the selection of research awards. Meritorious projects would first be selected using peer review; then awards would be granted to projects chosen from that pool of applicants using a lottery.

This proposal internalizes chance, which has been found to be prevalent and pervasive in the current award selection process. A modified lottery would increase the overall fairness of the process, both by distributing recognition to a wider pool of meritorious applicants and by mitigating, in the second stage of the selection process, biases such as the Matthew effect, personal favor, and unconscious bias. This reform should also reduce the costs of the selection process, which currently entails convening panels of experts for long deliberations. Those savings could be productively redirected to increase research funding.

I have also discussed above a number of variations on the basic modified lottery that could improve the management of grant selection, increase the fairness of the process, and tailor the process to the needs and budget constraints of agencies and their programs. Likewise, it would be possible to link the past performance of research grant recipients to the probability of success in the second-stage lottery, thus adding an incentive for research contractors to perform at the best of their ability.

Further work on this topic should include studying how to adapt this general proposal to the specific needs of research agencies, designing its gradual adoption, and planning evaluation programs for public accountability and to improve upon the original design.

Federally sponsored research is just too important not to leave it to chance—at least partially.

Appendix: A Stylized Representation of the Effect of Chance (in Award Selection) on Effort in Grant Performance

I will first state a few necessary assumptions:

- *A1*. Complete information: the agency knows the researchers' payoff functions.
- *A2*. Imperfect information: the agency can only partially observe the level of effort of researchers.
- *A3*. Decisions are based on a cost-benefit analysis, requiring merely a positive balance in the expected utility, which is not necessarily the same as maximizing expected utility.
- *A4.* The cost-benefit analysis for researchers includes the costs incurred today and the rewards that can possibly be received tomorrow. Subtracting the costs of a project in funding round t from the benefits from the next round $t + 1$ establishes the causal link between effort and reward.

Let the utility function of researchers (as a linear function of grant money and effort) be

$$U_S(w, e) = I(w) - C(e) = \rho w - e.$$

Utility gains for researchers are a function I(·) of the size of the grant w. The disutility of effort is a function C(e), where effort e can take two values, high or low. The function is defined such that $C(e_h) > C(e_l)$. The multiplier $\rho > 0$ reflects the fact that the benefits to researchers from obtaining a grant exceed the monetary value of it.

Within the time frame of the cost-benefit calculus, the researchers' choice is to play high or low effort in this round, and the agency's choice is to fund or not fund a new project the next round.

Figure 1 shows the normal-form representation of this game using a numerical example to illustrate the point visually. Let researchers' costs be $C(e_h) = 1$, $C(e_l) = 0$ and their benefits be 2 (with $w = 1$ and $\rho = 2$) if their project is funded but 0 otherwise (with $w = 0$). Recall that

$E[U_S(w, e)] = \rho w - e$. The cost for the agency is the monetary cost of the grant w, which is 1 when the agency funds the project and 0 when it does not. Let the agency derive benefits of 2 from high effort and 0 from low effort.

This is a classic prisoner's dilemma. It is well known that the pure strategy equilibrium of this game is suboptimal. In this case, that means researchers play low effort and the agency cuts funding at the end of the only period of publicly funded research:

$$\text{Pareto inferior outcome: } (e_h, e_l)_t \rightarrow (0, 0);\ t = 1 \text{ (only one period).}$$

But this game is not played only once. This is important because a wide variety of noncooperative games (including the prisoner's dilemma) can be induced into cooperative equilibriums if the game is set to be played indefinitely. As the players realize that by coordinating their actions they can achieve better outcomes, they build trust and learn to cooperate (see standard discussion in Gibbons 1992). Trust, in this context, is the belief that while one player plays cooperation, the other will follow suit. If researchers are reassured that the agency is committed to funding their research in every successive period and if the agency has reasons to believe that researchers will play high effort at every iteration of the game, then both will cooperate:

$$\text{Pareto optimal outcome: } (e_h, e_l)_t \rightarrow (1, 1);\ \text{all } t = 1, 2, \ldots, \infty.$$

To induce a Pareto optimal outcome, the interests of researchers and the agency must be aligned. An example of induced alignment is contingent executive compensation. The board of directors of a firm (principal) can offer the executive officer (agent) a compensation contract tied to long-term earnings or stock options instead of providing a fixed salary. In this manner, the interests of the agent become aligned with those of the principal.

Figure 1. The Prisoner's Dilemma in Science Funding

		agency	
		funding	not funding
researchers	high effort	1, 1	−1, 2
	low effort	2, −1	0, 0

The proposed modified lottery could achieve similar results to contingent executive compensation in research award selection. The funding agency promises to fund a project with probability α (the chances of the lottery) after projects are screened for quality and the screening includes a review of past performance. If α can be set a little higher for any project associated with distinguished past performance (set $\alpha = \alpha_1$), a level above the probability of all other finalists without such bona fides ($\alpha = \alpha_0$), then researchers may be induced to play high effort.

How can an agency approximate the necessary extra $\alpha_1 - \alpha_0$?

Researchers will play high effort if the expected utility of doing so is greater than the expected utility of playing low effort: $E[U_S(w, e_h)] > E[U_S(w, e_l)]$. In the linear functional form, $\alpha_1 \rho w - C(e_h) > \alpha_0 \rho w - C(e_l)$. Ergo,

$$\alpha_1 - \alpha_0 > \frac{C(e_h) - C(e_l)}{\rho w} > 0. \tag{1}$$

Since we defined $C(e_h) > C(e_l)$, the extra α must indeed be at least a little something. But how much? This condition shows that the agency needs to increase α for good performers at least as much as the difference between the cost-benefit ratio for high effort and the same ratio for low effort.

Let's take two numerical examples to figure out what this condition means.

- *Illustration A of equation (1).* If the costs, as a share of the grant amount, run at 10% for high effort and 5% for low effort and the multiplier of benefit is 5 (recall that ρ multiplies the total grant amount w), then the agency needs to inflate α by more than one percentage point to induce high effort.
- *Illustration B of equation (1).* However, if the costs, as a share of the grant amount, run at 40% for high effort and 20% for low effort and the multiplier is a mere 2, then the agency needs to prop α by ten percentage points to induce high effort.

If the agency has set out to select 100 finalists from the screening stage, numerical illustration A means that each high performer selected must displace one of those 100, and numerical illustration B suggests that each high performer must displace 10 others. The first scenario could be tenable, but the second scenario is hardly so. It all depends on how researchers define the cost of effort with respect to the grant amount and how important that particular agency's award is for their careers. Costs could be inferred from the portion of salary that researchers claim the award will cover. It is much harder to estimate the multiplier, or prestige factor, attached to a particular grant scheme.

This simple model suggests that setting a higher α to reward prior high performance is feasible but fraught with difficulties. Separating a pool for recognized high performers from the common lottery pool could be a superior alternative, because it seems fair to give all finalists within their respective pools equal chances to win. This conclusion also affirms the importance of reviewing the track record of applicants in the screening stage of the modified lottery.

References

Adam, David. 2019. “Science funders gamble on grant lotteries.” *Nature* 575(7785): 574–75.

Becattini, Francesco, Arnab Chatterjee, Santo Fortunato, Marija Mitrović, Raj Kumar Pan, and Pietro Della Briotta Parolo. 2014. “The Nobel Prize delay.” *arXiv:*1405.7136v1[physics-soc-ph]. Accessed February 20, 2021: https://arxiv.org/pdf/1405.7136.pdf

Bush, Vannevar. 1945. *Science—the endless frontier: A report to the president on a program for postwar scientific research*. Reprinted as part of the thirtieth anniversary observation of the NSF 1950–1980. Washington, DC: National Science Foundation.

Chubin, Daryl E., and Edward J. Hackett. 1990. *Peerless science: Peer review and US science policy*. Albany, NY: SUNY Press.

Cole, Stephen, Leonard Rubin, and Jonathan R. Cole. 1978. *Peer review in the National Science Foundation: Phase one of a study*, Committee on Science and Public Policy. Washington, DC: National Academy of Sciences.

———. 1981. *Peer review in the National Science Foundation: Phase two of a study*, Committee on Science and Public Policy. Washington, DC: National Academy of Sciences.

Cole, Stephen, and Gary A. Simon. 1981. “Chance and consensus in peer review.” *Science* 214(4523): 881–86.

Fang, Ferric C., and Arturo Casadevall. 2016. “Research funding: The case for a modified lottery.” *MBio* 7(2): e00422-16. doi:10.1128/mBio.00422-16.

Gibbons, R. 1992. *Game theory for applied economists*. Princeton, NJ: Princeton University Press.

Ginther, Donna K., Walter T. Schaffer, Joshua Schnell, Beth Masimore, Faye Liu, Laurel L. Haak, and Raynard Kington. 2011. “Race, ethnicity, and NIH research awards.” *Science* 333(6045): 1015–19.

Kaatz, Anna, Wairimu Magua, David R. Zimmerman, and Molly Carnes. 2015. “A quantitative linguistic analysis of National Institutes of Health R01 application critiques from investigators at one institution.” *Academic Medicine: Journal of the Association of American Medical Colleges* 90(1): 69–75.

Kotchen, Theodore A., Teresa Lindquist, Anita Miller Sostek, Raymond Hoffmann, Karl Malik, and Brent Stanfield. 2006. “Outcomes of National Institutes of Health peer review of clinical grant applications.” *Journal of Investigative Medicine* 54(1): 13–19.

Ley, Timothy J., and Barton H. Hamilton. 2008. “The gender gap in NIH grant applications.” *Science* 322(5907): 1472–74.

Merton, Robert K. 1968. “The Matthew effect in science: The reward and communication systems of science are considered.” *Science* 159(3810): 56–63.

Osterloh, Margit, and Bruno S. Frey. 2020. "How to avoid borrowed plumes in academia." *Research Policy* 49(1): 103831.

Pier, Elizabeth L., Markus Brauer, Amarette Filut, Anna Kaatz, Joshua Raclaw, Mitchell J. Nathan, Cecilia E. Ford, and Molly Carnes. 2018. "Low agreement among reviewers evaluating the same NIH grant applications." *Proceedings of the National Academy of Sciences* 115(12): 2952–57. doi:10.1073/pnas.1714379115.

Pier, Elizabeth L., Joshua Raclaw, Anna Kaatz, Markus Brauer, Molly Carnes, Mitchell J. Nathan, and Cecilia E. Ford. 2017. "'Your comments are meaner than your score': Score calibration talk influences intra- and inter-panel variability during scientific grant peer review." *Research Evaluation* 26(1): 1–14.

Silver, Andrew. 2019. "Taiwan considers double-blind peer review for grants." *Nature*, May 24, 2019. doi:10.1038/d41586-019-01651-3.

Travis, G. David L., and Harry M. Collins. 1991. "New light on old boys: Cognitive and institutional particularism in the peer review system." *Science, Technology, & Human Values* 16(3): 322–41.

Zuckerman, Harriet. 1977. *Scientific elite: Nobel laureates in the United States*. New York: Free Press.