Does Peer Review Penalize Scientific Risk Taking? Evidence from NIH Grant Renewals*

Pierre Azoulay
MIT and NBER
100 Main Street, E62-487
Cambridge, MA 02142

Wesley H. Greenblatt
Boston Children's Hospital
and Harvard Medical School
300 Longwood Avenue
Boston, MA 02115

October 23, 2026

Abstract

Science funders loudly profess their appetite for "high-risk, high-reward" research. Practicing scientists insist, just as loudly, that peer review punishes it. We study the determinants of renewal for more than 100,000 cycles of R01 grants from the National Institutes of Health between 1980 and 2015. We define scientific risk taking as an *ex ante* choice—the commitment of effort to a project whose distribution of outcomes is wider, less well understood, or less well charted than the alternatives the investigator could have pursued—and we construct four distinct measures of chosen ambition: extreme tail outcomes, disruptiveness, pivoting from an investigator's prior work, and standing out from the crowd in one's field. After controlling comprehensively for investigator, institution, and grant characteristics, we find that riskier grants are renewed at markedly lower rates across all four measures. The penalty is sharper for new investigators, and it tends to be sharper for more novel research.

Keywords: scientific risk, scientific productivity, peer review, NIH, government funding.

JEL Classification: O32, O38, H51

*We thank Alessandro Bonatti, Matt Clancy, Danielle Li, and Scott Stern, as well as seminar participants at the Max Planck Institute for Innovation and Competition, NBER Summer Institute, and NBER Productivity Seminar for useful feedback. We are grateful to Michael Frakes and Melissa Wasserman for sharing their data on the patents of biologic drugs, to Nicola Melluso for data on the reuse of new noun phrases, and to Roger Masclans for computing the commercial potential of publications that the public scores do not cover. The data and the code that reproduce every table and figure of the paper and its appendices are available at https://github.com/pierre-azoulay/nih-peer-review-risk.

1 Introduction

No funder of science has ever prided itself on supporting "low risk, low reward" projects, yet few ideas elicit as much agreement as the claim that peer review punishes risk taking. These laments appear in the popular press, in U.S. Congressional testimony, or the pages of *Nature*, and are authored by successful investigators and leading science policy makers (Kolata 2009, Kornberg 2007, Woolston 2014). Prominent scientists have penned editorials satirizing the behavior of risk-averse grant funders, and ponder whether Nobel Prize winners could receive funding for their seminal discoveries in the modern funding climate (Fields 2014, Petsko 2012). While economists would caution that a benevolent social planner might want to tame individual scientists' risk appetites, the proposition that current institutions do not adequately reward risky research strikes practicing scientists as self-evidently true.

Yet for all the colorful anecdotes, blue-ribbon panels, and editorials, there is surprisingly little systematic evidence that peer review punishes risk taking, and less evidence still that society as a whole would be better served if funders relaxed their conservatism. In a survey of investigators supported by *Fast Grants*, 78% said they would substantially change their research program if they were freed from the demands of funders—pursuing more ambitious research, pivoting to new topics, and testing hypotheses others see as unlikely to succeed (Collison et al. 2021). But whether peer review is the mechanism behind such self-censoring, and how much it moves the research frontier in the aggregate, the systematic evidence does not yet tell us. Azoulay et al. (2011) compare two grant funding mechanisms—the Howard Hughes Medical Institute Investigatorships and National Institutes of Health (NIH) R01 grants—and provide evidence that the former encourages more risk taking than the latter, but this does not, in itself, imply that peer review at NIH punishes risk taking. Other scholars have examined how novelty shapes grant funding outcomes (Ayoubi et al. 2021, Packalen and Bhattacharya 2020, Veugelers et al. 2026). Yet novelty, as we argue at some length below, is not the same thing as risk taking: investigators can pursue novel research that is not risky, and risky research that is not novel.

Part of what has held back empirical progress on this question is that "risk taking" has been used loosely in the science-policy literature, sometimes to refer to outcome variance, sometimes to novelty, sometimes to uncertainty, ambiguity, or complexity. Drawing on the project-management framework of Pich, Loch, and De Meyer (2002), we conceptualize scientific risk taking not as an attribute of a project but as an *act of chosen ambition*: the decision by an investigator to commit effort to a project whose distribution of outcomes is wider, less well understood, or less well charted than the alternatives she could have pursued. Section 2 develops this framework

and situates it in the literature on risk and uncertainty in science (Franzoni and Stephan 2023, Franzoni et al. 2022).

How much ambition a scientist is willing to accept depends on more than her private taste for risk; it depends, just as forcefully, on the institution that will decide whether her proposal is funded. In US biomedical research, that institution is the NIH, which allocates its grants through peer review: scientific proposals are evaluated by a panel of experts, with the resulting scores having a major influence on the likelihood of funding. At the end of a grant cycle, a funded scientist has the option of extending the research program by applying for an additional cycle of support. That renewal decision, a second encounter between the investigator's work and NIH, is what we study in this paper. At renewal, the study section reads a record of what the grant produced as well as a plan for the next cycle, and that record is also what allows us to measure the risk the investigator took. To study the decision with full precision, one would ideally have access to all proposals (regardless of funding status), reviewer scores, and grant outcomes in the form of publications and citations traceable to a particular funding stream. With access limited to publicly available data, we focus on funded applications and ask whether risk taking is associated with the likelihood that a grant is renewed competitively for an additional cycle. Our findings concern the competitive renewal: every grant we study first had to win its initial award, in a more selective competition,[1] whose treatment of risk we cannot observe. We link each grant cycle with the publications it produced, leveraging the requirement that authors acknowledge their NIH grant support in published work.

R01 grants are the "bread and butter" of academic biomedical research, and a requirement for scientists who aspire to establish and sustain a laboratory over the long run. We study how risk taking is penalized or rewarded in 103,484 NIH-funded R01 and R01-equivalent grant cycles that started between 1980 and 2015, after controlling for a comprehensive set of covariates for investigator demographics, institution quality, and grant characteristics (among them the number and quality of the publications produced during the cycle). Because a single measure is unlikely to capture the full range of what chosen ambition can mean in practice, we construct four measures, each corresponding to a different form of information inadequacy that an investigator can choose to accept. *Extreme tail outcomes* captures the classical variance notion: the acceptance of a project whose realizations may range from breakthrough to failure (Azoulay et al. 2011). *Disruption* captures the acceptance of a project whose realized direction breaks from the field's prevailing causal mapping (Funk and Owen-Smith 2017). *Pivoting* captures investigator-level

[1] From 2015 to 2025, 18 percent of new applications for R01-equivalent grants were funded, against 40 percent of applications for competitive renewal (Appendix A).

radical uncertainty: the scientist enters a topic where her own prior expertise is thinnest. *Standing out* captures ambiguity at the level of reception: the project lives in a sparsely populated region of the field's state space, so that even conditional on scientific success, how the work will be received is itself uncertain. Each measure is built from the publications that acknowledge the grant during the cycle that is up for renewal: the record of the work done under the grant, which the study section reads when it assesses the application for another cycle. We infer from that work the choice the investigator made when she committed to the project, and our claim concerns the risk that reviewers can see in the record when they assess the application; Section 3 states what this interpretation requires.

Across each of these measures, we find that risk taking is penalized. Grant cycles in the top decile of risk taking are renewed 4.5 to 8.3 percentage points less often than those in the bottom decile, depending on the measure, a gap of 11 to 19 percent of the mean renewal rate. Each measure keeps its association with renewal when the other three are held fixed, so the penalty on one measure does not merely reflect its correlation with another. In sharp contrast, novelty on its own is associated with *higher* renewal rates under our headline measure—direct evidence that novelty and risk taking are distinct constructs that it is best not to conflate, and that the penalty we document is not a re-badged novelty effect.

We use "penalty" throughout the paper in a descriptive sense: a negative partial association between chosen ambition and renewal, conditional on our controls. The word names the sign and magnitude of the relationship—nothing more. This use differs from that of Hill et al. (2025), who find that pivots lower the impact of research, an outcome desirable without qualification. Our outcome is a funding decision, and a lower probability of renewal is a loss only if the work that was not renewed would have been worth funding. The documentation of this penalty is, we believe, on firm empirical ground. Whether the penalty is *excessive*—whether peer review discourages ambition beyond what a well-informed social planner would choose—is a harder question, and the evidence we can bring to bear on it is necessarily more circumstantial. We present two pieces of evidence consistent with the claim that the penalty exceeds what appropriate conservatism would require. First, the gradient in the renewal-risk relationship is steeper for new investigators on three of our four measures and for more novel work on all four, although for novel work the difference is statistically significant only for disruption and standing out. These are the two groups for whom exploratory research is plausibly most valuable from a social standpoint, given the knowledge spillovers associated with early-stage basic research (Azoulay and Li 2022). Second, in the five years after the cycle, grants whose work was new to the investigator (above the median of pivoting) are followed by more patents arising from the grant, whether or not they

were renewed. That is what one would expect if study sections held this kind of risk to a higher bar; the other three measures show the pattern in part or not at all.

Our study makes three contributions. First, we commit to an explicit, choice-based definition of scientific risk taking and situate it within the information-adequacy framework of Pich et al. (2002). Extending the framework to the selection of a project lets us organize an eclectic set of risk-taking measures under a single umbrella, while being transparent about what each measure captures—and it cleanly separates risk taking from the related but distinct construct of novelty, which has received the bulk of prior empirical attention (Ayoubi et al. 2021, Boudreau et al. 2016, Packalen and Bhattacharya 2020, Teplitskiy et al. 2022, Veugelers et al. 2026). Second, rather than characterizing an investigator's prior research trajectory to predict future funding—an approach common in previous studies—we focus on risk measures directly tied to a specific grant proposal and the publications emerging from it. This distinction matters because researchers often manage multiple projects simultaneously, each carrying different levels of risk; an investigator's past work may not accurately reflect the level of risk in a particular grant proposal, and in some cases scientists may deliberately avoid proposing high-risk ideas to funders they suspect might penalize bold experimentation, routing such projects through alternative funding sources instead. Third, by examining the entire universe of R01 NIH grants over more than three decades, we study peer review's treatment of risk taking as a systemic feature of the funding system rather than as an artifact of any single competition.

Our closest intellectual antecedents are Boudreau et al. (2016), who experimentally assign grant proposals to peer evaluators, and Carson et al. (2023), who manipulate the distribution of reviewer scores to assess how peers (rather than funders) value disagreement among reviewers. Like ours, these studies exploit proposal-level variation to uncover the determinants of funding success. Unlike ours, they control the treatment of interest in a framed experimental design—an unambiguous methodological strength. However, they do not measure proposal-level risk, nor can they replicate the high stakes of NIH grant funding competitions. Our approach relies on observational data. While this rules out the identification of causal effects, it allows us to examine risk taking as the primary variable of interest in a real-world setting of substantial policy interest—the entire universe of R01 NIH grants over an extended time period.

The remainder of the paper is structured as follows. Section 2 develops our definition of scientific risk taking, situates it within the Pich et al. framework, and examines how peer review may influence the level of chosen ambition investigators are willing to propose. Section 3 introduces the setting, data, and measurement choices, including the four risk-taking measures and what their interpretation requires. In Section 4, we describe our empirical strategy and provide

descriptive statistics. Section 5 presents our main results. Section 6 concludes by discussing the broader implications of our findings for the design of peer review and the allocation of research funding.

2 Scientific risk taking and peer review

"Risk taking" has been used at various times in the science-policy literature to mean outcome variance, novelty, uncertainty about feasibility, or ambiguity about reception (Althaus 2005, Franzoni and Stephan 2023). When different participants in the debate use the same word for different phenomena, they can agree that something is wrong while disagreeing entirely about what it is. We take this conceptual slippage to be a first-order obstacle to empirical progress on the question of how peer review treats risk, and we confront it directly. We define scientific risk taking as an *ex ante* choice made by an investigator: the decision to commit effort to a project whose distribution of outcomes is wider, less well understood, or less well charted than the alternatives she could have pursued. We use "risk taking" and "chosen ambition" interchangeably. Three features of this definition matter for what follows.

Risk taking is an act, not an attribute of a project. It describes what a scientist *does* at the moment of commitment, not a fixed property of the question being studied. The same research question may represent a bold wager for one investigator and a routine extension for another—what separates them is not the question but the menu of projects each could plausibly have pursued instead. A scientist at the frontier of her field may propose a well-charted continuation of her existing program, or commit to an investigation that departs further from what she and her peers currently know; the choice is hers, and it is what we mean by chosen ambition.

The definition is deliberately broad. The literature on risk and uncertainty in science has long recognized that a unitary construct is inadequate to describe the forms of ignorance that pervade research (Franzoni and Stephan 2023, Franzoni et al. 2022). Following the lineage that runs from Knight (1921) through contemporary treatments, the distribution of outcomes a scientist faces may be characterized by *risk* in a narrow sense—known probabilities over known states—but may equally involve *uncertainty* (unknown probabilities over known states), or *radical uncertainty* in which even the space of possible outcomes is incompletely specified. In practice, most scientific projects involve some blend of all three, and the distinction among them is not always sharp even after the fact. Our definition encompasses all three.

This breadth has a cost for measurement, to which we return in Section 2.2. For now we note only that our definition differs from the tidier notion of risk that prevails in financial economics, where outcomes are well-defined, their distribution is at least estimable, and risk taking reduces to a choice of position along a mean-variance frontier. Science is seldom so tidy. Failed experiments can generate insights that raise the expected return of later work (Frankel et al. 2025); successful experiments may deliver discoveries whose ultimate value is not knowable for decades (Rosenberg 1996); and what counts as success depends on what the scientific community will eventually recognize. The scientist's choice of what to study is less a position on the frontier than a wager on where the frontier will be.

The definition separates risk taking from novelty. Novelty describes how far a project departs from prior art—the distance between its intellectual content and the existing literature—whereas risk taking describes a decision about committing to a less well-charted outcome distribution. A project can be novel without being risky: a well-understood extension into a previously unexplored but conceptually straightforward corner of the literature may be novel by construction yet carry little uncertainty about whether it will succeed or what success will look like. Conversely, a project can be risky without being novel: an attempt to overturn a long-established result may engage the same concepts, methods, and literature as the orthodoxy it challenges, yet commit the investigator to an outcome distribution ranging from vindication to career-damaging failure. Previous empirical work on the evaluation of scientific proposals has focused largely on novelty (Ayoubi et al. 2021, Boudreau et al. 2016, Packalen and Bhattacharya 2020, Veugelers et al. 2026), and novelty is indeed a feature of many risky projects—but the overlap is neither complete nor unidirectional, and collapsing the two obscures more than it clarifies.

The framework most naturally paired with this view of risk taking is Manso (2011), whose model of exploration versus exploitation formalizes the contrast between safer, incremental projects and riskier exploratory ones: exploration is assumed less likely to succeed but to carry higher upside conditional on success. Our notion of chosen ambition generalizes Manso's binary contrast along a continuum. The scientist is not choosing between two arms of a bandit; she is choosing a project from a much larger menu, and the dimensions along which projects differ include variance, uncertainty about probabilities, and uncertainty about the state space itself. Manso models one slice of this broader choice; our definition covers the others as well.

2.1 Why peer review may penalize chosen ambition

How much chosen ambition a scientist is willing to accept is shaped by many institutions, of which peer review is only one. The tolerance for early failure of those who fund the work (Manso 2011, Azoulay et al. 2011, Tian and Wang 2014), the time horizons over which contributions are evaluated (Lerner and Wulf 2007), the structure of the team itself (Wu et al. 2019), and the financial incentives tied to successful outcomes (Graff Zivin and Lyons 2021) all shape how much ambition researchers are willing to bear when selecting projects. Peer review enters this landscape as the institution that decides whether any given proposal will receive the resources required to carry it out. In biomedical research, an investigator who cannot sustain a flow of grant funding soon loses her laboratory, and the NIH R01 is the instrument through which that flow is negotiated—a position that gives peer review a degree of influence over project selection that few other institutions can match.

Whether peer review exercises that influence in favor of or against chosen ambition is not obvious on theoretical grounds. A reasonable starting point is that funders, by supporting a portfolio of projects with largely independent probabilities of success, are better positioned to absorb risk than the individual scientists they fund (Arrow and Lind 1970). The standard economic logic would suggest that the risk tolerance of the funder should exceed that of the researcher, and the rhetoric of funding agencies—which emphasizes their appetite for transformative, "high risk, high reward" projects—is consistent with that logic. But there are reasons, both empirical and mechanical, to doubt that peer review delivers on the rhetoric in practice (Azoulay and Li 2022, Franzoni et al. 2022). We discuss four, one at each level of the process: the investigator, the individual reviewer, the committee as a deliberative body, and the funding institution as a whole.

The investigator level: anticipatory self-censorship. Much of what peer review does to chosen ambition may happen before a proposal is ever read. Academic laboratories depend on external grants to fund their operations, and the consequences of failing to secure funding are severe: closure of the lab, dismissal of dependent personnel, and—in many cases—loss of the investigator's own job and career. Even a short gap is costly: in laboratories supported by a single R01 grant, an interruption of 30 days or more in the funding of its renewal raises the share of personnel who are not employed by 3 percentage points, or 40 percent, with the largest effects among trainees (Tham et al. 2026). To the extent that investigators *perceive* peer review to punish ambition, they will rationally adjust what they propose, advancing safer and more incremental projects than they would in the absence of the funding constraint (Langer 2012). The intense competition for funds that characterizes the current NIH landscape, in which most proposals are

rejected, compounds this pressure by pushing investigators toward projects with relatively quick and predictable payoffs (Alberts et al. 2014, Fang and Casadevall 2015). Whether or not peer reviewers actively penalize ambition at the evaluation stage, the *anticipation* that they might is sufficient to filter out the boldest proposals before they are submitted.

The reviewer level: evaluative bias under uncertainty. The second channel operates on the proposals that are actually submitted. Psychologists have long documented that individuals exhibit ambiguity aversion—a preference for outcomes whose probabilities are known over those whose probabilities are not—even when the expected value of the latter is higher (Fox and Tversky 1995, Mueller et al. 2012). Peer reviewers are individuals, and the proposals on which chosen ambition is highest are precisely those whose probabilities of success are hardest to estimate. Empirical evidence on the treatment of novelty in peer review is consistent with this mechanism: evaluators systematically assign lower scores to proposals that are more novel (Boudreau et al. 2016) and place heavy weight on feasibility at the expense of potential impact (Krieger et al. 2023, Lane et al. 2026). This channel is likely to bite hardest on proposals whose realizations involve *radical* uncertainty rather than narrow risk—projects whose prospects cannot be assessed by familiar rules of thumb because the state space of possible outcomes is not yet well mapped.

The committee level: deliberation and the weight of dissent. Even if individual reviewers were perfectly calibrated, committees convert disagreement into a negative collective signal. Two forces make this so. First, negative information is weighted more heavily in human judgment than equally strong positive information: "bad is stronger than good" is one of the more robust findings in social psychology (Baumeister et al. 2001). The effect has been documented experimentally in grant review itself: when evaluators are exposed to peers' scores before finalizing their own, negative peer scores trigger a downward revision while positive peer scores produce no symmetric upward movement (Lane et al. 2022). Second, mean-score aggregation near a strict funding payline is mechanically sensitive to downward outliers, because scoring scales are bounded above. A proposal that inspires *disagreement*—half the room excited, half unconvinced—therefore fares worse than one that inspires uniform mild approval, even when the first carries the higher expected value. The dynamic is vividly visible in adjacent gatekeeping settings: in editorial peer review at a general medical journal, manuscripts with uniformly negative reviews are rejected 88% of the time, those with uniformly positive reviews 20%, but *split* reviews 70%—nowhere near the midpoint (Kravitz et al. 2010). Chosen ambition is likely to provoke such splits: in venture competitions, the more distinctive a startup's proposition, the more its judges disagree (Gius 2025). A rule that ranks proposals by their mean score then penalizes chosen ambition as a side effect, independent of any individual reviewer's preferences (Carson et al. 2023). That institutional workarounds have

been developed—"golden tickets," which let a single reviewer advance a proposal that the rest of the panel would reject (Wang 2026), and NIH's Transformative Research Awards—is itself evidence that the underlying aggregation problem is widely recognized, even where it has not been fully solved.

The institutional level: political economy and the demand for accumulated success. Beyond the mechanics of deliberation, funding agencies themselves face political pressures that shape the portfolios their committees are encouraged to assemble. Public funders are accountable to political constituencies that demand evidence of return on investment, and "return" in this context is typically measured in publications, citations, and tangible achievements rather than in long-run knowledge spillovers (Clancy 2023). An agency that supports a portfolio of boldly ambitious projects—many of which will fail, by construction—must defend that portfolio to observers who may not appreciate that failure is the cost of ambition rather than its opposite. The safer equilibrium, from the agency's standpoint, is to support projects whose success is more predictable, even at the cost of lower expected long-run value.

A countervailing force operates on the supply side. Scientists may be systematically overconfident about their projects' probability of success or the value of their potential results (Gan et al. 2012, Sanchez and Dunning 2023), in which case some degree of conservatism in the evaluation process may be socially appropriate—a corrective for proposer optimism rather than a distortion. This is an instance of a broader question that the conceptual framework alone cannot resolve: even if peer review penalizes chosen ambition, it does not follow that the penalty is excessive. We return to this question in Sections 5 and 6.

These four channels are not mutually exclusive. They are, if anything, mutually reinforcing: self-censorship by investigators shapes the pool of proposals reviewers see; ambiguity aversion by reviewers shapes the scores the committee aggregates; dissent-sensitive aggregation shapes which proposals survive committee deliberation; and institutional pressures shape the incentives of every level below them in subsequent cycles. What they share is a common tendency to favor projects whose outcome distributions are tighter, whose probabilities are more legible, and whose state spaces are better mapped.

2.2 Operationalizing chosen ambition

A definition of scientific risk taking as broad as the one we have adopted generates an immediate problem of measurement. If chosen ambition encompasses variance, uncertainty, and radical uncertainty—and if the most ambitious projects typically involve some blend of all three—then

no single empirical proxy can capture the construct in its entirety, and an approach that relies on a single measure would be vulnerable to the charge that the finding reflects the peculiarities of the proxy rather than the phenomenon. The strategy we adopt is one of deliberate triangulation: we construct four distinct measures of chosen ambition, each partial and each capturing a different source of information inadequacy, and we take the *pattern* of findings across them—rather than the numerical estimate from any one of them—as the object of empirical interest.

Pich et al. (2002) treat projects as payoff functions defined over a state space F and a causal mapping M from actions to outcomes. In their account, a project team may face three distinct forms of information inadequacy. It faces *uncertainty* when F and M are both known but outcomes are stochastic—known probabilities over known states. It faces *ambiguity* when the state space F is itself incompletely known (Pich et al. use the phrase "unknown unknowns"), so that the team cannot enumerate the events that might bear on success. And it faces *complexity* when the causal mapping M is too richly interconnected for the team to compute the optimal action even given full information about F. Pich et al. take the project as given and ask how a manager should cope with whichever form of information inadequacy she has been handed. No one hands our scientist anything: she chooses the information inadequacy she will face when she chooses her project. Risk taking, in our usage, is the act of selecting a project whose (F, M) profile is thinner than the alternatives—whether the thinness comes from variance around known outcomes, from unknown probabilities, or from an incompletely specified state space. This is a modest extension of the Pich et al. framework, and it provides the vocabulary we need to describe what each of our four measures is attempting to capture.

A few historical illustrations may help fix intuitions. The effort to synthesize insulin using recombinant DNA techniques in the late 1970s was a commitment with a relatively well-characterized F: the scientific community knew what success would look like and had identified the technical obstacles in advance. What it involved was substantial uncertainty around probabilities, and intense competitive pressure (Hall 1987). Early work on cancer immunotherapy, by contrast, faced mostly ambiguity: for decades it was not even clear whether therapies that selectively targeted cancer cells could in principle exist, and the space of possible successful treatments was not well defined until late in the research program (Littman 2015). Ignaz Semmelweis showed in the late 1840s that physicians who disinfected their hands prevented deaths from childbed fever in a Vienna maternity clinic. His work faced ambiguity of a third kind altogether: the science was straightforward, but the medical establishment rejected his findings for reasons that had little to do with their empirical validity (Nuland 2003). Each of these examples involves a scientist

committing effort to a project whose information profile was thinner, in a different respect, than the alternatives she could have pursued.

Each of our four measures is built from the publications that acknowledge the grant during its cycle: their topics, their distance from the investigator's earlier work and from the funded projects of her study section, and the citations they go on to receive. With respect to the renewal decision, this record is *ex ante*: it is the evidence of progress that reviewers have before them, and work about to appear typically figures in the renewal application itself. With respect to the investigator's commitment, by contrast, no one outside NIH observes the proposal she wrote, and we infer her choice from the work it produced. Our claim therefore concerns the risk that reviewers can see in the record when they decide, and Section 3 states what this interpretation requires, along with the technical details of each measure.

Extreme tail outcomes captures the acceptance of a project whose range of potential impact is wide. Grants whose publications include both highly cited and completely ignored ones score high on this measure; grants whose publications cluster tightly around their mean score low. Of our four measures, this one comes closest to the classical variance notion of risk and to the *uncertainty* rung of the Pich et al. framework: the distribution of outcomes is characterized by high variance around a roughly known mean, and the investigator's willingness to accept that variance—legible to reviewers at the time of decision through signals such as "big if true" or "interesting but not clear it will work out"—is what we are attempting to estimate.

Disruption measures the extent to which a grant's publications redirect the citation network of their field, building on the disruptiveness index of Funk and Owen-Smith (2017). A highly disruptive paper is one whose subsequent citations do not also cite its predecessors—that is, one whose readers treat it as a break from prior work rather than as a continuation. Read through our framework, a disruptive grant is one whose investigator accepted a bet on a revised causal mapping M: the field's prevailing account of how causes and effects fit together is being challenged, and the grant cycle's outputs either vindicate the bet or they do not. Disruption thus straddles the uncertainty and ambiguity rungs: it captures variance around a known outcome space when the bet succeeds in a recognizable way, and revision of the outcome space itself when it succeeds dramatically. The citations behind extreme tail outcomes and disruption are the exception to the *ex ante* character of our record, since they accrue after publication and in part after the decision; we use them to stand for the likely impact of the work, which reviewers can estimate, if noisily, from the publications themselves.

Pivoting measures the distance between the content of a grant cycle's publications and the content of the investigator's own prior work. A high-pivoting grant is one in which the investigator has moved substantially from her established area of expertise into a topic where her own prior M—her accumulated knowledge of what works, what doesn't, and what to expect—is least reliable. Unlike the previous two measures, which concern information inadequacy in the field, pivoting concerns the investigator herself. A scientist can pivot into a topic that the field collectively understands well and still be taking on substantial ambition at the individual level, because she is operating outside the zone where her own intuitions are well-calibrated.

Standing out measures the distance between the investigator's portfolio of work and the portfolios of her peer competitors for funding—the other principal investigators funded by the same institute after review by the same study section. This measure requires a more careful defense than the others, because at first glance it may appear to be a relabeling of novelty: the same distance metric that is used elsewhere in the literature to measure the originality of a paper relative to the prior knowledge base is used here to measure the distinctiveness of an investigator relative to her funding cohort. The formula is indeed similar. But the construct it is estimating is not. Novelty asks a question about intellectual content: how far does the work depart from the prior literature? Standing out asks a question about *institutional fit*: how poorly does the investigator's work map onto the administrative partition of the knowledge space that NIH uses to organize its peer review? An investigator can work on an intellectually well-established topic and still stand out, if her work straddles multiple study sections' jurisdictions or does not map cleanly onto any single committee's expertise. Conversely, she can work on a highly novel topic and not stand out, if that topic happens to fit neatly into the coverage of a particular study section. What standing out measures is the investigator's acceptance, at the moment of commitment, that her proposal will face reviewers whose frame of reference is not her own, a form of chosen ambition distinct from the intellectual novelty of the underlying work. The empirical case for this distinction—that standing out and novelty are only weakly correlated in our estimation sample, and that they carry opposite signs in their relationship to grant renewal—is presented in Sections 4 and 5.

Our definition also tells us which proxies to accept as measures of risk taking. Almost any proxy that compares a grant's work with a reference will single out some work as different, and a difference is not risk by itself. As a test, the definition asks three things of a proxy: that it describe the project the investigator committed to; that it compare that project with a reference that can stand for the alternatives on her menu, which we never observe; and that the difference bear on how well charted the project's outcomes were, not only on its content. Pivoting compares

the work with the line of research the investigator had been pursuing, standing out with the other projects funded through her study section, and disruption with the literature on which the work builds, which she could have consolidated instead. Extreme tail outcomes has no such reference: its only comparison is with other grant cycles. It answers instead to the first term of our definition, a distribution of outcomes that is wider, and of our four measures it corresponds most directly to the everyday meaning of risk: lines of work that end in success beside lines that end in failure. The same test excludes three other proxies. The novelty of the work compares it with all of prior science, which is not the investigator's menu; a result can be new to science and well charted for the scientist who produced it. The distance between the investigator and her collaborators, and the breadth of her earlier work, describe the people who carry out the project, not the project. We treat them as conditions under which risk is taken and ask whether the penalty differs with them (Appendices I and J); Appendix F examines novelty.

Two concessions follow. The first concerns *completeness*. We do not claim that these four measures exhaust the construct of chosen ambition, and we would be surprised if any finite set of observational proxies could. The Pich et al. framework itself implies that ambiguity—information inadequacy concerning the state space—is only partially observable: unknown unknowns cannot be measured directly, only inferred from what the project turned up. Our measures miss some forms of risk taking that our definition admits. Franzoni and Stephan (2023) distinguish four kinds of risk in science: that nature answers the question unfavorably, that the planned experiments cannot be carried out, that a competitor gets there first, and that finished work is poorly received. We measure the first and the last in part, the second only through the failures that reach print, and the third not at all. Nor can a measure built from publications record a bet that produced none: one grant cycle in six has no acknowledging publication, and therefore no measure (Appendix A). What we claim is coverage along four conceptually distinguishable dimensions, not saturation across the construct as a whole.

The second concerns *purity*. None of our four measures captures a single rung of the information-adequacy ladder to the exclusion of the others. Extreme tail outcomes correspond most naturally to the uncertainty rung but also register some ambiguity. Disruption straddles uncertainty and ambiguity by design. Pivoting and standing out trace information inadequacy to different sources—the investigator and the reviewing institution, respectively—but both involve blends in their content. We describe each measure by its *dominant* rung rather than by a clean one-to-one correspondence, and we do not claim that the four together provide an orthogonal decomposition of the construct.

What we do claim is that this set of four measures, taken together, permits a form of triangulation that no single measure could deliver on its own. If peer review penalizes one particular form of chosen ambition—narrow variance-risk, say—and is neutral toward the others, the penalty should show up on one measure and be absent on the others. If, instead, peer review penalizes chosen ambition across the board, the penalty should appear on all four. The *pattern* across the four measures is therefore informative in a way the numerical estimate from any one of them would not be. Whether that pattern appears in the data is an empirical question, to which we turn.

3 Setting, data, and measurement

3.1 The National Institutes of Health

The NIH plays a prominent role in funding biomedical research. In 2012, it comprised 27% of all biomedical research funding in the United States and nearly two-thirds of all public and philanthropic research funding (Moses et al. 2015). In 2021, the NIH spent $32 billion to support extramural research projects (Lauer 2022). While the NIH funds clinical research (including clinical trials, health services research, and behavioral studies), the bulk of their funding supports basic research aimed at advancing the frontier of scientific understanding. With research project grants having an average size of over $580,000 in 2021, NIH grants are a major source of funding for the researchers they support. Not only does this funding play an important role in supporting the purchase of materials and the salaries of lab personnel, many academic researchers are in "soft money" positions, funding their own salary out of awarded grants. The ability to successfully raise grant funding is an important metric upon which academic employment and promotion decisions are made.

Understanding the process the NIH uses to evaluate and fund grant applications can provide useful context for our empirical analysis. Briefly, the NIH is comprised of 27 separate institutes or centers (IC) which are typically organized around a set of diseases (e.g., National Cancer Institute) or organ systems (e.g., National Heart, Lung and Blood Institute). Each IC is responsible for funding research that is potentially relevant to its mission. Scientific evaluation of grant proposals occurs primarily in standing review committees known as study sections (e.g., "Cellular and Molecular Immunology" or "Atherosclerosis and Vascular Inflammation"). Each study section may review and score grants from multiple ICs, which are then funded by the ICs in order of their

study section score until the IC exhausts its funds.[2] The amount of resources devoted to reviewing grant proposals—both from the NIH with staff to manage the process and from study section members—is substantial. In 2021, the NIH reviewed nearly thirty-eight thousand applications for R01-equivalent grants (Lauer 2022).

The NIH is an attractive setting for studying how peer review might respond to risk taking. First, professional norms lead most academic research to be published in peer-reviewed journals. When combined with rich bibliometrics in biomedicine, this provides a robust "paper trail" we can follow and characterize to measure the degree of grant risk taking. Second, as NIH research is a source of significant improvements in health and positive economic spillovers, understanding how the NIH may more effectively utilize its funds to support breakthrough innovation is of considerable policy interest (Azoulay et al. 2019, Fleming et al. 2019).

3.2 Data

As we do not observe grant applications, we focus on whether an awarded grant was renewed. We rely on the NIH's Consolidated Grant Applicant File to identify all those researchers who received an NIH grant, 1980-2015. Our sample consists of grants first funded from 1980 to 2015, and of the cycles of those grants that started by 2015. We start in 1980 because we cannot reliably link publications to the NIH grants that supported them before that date. We stop with the cycles that started in 2015 so that we observe whether every cycle was renewed: the last of them ended in 2022. We limit our analysis to R01-equivalent grants.[3] These investigator-initiated grants are a major funding mechanism to support the work of principal investigators who have achieved career independence and run their own laboratory. We also exclude grant awards from special emphasis study sections, which may not have an option for grant renewal. Every one of our measures is computed from the publications that acknowledge the grant during the cycle, so a cycle that produced no such publication has no measure and is excluded from every estimation sample. One cycle in six (15.9 percent) falls into this category, and these cycles are renewed far less often than the others (23.3 against 41.9 percent). They do not come from investigators with safer records (Appendix A), and our estimates describe the penalty among cycles that produced at least one acknowledging publication. We focus our analysis on the grant cycle, which corresponds to a single competitive review of a funding proposal. A cycle starts the year a grant proposal is

[2] Institute directors do have discretion to fund applicants out of order considering factors such as fit with the institute's mission, the overall portfolio of projects being supported, and their own evaluation of application quality.

[3] We consider R01-equivalent grants to be those with an NIH activity code of R01, R37, R29, or R23. Appendix A lists the mechanisms that NIH also counts as R01-equivalent and that we exclude, with the reasons.

funded and ends either at the conclusion of grant funding or when a proposal for grant renewal is competitively reviewed and funded. Investigator characteristics are explicitly considered during renewal decisions, and our measures and controls describe one investigator. We therefore drop the cycles whose starting award names several principal investigators, and those whose principal investigator changes before the cycle ends; the outcome is whether the grant is renewed, whoever leads it next (Appendix A). This gives a final sample of 103,484 grant cycles from 69,381 grants and 40,088 investigators.

Our publication data is from *PubMed*, an online resource from the National Library of Medicine to support access to the biomedical research literature that indexes over 32,000 journals and over 39 million articles published through 2025. The citations these articles make and receive are from OpenAlex. We limit our analysis to original research publications; this excludes other types of publications such as review articles, letters, and editorials. For each grantee, we identify their career publications in *PubMed* using Author-ity (Torvik and Smalheiser 2021, 2009). Author-ity disambiguates authors using the insight that authors of the same name sharing coauthors, affiliations, keywords, and other such characteristics are likely, in fact, to be the same person. Author-ity has been shown to be highly accurate (Appendix A).

We use NIH RePORTER to link publications attributable to each NIH grant. This data relies on the grant having been listed in the publication's acknowledgement section as a funding source, which is required as a condition of accepting NIH funding. We consider publications attributable to a grant cycle if they acknowledge the grant and appeared between the first and the last funded year of the cycle.[4] As investigators may have different degrees of risk taking across different projects and funding sources, in all cases, our measures of risk taking are calculated solely using publications during a grant cycle that acknowledge funding from the grant rather than also including publications which do not acknowledge funding from the grant.

To measure research topic and methodology, we capitalize on Medical Subject Headings (MeSH). MeSH is a hierarchical controlled vocabulary created and maintained by the National Library of Medicine used for indexing and searching the biomedical literature. In the 2026 edition of MeSH, there are over 31,000 terms, with an average of about 10 MeSH keywords associated with each *PubMed* publication indexed with MeSH. Crucially, MeSH are assigned by professional indexers. While it is difficult to assess accuracy as there is no single "correct" way to index articles, MeSH have generally been found to be consistent across indexers (Funk and Reid 1983).

[4]Publications that appear after the end of the cycle were not available to the study section when it reviewed the renewal, so they inform neither our outcome nor our measures; when the grant is renewed, they belong to its next cycle.

In addition to MeSH associated with grant publications, we use the National Library of Medicine Medical Text Indexer to assign MeSH to grant abstracts. Grant abstract text is available from the NIH RePORTER.

3.3 Measuring risk taking

A key consideration in our measurement strategy is that we construct our risk measures from the publications that acknowledge the grant during the cycle. These publications are the record that study sections evaluate at renewal, and our claim concerns the risk visible in that record, not the risk of the renewal application, which we do not observe. NIH instructs study sections to consider the progress made in the last funding period, and much of that progress is reported in these publications. The record is also what an investigator anticipates when she decides how ambitious to be, so that a penalty at renewal weighs on her decision to take risk in the first place. The interpretation therefore does not require that risk taking stay constant over a cycle, nor that the work proposed for the next cycle be as risky. It requires only that risk taking not change during the cycle in a way that tracks the prospect of renewal. Appendix G examines this requirement. The association between risk and renewal is already present in the publications of a cycle's first two years, and risk measured on the investigator's publications from before the cycle began, which nothing during the cycle can have shaped, is also associated with lower renewal. A second requirement would remain even if the applications were available: what our measures record must be correlated with what reviewers perceive when they evaluate the renewal. We cannot observe that perception; the requirement is only that our measures correlate with features of the work that reviewers do weigh, such as its feasibility or its distance from the field's mainstream.

First, risk taking may lead to greater variance in outcomes, increasing both the likelihood of a breakthrough and the likelihood of failure (Azoulay et al. 2011). We capture this dynamic with a measure we term "extreme tail outcomes," defined as the difference between the highest and lowest vintage-adjusted citation percentiles among publications from the grant cycle. We measure the standing of each publication by the citations it receives in the year it appeared and the five that follow, ranked against all *PubMed* publications of the same year. The window keeps the measure close to the record that study sections evaluate: for a publication that appears late in the cycle, most of these citations are made after the renewal decision, so that the percentile stands for the reviewers' reading of the work more than for its eventual impact, which for unconventional work can take longer to be recognized (Wang et al. 2017). Additionally, using percentiles enables meaningful comparisons across grants awarded in different time periods, ensuring that differences

in citation practices over time do not distort our measure.[5] Appendix B decomposes the measure. Its association with renewal comes mostly from the cycle's weakest publications, not from its best one. A weak publication can be a minor piece of a project that did well or what remains of a line of work that failed, and each accounts for about half of the measure's lower tail.

Research can be either consolidative, reinforcing earlier findings, or disruptive, challenging the status quo. Our second measure leverages this distinction with the disruption index proposed by Funk and Owen-Smith (2017), which Wu et al. (2019) brought to scientific publications and which Park et al. (2023) used to document a decline in the disruptiveness of papers and patents. This index classifies a paper as highly disruptive if future research cites it while largely ignoring its references, whereas highly consolidative research is characterized by future studies citing both the focal paper and its references. Our measure of "disruption" for a grant cycle is the highest percentile of the disruption index for the publications emerging out of the initial round of funding (calculated relative to all publications appearing in *PubMed* in the same year). Appendix C explains how the index is computed, what a high value does and does not mean, and how the association between disruption and renewal fares under alternative versions of the measure. In every version of the measure we examine, more disruptive cycles are less likely to be renewed.

As a third measure, we examine the intellectual distance between a grant cycle's publications and the investigator's prior work. This approach recognizes that some aspects of risk stem not just from the project itself but also from who is conducting the research. Scientists who integrate their existing expertise with new knowledge—whether by addressing novel topics, adopting new techniques, or working with unfamiliar model organisms—are taking on risk. Such shifts may yield valuable discoveries but can also devalue prior expertise and invite skepticism from peer evaluators (Arts and Fleming 2018, Hill et al. 2025). We term this measure "pivoting." To compute it, we calculate the fraction of MeSH term pairs in a grant cycle's publications that are new relative to those in the investigator's prior work during the preceding five years: her original research publications in English, in any authorship position.[6] Appendix D dissects the measure: new pairs that recombine descriptors the investigator already uses are not penalized, while pivots into unfamiliar territory are, the more so the more distant the new descriptors are from her earlier work. This is the gradient that Hill et al. (2025) document for the impact of research: the further scientists move from their prior work, the lower the impact of what they produce.

[5] A limitation of this approach is that we infer failure from exceptionally low citation rates. While this is a reasonable proxy, truly failed projects are more likely to remain in a file drawer rather than being published in a minor outlet.

[6] Counting every authorship position is the conservative choice for this measure: each earlier publication adds pairs to the investigator's record and makes a pivot harder to register.

Our final measure of risk taking captures intellectual distance from what other NIH-funded investigators are studying within the same scientific domain. This measure reflects the risks of pursuing contrarian research—projects that diverge from the mainstream of one's peers—thereby increasing career risk (Fang and Casadevall 2015). We term this measure "standing out" and operationalize it as the fraction of MeSH term pairs in a grant cycle's publications that appear exclusively in the focal investigator's work, meaning no other funded researchers in the same IC-study section-year use these term pairs in their grant proposal abstracts.[7,8] To check that the measure does not simply flag specialized work, which would stand out against any study section, we compute it against five other study sections of the same year, drawn at random: against them, standing out is not penalized; the penalty appears only against the investigator's own study section, and it comes from combinations of topics that science had used for years and that the funded projects of her study section did not combine (Appendix E).[9]

3.4 Measuring novelty

To empirically distinguish novelty from risk taking, we construct a separate measure of novelty. Prior research has taken various approaches, including analyzing unusual or intellectually distant combinations of references (Uzzi et al. 2013, Wang et al. 2017), uncommon combinations of patent classes (Fleming 2001), citation network centrality (Shibayama and Wang 2020), and the age of keywords or keyword combinations (Arts et al. 2025, Packalen and Bhattacharya 2020). Other studies have used domain-specific measures, such as novelty in chemical structures (Krieger

[7]The relevant sample size varies across each of these measures, from 83,052 cycles for standing out to 101,435 for pivoting (Table 2). Extreme tail outcomes is only defined for those grant cycles resulting in at least two publications with a citation percentile. Disruption is only defined for those grant cycles resulting in at least one publication with at least one reference and at least one forward citation in *PubMed*. Pivoting is defined only if there is at least one publication with MeSH terms in both the current grant cycle and the preceding five years. Standing out is defined only if the publications of the current cycle yield more than five MeSH term pairs and sufficient funded grant proposal abstracts in the same IC-study section-year are available.

[8]We compare an investigator's grant cycle publication MeSH to that for all available awarded grant proposal abstracts, including those that are not R01-equivalent grants, in the same IC-study section-year. The comparison group holds 159 abstracts at the median for the cycles evaluated in the 1980s, and 108 for those evaluated in the 2010s. We limit our measure to those grant cycles with at least 20 other abstracts in the same IC-study section-year, a condition that excludes about a fifth of the cycles (Appendix E). We prefer this approach to alternatives, as the comparison group is thicker for those with sufficient available data, facilitating our focus on the tail-measures of intellectual distance from other investigators. Similar results are obtained using other thresholds for the number of abstracts in the same IC-study section-year.

[9]Pivoting and standing out are both built from MeSH descriptors. Distances between embeddings of the publications' text, computed with a language model, are a natural alternative for both. The choice of representation matters, however: for patents, Ganguli et al. (2024) find that a representation chosen without validation against alternatives can leave a trend in similarity ambiguous, or even reverse it. Validating such representations for biomedical publications since 1980, whose early abstracts are often short, is a task for future work.

et al. 2022, Rzhetsky et al. 2015). Our measure is the share of new noun phrases in a cycle's publications (Arts et al. 2025). A noun phrase is a noun together with the words that specify what it refers to ("vascular endothelial growth factor"); noun phrases name the objects, methods, and concepts of a field, and a new one often marks a new idea. Arts et al. extract the noun phrases of the titles and abstracts of some 75 million publications and record the first publication to use each. We count a phrase as new only if a later publication that shares none of its authors uses it as well (Appendix F). For each cycle, we divide the number of new noun phrases in the acknowledging publications by the number of their noun phrases. Because a phrase must be used again to count as new, the measure declines toward the end of the period, and we compare cycles evaluated in the same year. The measure is built without MeSH descriptors, so that it shares no raw material with pivoting and standing out, and it is nearly uncorrelated with both (Table 1d). Appendix F reports three alternatives, among them the share of new MeSH pairs of Boudreau et al. (2016).

4 Research design and a descriptive look at the data

We observe the universe of NIH grants actually awarded but unfortunately not the grant applications that NIH ultimately declines to fund. Because of this, we empirically focus on NIH R01-equivalent grants that were awarded and measure whether or not the grant was renewed. This approach lets us reliably observe all investigators who were at risk of having their grant renewed. In this population, a grant may not be renewed either because the investigator applied for but did not receive a competitive renewal, or because they never applied for a renewal. Given the career importance of grant funding in maintaining a lab, we expect most investigators will apply for renewal.[10] Furthermore, if an investigator with risky research is dissuaded from applying for grant renewal due to their perception of a low likelihood of renewal and the time involved in applying, the implications for our research question are very similar to having applied and not been selected. As our sample evaluates the degree of risk taking among scientists all of whom have an R01-equivalent grant, this suggests these scientists will have similarly benefited from the act of applying for and receiving an R01-equivalent grant (Ayoubi et al. 2019, Jacob and Lefgren 2011).

[10] Applications for competitive grant renewal, using the NIH's more expansive definition of R01-equivalency (see footnote 3) were funded at a rate of 34.5% in 2015 (Lauer 2016); the renewal rate in our sample in the same year, 27.1%, is about four fifths of that rate, consistent with the assumption that most grantees do apply for renewal.

Using this research design, the association between risk taking and renewal could reflect omitted variables: unmeasured factors correlated with both our proxies for risk taking and a grant's likelihood of renewal. To mitigate this concern, we incorporate a rich set of controls for investigator, institution, and grant cycle characteristics.[11] Every regression includes controls for the number and the quality of the publications produced during the cycle, so that it compares cycles whose records of progress are of the same size and of similar quality. Study sections evaluate that record when they decide on renewal, and renewed cycles produce more publications that acknowledge the grant than the others (Table 1b). Without these controls, two of our measures would largely count publications, since the spread of citation percentiles and the highest disruption index among a cycle's publications both grow with their number. The question we ask is whether, among cycles that produced as much, in journals of similar standing, those whose work was riskier are less likely to be renewed.

We know of no source of variation in risk taking that is unrelated to everything else that bears on renewal, and the obstacles to finding one are not peculiar to our study. NIH has never tested by experiment how it allocates its funds, despite calls to do so (Azoulay 2012), and the data that might stand in for an experiment are closed to researchers. We cannot see the applications that NIH declines to fund; of those it funds, we can read only the abstracts; and the distribution of reviewers' scores is destroyed to protect their confidentiality, so that not even NIH's staff can examine it.[12]

4.1 Methodological considerations

Our primary estimating equation relates the dependent variable of interest, $renewal_{ij,t(j)}$, for investigator i, grant j, and grant cycle $t(j)$ to risk taking assessed for the preceding cycle of the same grant j, $t(j)-1$. Formally, we estimate variations of:

$$E\left[renewal_{ij,t(j)} = 1 \,\middle|\, X_{ijt}\right] = \beta_0 + \beta_1\, risktaking_{ij,t(j)-1} + X_{ij,t(j)}$$

[11] The NIH instructs reviewers to score each application based on five criteria: significance, innovation, approach, investigator, and environment. While the first three assess the importance, novelty, and feasibility of the proposed project itself, the last two evaluate investigator qualifications and ability of the institution to support the work, respectively.

[12] Boudreau et al. (2016) and Lane et al. (2022) estimate causal effects by running experiments within grant competitions at a research university, at a price paid in scale and in stakes. Our estimates are conditional correlations, and we present them as such. The appendices test alternative interpretations of each measure and rule out several; they do not make our estimates causal.

where $renewal$ is an indicator variable equal to 1 if grant j was renewed for the next cycle $t(j)$, and 0 otherwise. We include high dimensional vectors of controls for investigator and grant cycle characteristics $X_{ij,t(j)}$. This vector includes sets of indicator variables for investigator age, degree and gender; log investigator career citations, log prior investigator NIH funding, an indicator variable for no prior investigator NIH funding, and log institutional NIH total funding as measures of investigator and institutional quality; and sets of indicator variables for year of evaluation, initial cycle length, and IC-study section fixed effects. It also includes the controls for research quality and quantity during cycle $t(j) - 1$: log of mean journal impact factor, sets of indicator variables for the number of publications during the grant cycle (separately for those acknowledging and not acknowledging funding from the grant), and indicator variables for having publications in highly selective journals.[13,14] We use robust standard errors with two-way clustering at the investigator and grant level throughout.

Our primary statistical model is a linear probability model (LPM). While LPM can produce fitted values outside the interval [0,1]—which makes it unsuitable for predictive purposes—our objective is to estimate the marginal effects from the conditional expectation function. In this context, the LPM is typically more appropriate (Angrist and Pischke 2009, pp. 94-106). Specifically, LPM offers a straightforward estimation of marginal effects and is computationally efficient when dealing with a high-dimensional vector of covariates and fixed effects, which is crucial for our analysis.

4.2 Descriptive statistics

Tables 1a and 1b present descriptive statistics for investigators and grant cycles, respectively. Reflecting the intense competition for R01-equivalent grants, the sample consists of highly accomplished scientists. On average, investigators begin the grant cycle 18.0 years after completing their terminal degree, with 54 prior publications and 1,777 career citations. Approximately 5% of investigators in the sample have achieved elite status, defined as having won a Nobel Prize, Lasker Award, Howard Hughes Medical Institute Investigatorship, or membership in the Institute of Medicine or National Academy of Sciences.

The majority of an investigator's research output is not directly funded by the focal R01-equivalent grant. During a cycle, investigators publish on average 6.8 papers that acknowledge

[13] We use journal impact factor as a proxy for publication quality in grant cycle publications, as this metric is visible to peer review committees at the time of evaluation, whereas early citations provide a noisy signal of quality.

[14] For this measure we include those publications which were published in *Nature*, *Cell*, *Science*, and the *New England Journal of Medicine*.

the grant and 11.8 that do not, and in the average cycle, 66% of the latter still list NIH support, either through other NIH grants to the investigator or to coauthors (Table 1b). Overall, 41.9% of grant cycles were renewed, with renewal rates higher for subsequent cycles (50.6%) compared to the first cycle (35.5%).[15] Renewed grant cycles were associated with younger investigators (17.1 vs. 18.6 years of career experience), greater research productivity during the cycle (8.0 vs. 6.0 publications and 242 vs. 160 five-year citations, limiting the comparison to articles directly acknowledging the focal grant), and fewer publications that do not acknowledge the focal grant (11.5 vs. 11.9). Figure 1 provides an histogram of the distribution of the number of publications per grant cycle, separately for renewed and terminated grants.

Table 1c presents descriptive statistics for the four measures of risk taking, as well as novelty. Without controls, the differences between renewed and other cycles are small: the median of each measure is slightly lower among renewed cycles, and the means are nearly equal. Figure 2 displays histograms for each of the four measures by grant cycle renewal status, providing initial evidence that renewal rates vary across different levels of risk taking. Finally, Table 1d presents correlations between measures of risk taking and novelty. Two pairs are correlated, each built from the same raw material: extreme tail outcomes and disruption (0.39), both computed from citations, and pivoting and standing out (0.24), both computed from MeSH descriptors. The other correlations are below 0.10 in absolute value, consistent with each measure capturing a distinct facet of risk taking. Novelty is only weakly correlated with any of them, between -0.16 and 0.08, consistent with our view that work new to science need not be risky work.

5 Results

5.1 Risk taking and grant renewal

Table 2 reports the association between each of the four measures of risk taking and the likelihood of grant renewal. For every measure, cycles in the top decile are renewed less often than those in the middle eight deciles, by 2.5 to 4.2 percentage points, or 6 to 10 percent of the mean renewal rate; cycles in the bottom decile are renewed more often, by 2.3 to 5.1 points, except for pivoting, whose bottom decile does not differ significantly from the middle.[16]

[15] In contrast, the success rate of applicants for a new R01-equivalent grant is lower. In 2015, 16.1% of applications for a new R01-equivalent grant were funded (Lauer 2016).

[16] For extreme tail outcomes, pivoting, and standing out, percentiles are calculated relative to other grant cycles in the same evaluation year. For disruption, we emphasize work that is disruptive in an absolute sense and so

We further investigate the full distribution of risk-taking levels and their association with grant renewal, as depicted in Figure 3. In these specifications, separate indicator variables are included for each ventile of the risk measure. Because they impose no functional form, these specifications show whether the penalty arises mainly in the tails of the distribution or across its whole range. Consistent with Table 2, we find higher renewal rates for lower levels of risk taking and lower renewal rates for higher levels. Notably, several measures show a nearly monotonic decrease in renewal rates as risk taking increases.

Because two pairs of measures are correlated (Table 1d), the penalty on one measure could be the penalty on its partner. Appendix Table H1 enters the four measures in the same regression, as continuous variables, on the 69,826 cycles for which all four are defined. Each keeps a negative association with renewal, significant at the 1 percent level, and holding the other three fixed reduces its coefficient by 15 to 40 percent. A one-standard-deviation increase in each measure is associated with a probability of renewal lower by 1.1 to 2.9 percentage points, against a mean renewal rate of 45.1 percent.

Additionally, we examine how the penalty varies across NIH's institutes and over time (Appendix K). The penalty is a feature of NIH peer review as a whole, and not a recent one. In separate estimates for each of the eleven largest institutes, all 44 coefficients are negative (Appendix Figure K1); estimated year by year, the coefficient of each measure is negative in at least 34 of the 38 evaluation years from 1982 to 2019, and no year has a positive coefficient that is statistically significant (Appendix Figure K2). The size of the penalty varies across institutes and over time, but the variation differs from one measure to the next, and it is not consistently related to the scarcity of money: only for disruption is the penalty significantly smaller where the institute renewed more of its other cycles in the same year (Appendix Table K1).

5.2 Novelty and risk taking

We next examine whether novelty confounds the relationship between risk taking and grant renewal. We estimate models similar to those shown in Figure 3, but with the addition of indicator variables for each ventile of the novelty measure. The results of this analysis are presented in Figure 4. First, novel work is not penalized the way risky work is: cycles more novel than the median are renewed 1.9 to 2.0 percentage points more often than the others (an increase of 4.2 to 4.6 percent relative to the mean renewal rate), and holding risk taking fixed reduces that difference by an eighth at most (Appendix Table F1). Second, holding novelty fixed leaves the

use the percentile of the most disruptive paper, calculated relative to all publications in *PubMed* appearing in the same year.

penalty on each measure of risk taking nearly unchanged, its coefficient changing by about 3 percent at most. Appendix F repeats the analysis with three alternative measures of novelty: depending on the measure, more novel work is renewed more often than less novel work, or about as often, and it is never penalized the way risky work is.[17]

5.3 Is the risk penalty excessive?

Whether the penalty is excessive depends on an assumption that our data cannot test: that riskier projects carry, on average, a higher expected return, so that a funder who declines them gives up value. Unlike in finance, the relationship is probably not systematic. There are celebrated cases in which high value and a high probability of success coexist: the Human Genome Project was very likely to succeed when it was launched, given earlier advances in sequencing, and its results were expected to matter greatly for human health and for later science. And some risky projects are simply poor bets. We therefore hold the assumption as a central tendency. Without it, the penalty is a fact about peer review; with it, the penalty is also a cost to society.

Even under this assumption, a planner may want some penalty on risk. If scientists are overconfident about their projects, some conservatism in review may correct for their optimism (Section 2.1), so the size of the penalty says little about whether it is excessive. Where the penalty is steepest says more. When prior beliefs about the value of a project are diffuse, penalizing risk is most likely to stifle valuable exploration, and it would seem perverse for a planner to discourage risk taking where a project's expected impact is most uncertain. Prior beliefs are least informative for new investigators, whose track record is short, and for novel work, where earlier knowledge is sparse. A planner would therefore penalize risk less in both groups. New investigators instead pay a steeper penalty on three of our four measures, by 26 to 110 percent (Figure 5 and Appendix Table I1),[18] while the penalty differs little with the investigator's other traits (Appendix I). For novel work, the estimates are of the same sign but less precise: among cycles whose work is more novel than the median, the penalty is steeper on each of our four measures, by 9 to 55 percent, although only the differences for disruption and standing out are statistically significant (Appendix Table F1). Under the assumption, a penalty that is no milder, and for new investigators steeper, where prior beliefs are least informative is difficult to reconcile with a reasonable model of optimal scientific risk taking.

[17] Under semantic distance, the most novel cycles are renewed somewhat less often than those in the middle of the distribution, by at most 5 percent of the mean rate, against 8 to 15 percent for the riskiest ventile of each measure of risk taking (Appendix Figure F2, Panel C).

[18] New investigators may need to show some degree of pivoting from the research done during the training phase of their career to establish independence and a distinct research identity from their mentors.

We ask what became of the work that study sections renewed and of the work that was not renewed (Table 3). We call a cycle terminated when it was not renewed and its grant was not revived later. The logic is that of Becker (1957). If reviewers hold risky work to a higher bar, the risky cycles that clear it should outperform the other renewed cycles, and the risky terminated cycles should outperform the other terminated cycles, since some of them would have cleared a common bar. If reviewers instead apply a common bar to work whose outcomes are more dispersed, the risky renewed cycles should again outperform the other renewed cycles, but the risky terminated cycles should do worse than the other terminated cycles, since more of them fall far short of the bar. The table compares the cycles above and below the median of each measure on outcomes recorded in the five years after the cycle's last payment, none of them built from the publications behind our measures (their descriptive statistics are presented in Table 1e, and their sources and construction in Appendix L). For renewed cycles, these five years include work funded by the renewal, so each comparison is made among renewed cycles or among terminated ones, never across them. These comparisons are of averages, not of the cycles at the margin of the decision, and averages also depend on how outcomes are distributed on each side of the bar, so the evidence they offer is circumstantial.

The pattern a higher bar predicts is clearest for pivoting. Cycles above the median of pivoting are followed by 17 to 25 percent more patents arising from the grant, whether or not the grant was renewed. Among renewed cycles, those above the median of disruption are also followed by more patents, and their later publications are cited more often by patents (Marx and Fuegi 2020, 2022) and by patents of greater combined value (Kogan et al. 2017); for first cycles, they are also cited more often by drug- and device-related patents (Durvasula et al. 2023, Frakes and Wasserman 2026). The pattern is not general, however: terminated cycles above the median of disruption are not followed by significantly more patents than the others, so the comparison cannot tell a higher bar from a common one. On the outcome closest to what study sections are asked to judge, the scientific influence of the work, the measures disagree: the publications that follow renewed cycles above the median of disruption are cited somewhat more often, but those that follow cycles above the median of extreme tail outcomes are cited less often, renewed or terminated. We read Table 3 as evidence that some of the work reviewers penalize, in particular work that is new to the investigator, is later put to valuable technological use, but not as evidence that the penalty on risk taking is excessive in general.

5.4 Robustness checks and ancillary results

We conduct several robustness checks to assess the sensitivity of our estimates to alternative assumptions and subsamples. First, Appendices B to E examine the components and variants of each measure. For extreme tail outcomes, these are the level and the spread of a cycle's citation percentiles, and its publications grouped into projects (Appendix B); for disruption, eleven versions of the index, versions that hold fixed the citations or the references of the cycle's publications, and other summaries of those publications (Appendix C); for pivoting, near and far pivots, the distance of the new descriptors in the MeSH tree, pivots out of fading fields, and pivots proposed in the funded application (Appendix D); and for standing out, its unique pairs divided by whether they appear on the investigator's earlier publications and by their age in *PubMed*, and standing out computed against placebo study sections (Appendix E). Pivots out of fields that had been fading before the cycle began are penalized as much as other pivots, and so are pivots that the investigator proposed in the application that won the grant; what is penalized is the move into descriptors unfamiliar to the investigator (Appendix D). As risk-taking measures may be undefined for some grant cycles, the sample used for analysis varies across each risk-taking measure (see footnote 7 for details). Our results are robust to using a subsample for which all four risk-taking measures are defined for all grant cycles (Appendix Table A3).

Publications are harder to attribute correctly when an investigator's name is common. We therefore split the sample at the median frequency of the investigator's last name in *PubMed*: the penalty is similar in both halves (Appendix Table A3, columns 3 and 4). Investigators whose research lies mostly outside biomedicine may not seek renewal of an NIH grant at all, and their exit could look like a penalty. Every measure is penalized both among investigators whose share of publications acknowledging NIH funding is below the median and among those above it (Appendix Table A4).

Finally, Appendix Table H2 reports the test of Oster (2019), which extrapolates from the controls we observe to those we do not. Adding the controls of Table 2 takes every coefficient further from zero, so that Oster's δ is negative for all four measures, between -2.9 and -30.1: unobserved confounders resembling the observed controls would enlarge the penalty. The test rests on two assumptions that cannot be checked, which Appendix H discusses.

6 Discussion

Using a comprehensive dataset of awarded NIH R01-equivalent grants and multiple measures of risk taking, we find that, when a grant comes up for renewal, ambition counts against it: on each of our four measures, the cycles whose work took the greater risks are renewed less often. This provides systematic evidence that NIH peer review penalizes risk taking. We also show that novelty functions differently from our risk-taking measures, cautioning against the practice of using the former as a proxy for the latter. Given the NIH's central role in U.S. biomedical research funding, these findings have significant policy implications for peer review and innovation funding.

The distinction between novelty and risk taking has a practical side. Since 1997, NIH has asked reviewers to score the innovation of each application (National Institutes of Health 1997), and in our data the novelty of a cycle's work is associated with a higher probability of renewal, or with none, depending on the measure. What reviewers appear to discount is work whose outcomes are less well charted, for the field, for the investigator, or for her study section. If so, instructions or programs that reward novelty would not by themselves reduce the penalty on risk taking: that would require a change in how reviewers weigh the prospect of failure, a different matter from their appetite for new ideas.

Our estimates are conditional correlations about a single decision, the competitive renewal of grants that were funded, and they apply to new applications or to other funders only if the mechanisms of Section 2.1 operate there as well. Without the applications that NIH turned down, we cannot sign the bias that comes from studying only grants that won their initial award, since it depends on how the competition for that award treats risk (Appendix A). If, as we suspect, that competition also penalizes risk, the risky grants we observe had to clear a higher bar to be funded and may be stronger in ways our controls do not capture; the penalty we estimate would then *understate* the one we would find if we could observe every application, funded or not. A renewal decision also cannot show what happens before a proposal is written. Investigators aware of the NIH's risk aversion may self-select out of the funding process altogether, particularly for their high-risk projects (Langer 2012). Finally, young researchers training under NIH-funded mentors may internalize a preference for safer projects, reinforcing long-term conservatism in research agendas.

Each of our risk-taking measures admits alternative readings. A pivot may be a sound retreat from a field that is drying up rather than a bet on a new one; the spread of a cycle's citations

may reflect how its results were divided into publications as much as the risk of the research; disruption may owe something to how citations accumulate after the fact; and work that few others in a study section pursue may be a niche rather than a bet. Appendices B to E take each measure apart with these objections in mind. Each measure, in short, invites a quibble. Yet despite their idiosyncratic flaws and the fact that they are constructed in distinct ways—sometimes even drawing on different underlying data—the results remain remarkably consistent across measures. This consistency is reassuring: if each metric is flawed in its own way, much like Tolstoy's unhappy families, they are at least flawed differently. And yet, they tell the same story.

Both high-risk, high-reward and low-risk, incremental research play vital roles in scientific progress. If riskier projects carry a higher expected return on average, there are reasons to believe that NIH should support more risky research than it does. Established research fields naturally lend themselves to incremental work, while emerging areas require greater exploration and risk taking. Yet the penalty does not ease for more novel work: among cycles whose work is more novel than the median, it is steeper on each of our four measures, though significantly so only for disruption and standing out. If the value of exploration is highest there, as we suspect, the penalty is a potential source of welfare loss, and Table 3 shows that some of the penalized work, in particular work that is new to the investigator, is later put to valuable technological use. NIH risk aversion may also disproportionately affect research areas with limited private-sector incentives, where public support is most essential (Azoulay and Li 2022). Governments are uniquely positioned to bear risks too large for private industry, especially given declining corporate investment in science (Arora et al. 2018, Fleming et al. 2019). NIH's size and influence mean its hesitancy to fund risky fundamental research is unlikely to be offset by other funding sources, potentially slowing technological progress (Babina et al. 2023). The same logic applies to NIH-funded clinical research, where industry has limited incentives to invest in high-risk studies with uncertain commercial returns (Greenblatt et al. 2023). While private philanthropies provide alternative funding, their priorities may not align with the greatest public needs (Murray 2013).

Recognizing these challenges, the NIH has created dedicated high-risk, high-reward funding mechanisms, such as the Pioneer Awards, Director's Transformative R01 Awards, and New Innovator Awards. From fiscal year 2011 to 2016, these programs and the Early Independence Award made 514 new awards, against 22,559 new R01 awards (National Institutes of Health 2019). Several private science funders—including the Howard Hughes Medical Institute, Open Philanthropy Project, Fast Grants, Chan Zuckerberg Initiative, and Wellcome Trust—have also launched initiatives to support high-risk research. Whether such programs encourage risk taking

depends on their design more than on their names. A Pioneer Award provides $700,000 a year in direct costs, for at most five years, and cannot be renewed (National Institutes of Health 2026), after which its holder must again compete for R01 grants. Five years without the prospect of renewal is a short horizon for work whose outcomes are uncertain, and an award of this kind may reward past ambition more than it encourages new ambition. Labels can also mislead. The highly cited ecologists surveyed by Gravem et al. (2017) agreed, with strong consensus, that asking applicants to state the transformative potential of their proposals leads them to overstate the importance of their work. Our measures are built from the work done during a cycle, not from what the application promised, and it is the ambition of that work that is associated with a lower probability of renewal. No program becomes transformative by being called so; what matters is what it asks of applicants, how it ranks them, and how long it waits for results.

A fundamental challenge in motivating risk taking in science is that, *ex post*, a failed but risky research project and a low-quality idea may generate similar outcomes—no publication, patent, or product. While fame and fortune can both act as powerful motivators for pursuing risky science, investigators may still hesitate if failed risky projects lead to job loss or lab closure. Without compromising high standards and accountability, motivating risk taking may require simultaneous adjustments to multiple features of the funding environment, including feedback, time horizon, and job security (Manso 2011).

Which adjustments, and in what combination, are questions that experiments could answer, and three design choices are natural places to begin. The first is the weight that reviewers give to preliminary results. In the words of one Nobel laureate, funds are awarded "for compelling ideas, supported by preliminary evidence, creating a high likelihood of success," and investigators obtain that evidence by diverting some of the funds awarded for other work to test new ideas (Kornberg 2007, p. 8). A review that rewards preliminary evidence favors work whose risk has been reduced by the time it is judged. The second is how the scores of reviewers are combined into a ranking. Scores have a mixed record as predictors of the productivity and impact of funded work (Danthi et al. 2014, Li and Agha 2015), and their mean ignores how much reviewers disagree. Asked to rank proposals whose scores had been assigned at random, scientists funded by NIH favored, at the same mean score, the proposals whose scores were more dispersed; applied to the scores of earlier experiments that mimicked NIH review, their preferences would change which proposals are funded in 4 to 6 percent of cases, and in about 10 percent with the preferences of the scientists working closest to the proposals' fields (Carson et al. 2023). In venture competitions, startups that divide the judges are more likely to succeed than startups with the same average score (Gius 2025). Whether such a rule would select better science is unknown, and the scores

of past applications cannot settle it, since NIH does not keep them in a form that would allow the comparison. The third is the length of the award. An investigator who knows she will have time to absorb a setback, and to change course in its light, can afford to plan more ambitious work than one who must show results by the next renewal. Scientists appointed by HHMI, whose awards are long and tolerant of early failure, produce more hits and more flops than comparable scientists funded by NIH, although HHMI differs from NIH in more than the length of its awards (Azoulay et al. 2011).

The same design choices arise wherever an organization decides whether to keep funding work whose outcomes are uncertain. Venture capital firms invest in stages, starting small and cutting their losses as early as they can, because even experienced investors cannot tell in advance which ventures will succeed: in the records of one large firm, the scores its partners gave each venture at the first investment had a correlation of 0.1 with the venture's eventual return, and across the industry, the 6 percent of ventures that returned more than five times their cost produced about half of all gross returns (Kerr et al. 2014). Drug developers face the same decision at each stage of development, and they are more selective with novel candidates: compared with candidates assessed in the same year for the same disease, novel ones are less likely to be taken into human trials, and those that reach the market earn more revenue than incremental drugs (Frankel et al. 2025). Even in a program that Novartis created to fund transformative projects, scientists asked to divide a hypothetical budget among the finalists gave more weight to execution than the firm's own criteria did, and favored safer projects over those they judged riskier and more promising, by enough to change which projects the firm's criteria would have funded (Krieger et al. 2023). Testing which design encourages ambition requires data on these decisions and on their outcomes. Novartis opened its program to researchers, but access of this kind is much harder to obtain in firms than in grant-making bodies, because the information is commercially sensitive. A firm that found a design that raised the ambition of its project teams, or a venture capital firm that found one for the founders it backs, would have every incentive to keep it to itself. Among the organizations that make such decisions, NIH is unusual in having both the scale to learn by experiment which designs encourage ambition and the duty, as a public agency, to publish what it learns.

References

Alberts B, Kirschner MW, Tilghman S, Varmus H (2014) Rescuing US biomedical research from its systemic flaws. *Proc. Natl. Acad. Sci. USA* 111(16):5773–5777.

Althaus CE (2005) A disciplinary perspective on the epistemological status of risk. *Risk Anal.* 25(3):567–588.

Angrist JD, Pischke JS (2009) *Mostly Harmless Econometrics: An Empiricist's Companion* (Princeton University Press, Princeton, NJ).

Arora A, Belenzon S, Patacconi A (2018) The decline of science in corporate R&D. *Strategic Management J.* 39(1):3–32.

Arrow KJ, Lind RC (1970) Uncertainty and the evaluation of public investment decisions. *Amer. Econom. Rev.* 60(3):364–378.

Arts S, Fleming L (2018) Paradise of novelty—or loss of human capital? Exploring new fields and inventive output. *Organ. Sci.* 29(6):1074–1092.

Arts S, Melluso N, Veugelers R (2025) Beyond citations: Measuring novel scientific ideas and their impact in publication text. *Rev. Econom. Statist.*, ePub ahead of print January 28, https://doi.org/10.1162/rest_a_01561.

Ayoubi C, Pezzoni M, Visentin F (2019) The important thing is not to win, it is to take part: What if scientists benefit from participating in research grant competitions? *Res. Policy* 48(1):84–97.

Ayoubi C, Pezzoni M, Visentin F (2021) Does it pay to do novel science? The selectivity patterns in science funding. *Sci. Public Policy* 48(5):635–648.

Azoulay P (2012) Turn the scientific method on ourselves. *Nature* 484(7392):31–32.

Azoulay P, Graff Zivin JS, Li D, Sampat BN (2019) Public R&D investments and private-sector patenting: Evidence from NIH funding rules. *Rev. Econom. Stud.* 86(1):117–152.

Azoulay P, Graff Zivin JS, Manso G (2011) Incentives and creativity: Evidence from the academic life sciences. *RAND J. Econom.* 42(3):527–554.

Azoulay P, Li D (2022) Scientific grant funding. Goolsbee A, Jones BF, eds., *Innovation and Public Policy* (University of Chicago Press, Chicago), 117–150.

Babina T, He AX, Howell ST, Perlman ER, Staudt J (2023) Cutting the innovation engine: How federal funding shocks affect university patenting, entrepreneurship, and publications. *Quart. J. Econom.* 138(2):895–954.

Baumeister RF, Bratslavsky E, Finkenauer C, Vohs KD (2001) Bad is stronger than good. *Rev. General Psych.* 5(4):323–370.

Becker GS (1957) *The Economics of Discrimination* (University of Chicago Press, Chicago).

Bornmann L, Devarakonda S, Tekles A, Chacko G (2020) Are disruption index indicators convergently valid? The comparison of several indicator variants with assessments by peers. *Quant. Sci. Stud.* 1(3):1242–1259.

Boudreau KJ, Guinan EC, Lakhani KR, Riedl C (2016) Looking across and looking beyond the knowledge frontier: Intellectual distance, novelty, and resource allocation in science. *Management Sci.* 62(10):2765–2783.

Carson RT, Graff Zivin J, Shrader JG (2023) Choose your moments: Peer review and scientific risk taking. NBER Working Paper 31409, National Bureau of Economic Research, Cambridge, MA.

Clancy M (2023) Biases against risky research. *New Things Under the Sun* (March 31), https://www.newthingsunderthesun.com/pub/41nqfjgh.

Collison P, Cowen T, Hsu P (2021) What we learned doing Fast Grants. *Future* (June 15), https://future.com/what-we-learned-doing-fast-grants/.

Danthi N, Wu CO, Shi P, Lauer M (2014) Percentile ranking and citation impact of a large cohort of National Heart, Lung, and Blood Institute-funded cardiovascular R01 grants. *Circulation Res.* 114(4):600–606.

Durvasula M, Hemphill CS, Ouellette LL, Sampat B, Williams HL (2023) The NBER Orange Book dataset: A user's guide. *Res. Policy* 52(7):104791.

Fang FC, Casadevall A (2015) Competitive science: Is competition ruining science? *Infect. Immun.* 83(4):1229–1233.

Fields S (2014) Would Fred Sanger get funded today? *Genetics* 197(2):435–439.

Fleming L (2001) Recombinant uncertainty in technological search. *Management Sci.* 47(1):117–132.

Fleming L, Greene H, Li GC, Marx M, Yao D (2019) Government-funded research increasingly fuels innovation. *Science* 364(6446):1139–1141.

Fox CR, Tversky A (1995) Ambiguity aversion and comparative ignorance. *Quart. J. Econom.* 110(3):585–603.

Frakes MD, Wasserman MF (2026) Unveiling the patent landscape of biologic drugs. *Northwestern Univ. Law Rev.* 120(5):1315–1374.

Frankel AP, Krieger JL, Li D, Papanikolaou D (2025) Evaluation and learning in R&D investment. NBER Working Paper 31290, National Bureau of Economic Research, Cambridge, MA.

Franzoni C, Stephan P (2023) Uncertainty and risk-taking in science: Meaning, measurement and management in peer review of research proposals. *Res. Policy* 52(3):104706.

Franzoni C, Stephan P, Veugelers R (2022) Funding risky research. *Entrepreneurship Innovation Policy Econom.* 1:103–133.

Funk ME, Reid CA (1983) Indexing consistency in MEDLINE. *Bull. Medical Library Assoc.* 71(2):176–183.

Funk RJ, Owen-Smith J (2017) A dynamic network measure of technological change. *Management Sci.* 63(3):791–817.

Gan HK, You B, Pond GR, Chen EX (2012) Assumptions of expected benefits in randomized phase III trials evaluating systemic treatments for cancer. *J. Natl. Cancer Inst.* 104(8):590–598.

Ganguli I, Lin J, Meursault V, Reynolds NF (2024) Patent text and long-run innovation dynamics: The critical role of model selection. NBER Working Paper 32934, National Bureau of Economic Research,

Cambridge, MA.

Gius L (2025) Disagreement predicts startup success: Evidence from venture competitions. *Strategy Sci.* 10(2):93–108.

Graff Zivin J, Lyons E (2021) The effects of prize structures on innovative performance. *AEA Papers Proc.* 111:577–581.

Gravem SA, Bachhuber SM, Fulton-Bennett HK, Randell ZH, Rickborn AJ, Sullivan JM, Menge BA (2017) Transformative research is not easily predicted. *Trends Ecol. Evol.* 32(11):825–834.

Greenblatt W, Gupta C, Kao J (2023) Drug repurposing during the COVID-19 pandemic: Lessons for expediting drug development and access. *Health Affairs* 42(3):424–432.

Hall SS (1987) *Invisible Frontiers: The Race to Synthesize a Human Gene* (Atlantic Monthly Press, New York).

Hill R, Yin Y, Stein C, Wang X, Wang D, Jones BF (2025) The pivot penalty in research. *Nature* 642(8069):999–1006.

Jacob BA, Lefgren L (2011) The impact of research grant funding on scientific productivity. *J. Public Econom.* 95(9–10):1168–1177.

Kerr WR, Nanda R, Rhodes-Kropf M (2014) Entrepreneurship as experimentation. *J. Econom. Perspect.* 28(3):25–48.

Knight FH (1921) *Risk, Uncertainty and Profit* (Houghton Mifflin, Boston).

Kogan L, Papanikolaou D, Seru A, Stoffman N (2017) Technological innovation, resource allocation, and growth. *Quart. J. Econom.* 132(2):665–712.

Kolata G (2009) Grant system leads cancer researchers to play it safe. *New York Times* (June 28), https://www.nytimes.com/2009/06/28/health/research/28cancer.html.

Kornberg RD (2007) Testimony. Hearing "2006 Nobel Laureates" before the Subcommittee on Science, Technology, and Innovation of the Committee on Commerce, Science, and Transportation, U.S. Senate, S. Hrg. 110-1209 (May 2), https://www.govinfo.gov/app/details/CHRG-110shrg79907.

Kravitz RL, Franks P, Feldman MD, Gerrity M, Byrne C, Tierney WM (2010) Editorial peer reviewers' recommendations at a general medical journal: Are they reliable and do editors care? *PLoS One* 5(4):e10072.

Krieger J, Li D, Papanikolaou D (2022) Missing novelty in drug development. *Rev. Financial Stud.* 35(2):636–679.

Krieger J, Nanda R, Hunt I, Reynolds A, Tarsa P (2023) Scoring and funding breakthrough ideas: Evidence from a global pharmaceutical company. Working Paper 23-014, Harvard Business School, Boston.

Lane JN, Szajnfarber Z, Crusan J, Menietti M (2026) Beyond feasibility filters: How expertise heterogeneity enables innovation recognition. *Strategic Management J.* 47(5):1368–1432.

Lane JN, Teplitskiy M, Gray G, Ranu H, Menietti M, Guinan EC, Lakhani KR (2022) Conservatism gets funded? A field experiment on the role of negative information in novel project evaluation. *Management Sci.* 68(6):4478–4495.

Langer JS (2012) Enabling scientific innovation. *Science* 338(6104):171.

Lauer M (2016) Are attempts at renewal successful? *NIH Extramural Nexus* (February 16), https://nexus.od.nih.gov/all/2016/02/16/are-attempts-at-renewal-successful/.

Lauer M (2022) FY 2021 by the numbers: Extramural grant investments in research. *NIH Extramural Nexus* (March 7), https://nexus.od.nih.gov/all/2022/03/07/fy-2021-by-the-numbers-extramural-grant-investments-in-research/.

Lerner J, Wulf J (2007) Innovation and incentives: Evidence from corporate R&D. *Rev. Econom. Statist.* 89(4):634–644.

Li D, Agha L (2015) Big names or big ideas: Do peer-review panels select the best science proposals? *Science* 348(6233):434–438.

Littman DR (2015) Releasing the brakes on cancer immunotherapy. *Cell* 162(6):1186–1190.

Manso G (2011) Motivating innovation. *J. Finance* 66(5):1823–1860.

Marx M, Fuegi A (2020) Reliance on science: Worldwide front-page patent citations to scientific articles. *Strategic Management J.* 41(9):1572–1594.

Marx M, Fuegi A (2022) Reliance on science by inventors: Hybrid extraction of in-text patent-to-article citations. *J. Econom. Management Strategy* 31(2):369–392.

Masclans R, Hasan S, Cohen WM (2025) Measuring the commercial potential of science. *Strategic Management J.* 46(9):2199–2236.

Moses H III, Matheson DHM, Cairns-Smith S, George BP, Palisch C, Dorsey ER (2015) The anatomy of medical research: US and international comparisons. *JAMA* 313(2):174–189.

Mueller JS, Melwani S, Goncalo JA (2012) The bias against creativity: Why people desire but reject creative ideas. *Psych. Sci.* 23(1):13–17.

Murray F (2013) Evaluating the role of science philanthropy in American research universities. *Innovation Policy Econom.* 13:23–60.

National Institutes of Health (1997) Review criteria for and rating of unsolicited research grant and other applications. *NIH Guide for Grants and Contracts* 26(22), Notice NOT-97-010 (June 27), https://grants.nih.gov/grants/guide/notice-files/not97-010.html.

National Institutes of Health (2019) Report of the ACD working group on high-risk, high-reward research. Report, Advisory Committee to the Director, National Institutes of Health, Bethesda, MD, https://acd.od.nih.gov/documents/presentations/06132019HRHR_B.pdf.

National Institutes of Health (2026) NIH Director's Pioneer Award (DP1 clinical trial optional). Notice of funding opportunity RFA-RM-27-001 (June 29), https://files.simpler.grants.gov/opportunities/f3b0799c-6539-4045-8b85-42d7a9ef4dd5/attachments/ed6ead5d-829a-4e73-8c7a-e657936a9dc1/RFA-RM-27-001-Full-Announcement.html.

Nuland SB (2003) *The Doctors' Plague: Germs, Childbed Fever, and the Strange Story of Ignác Semmelweis* (W. W. Norton, New York).

Oster E (2019) Unobservable selection and coefficient stability: Theory and evidence. *J. Bus. Econom. Statist.* 37(2):187–204.

Packalen M, Bhattacharya J (2020) NIH funding and the pursuit of edge science. *Proc. Natl. Acad. Sci. USA* 117(22):12011–12016.

Park M, Leahey E, Funk RJ (2023) Papers and patents are becoming less disruptive over time. *Nature* 613(7942):138–144.

Petsko GA (2012) Goodbye, Columbus. *Genome Biol.* 13(5):155.

Pich MT, Loch CH, De Meyer A (2002) On uncertainty, ambiguity, and complexity in project management. *Management Sci.* 48(8):1008–1023.

Rosenberg N (1996) Uncertainty and technological change. Fuhrer JC, Little JS, eds., *Technology and Growth: Conference Proceedings*, Conference Series No. 40 (Federal Reserve Bank of Boston, Boston), 91–110.

Rzhetsky A, Foster JG, Foster IT, Evans JA (2015) Choosing experiments to accelerate collective discovery. *Proc. Natl. Acad. Sci. USA* 112(47):14569–14574.

Sanchez C, Dunning D (2023) Are experts overconfident? An interdisciplinary review. *Res. Organ. Behav.* 43:100195.

Shibayama S, Wang J (2020) Measuring originality in science. *Scientometrics* 122(1):409–427.

Teplitskiy M, Peng H, Blasco A, Lakhani KR (2022) Is novel research worth doing? Evidence from peer review at 49 journals. *Proc. Natl. Acad. Sci. USA* 119(47):e2118046119.

Tham WY, Staudt J, Perlman ER, Cheng SD (2026) Scientific talent leaks out of funding gaps. *Rev. Econom. Statist.*, ePub ahead of print September 8, https://doi.org/10.1162/rest.a.1865.

Tian X, Wang TY (2014) Tolerance for failure and corporate innovation. *Rev. Financial Stud.* 27(1):211–255.

Torvik V, Smalheiser N (2021) Author-ity 2018—PubMed author name disambiguated dataset. Dataset, Illinois Data Bank, University of Illinois Urbana-Champaign, https://doi.org/10.13012/B2IDB-2273402_V1.

Torvik VI, Smalheiser NR (2009) Author name disambiguation in MEDLINE. *ACM Trans. Knowledge Discovery Data* 3(3):11.

Uzzi B, Mukherjee S, Stringer M, Jones B (2013) Atypical combinations and scientific impact. *Science* 342(6157):468–472.

Veugelers R, Wang J, Stephan P (2026) Do funding agencies select and enable novel research: Evidence from ERC. *Econom. Innovation New Technol.* 35(6):1109–1126.

Wang H (2026) Funding the runners-up beats a golden ticket. Preprint, submitted September 17, https://doi.org/10.48550/arXiv.2609.19552.

Wang J, Veugelers R, Stephan P (2017) Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Res. Policy* 46(8):1416–1436.

Woolston C (2014) Sanger's legacy stirs up digital chatter. *Nature* 511(7507):9.

Wu L, Wang D, Evans JA (2019) Large teams develop and small teams disrupt science and technology. *Nature* 566(7744):378–382.

Table 1a: Descriptive Statistics—Investigator Characteristics

	Mean	Median	Std. Dev.	Min.	Max.
Female	0.24	0.00	0.43	0.00	1.00
Degree year	1983	1983	12	1932	2013
Degree: PhD	0.68	1.00	0.47	0.00	1.00
Degree: MD	0.20	0.00	0.40	0.00	1.00
Degree: MD/PhD	0.11	0.00	0.31	0.00	1.00
Degree: Other	0.02	0.00	0.13	0.00	1.00
Nb. of R01-equivalent grants	1.74	1.00	1.17	1.00	14.00
Nb. of R01-equivalent grant cycles	2.58	2.00	2.26	1.00	22.00
Career NIH R01 funding (2024 dollars)	12,874,251	8,896,409	13,231,382	0	316,013,294
Career NIH funding (2024 dollars)	21,288,397	11,260,949	34,874,456	154,034	2,277,009,052
Career elite status	0.05	0.00	0.22	0.00	1.00

Notes: One observation per investigator of the analysis sample with at least one cycle that has a publication acknowledging the grant. Grants and cycles are counted within the sample (R01-equivalent grants first funded 1980 to 2015). Career funding sums the investigator's NIH research funding over all years, in 2024 dollars. Career elite status: ever a Nobel Prize, a Lasker Award, an HHMI investigatorship, or membership in the Institute of Medicine or the National Academy of Sciences.

Table 1b: Descriptive Statistics—Grant Cycle

	Renewed grant cycles					Not renewed grant cycles				
	Mean	Median	Std. Dev.	Min.	Max.	Mean	Median	Std. Dev.	Min.	Max.
Grant cycle number	1.91	1.00	1.20	1.00	9.00	1.61	1.00	1.08	1.00	10.00
First cycle of grant	0.50	1.00	0.50	0.00	1.00	0.65	1.00	0.48	0.00	1.00
Year started	1996	1996	9	1980	2015	2000	2001	10	1980	2015
Cycle length (years)	4.35	4.00	1.14	2.00	14.00	4.37	4.00	1.05	2.00	17.00
Time, degree to cycle start (years)	17.14	16.00	8.73	-16.00	64.00	18.57	17.00	9.61	-23.00	62.00
Investigator publications before the cycle	51.5	33.0	56.9	0.0	1,148.0	55.4	36.0	62.6	0.0	1,196.0
Investigator citations before the cycle	1,662	628	3,336	0	89,394	1,859	711	3,712	0	137,911
Publications acknowledging grant										
Nb. of publications	8.01	6.00	8.05	1.00	587.00	5.99	4.00	6.56	1.00	153.00
Journal impact factor	6.45	5.46	4.01	0.28	43.28	5.61	4.76	3.71	0.14	43.28
Citations (five years)	242	127	385	0	14,782	160	71	323	0	34,306
Highly selective journal	0.13	0.00	0.34	0.00	1.00	0.06	0.00	0.24	0.00	1.00
Publications not acknowledging grant										
Nb. of publications	11.51	7.00	14.61	0.00	362.00	11.95	7.00	15.27	0.00	316.00
Journal impact factor	6.35	5.33	3.96	0.22	43.28	5.52	4.73	3.32	0.13	43.28
Citations (five years)	350	135	692	0	15,059	331	127	664	0	34,419
Highly selective journal	0.20	0.00	0.40	0.00	1.00	0.13	0.00	0.33	0.00	1.00
Share listing an NIH grant	0.65	0.71	0.31	0.00	1.00	0.67	0.75	0.31	0.00	1.00

Notes: The unit of analysis is the grant cycle, among cycles with at least one publication acknowledging the grant. Grant cycle number minus one is the number of times the grant has been competitively renewed before. Time from degree to cycle start is measured from the investigator's last terminal degree. Investigator publications before the cycle count the investigator's English original research in Author-ity published before the cycle's first year; investigator citations are the OpenAlex citations from *PubMed* papers to the same publications through the year before the cycle's first year. Publications acknowledging the grant are those published between the cycle's first year and its last year of payments; publications not acknowledging the grant are all the investigator's other publications of the same years, whether they acknowledge another grant or none. The share listing an NIH grant is the share of these other publications whose *PubMed* grant list names an NIH grant. Citations to the cycle's publications are five-year OpenAlex citations.

Table 1c: Descriptive Statistics—Measures of Risk Taking and Novelty

	Renewed grant cycles					Not renewed grant cycles				
	Mean	Median	Std. Dev.	Min.	Max.	Mean	Median	Std. Dev.	Min.	Max.
Extreme tail outcomes	36.52	33.93	22.68	0.00	94.70	37.23	35.18	22.64	0.00	94.66
Disruption	59.74	62.48	23.00	2.64	96.94	59.60	62.80	23.92	2.64	96.94
Pivoting	0.74	0.75	0.14	0.00	1.00	0.74	0.76	0.15	0.00	1.00
Standing out	0.80	0.81	0.10	0.00	1.00	0.80	0.82	0.11	0.00	1.00
Novelty (share of new noun phrases)	0.04	0.04	0.04	0.00	1.00	0.04	0.03	0.04	0.00	1.00

Notes: The unit of analysis is the grant cycle, among cycles with at least one publication acknowledging the grant; each measure is defined only for the cycles with qualifying publications. Extreme tail outcomes and disruption are in percentile points.

Table 1d: Descriptive Statistics—Correlations across Measures of Risk Taking and Novelty

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) Extreme tail outcomes	1.000						
(2) Disruption	0.389	1.000					
(3) Pivoting	-0.099	0.011	1.000				
(4) Standing out	0.087	0.030	0.237	1.000			
(5) Novelty	-0.166	-0.073	0.084	-0.048	1.000		
(6) Nb. of publications	0.467	0.347	-0.141	0.077	-0.065	1.000	
(7) Citations	0.184	0.182	-0.047	0.059	0.018	0.614	1.000

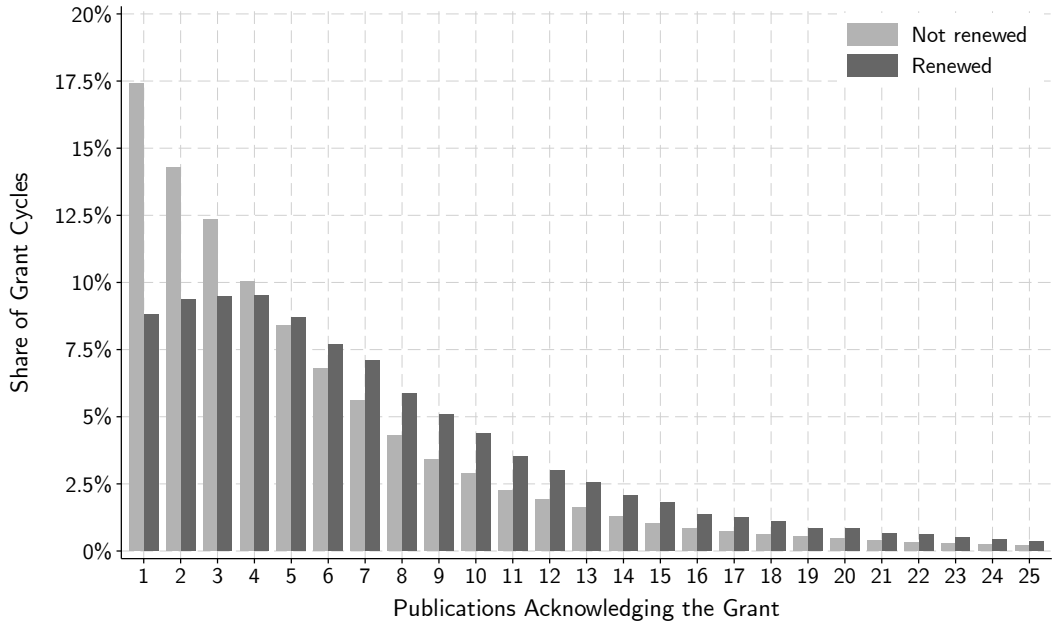
Notes: Pearson correlations over the 69,545 cycles for which every variable is defined. The number of publications and the five-year citations refer to the publications acknowledging the grant during the cycle.

Table 1e: Descriptive Statistics—Downstream Outcomes of Renewed and Terminated Grant Cycles

	Mean	Median	Std. Dev.	Min.	Max.	Cycles
A. Renewed cycles						
Article Citations	280.4	164.0	432.5	0.0	34,407.0	42,127
Direct Grant Patents	0.22	0.00	0.95	0.00	46.00	42,127
Patent Citations	24.05	4.00	103.22	0.00	9,911.00	42,127
FDA Patent Citations	0.63	0.00	4.27	0.00	198.00	42,127
KPSS Patent Value	359.8	0.0	2,187.6	0.0	160,502.4	42,127
Commercial Potential	0.562	0.626	0.254	0.000	0.999	39,083
B. Terminated cycles						
Article Citations	93.5	43.0	169.3	0.0	5,263.0	57,597
Direct Grant Patents	0.10	0.00	0.56	0.00	22.00	57,597
Patent Citations	5.57	0.00	34.93	0.00	4,656.00	57,597
FDA Patent Citations	0.16	0.00	2.09	0.00	213.00	57,597
KPSS Patent Value	77.8	0.0	1,007.3	0.0	173,181.0	57,597
Commercial Potential	0.554	0.613	0.280	0.000	0.999	46,062

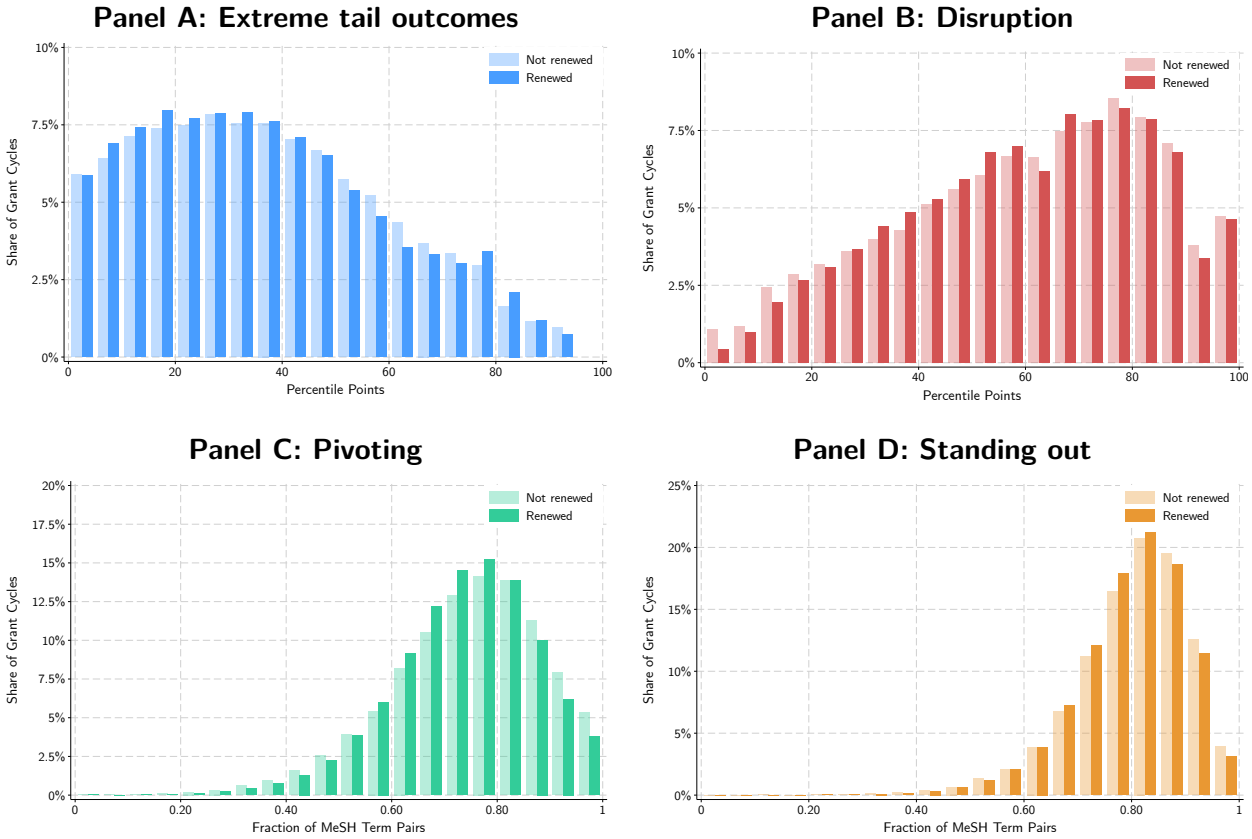
Notes: The unit of analysis is the grant cycle: the cycles of Table 3, that is, cycles evaluated by 2020 with at least one publication acknowledging the grant and a known number within the grant, renewed (Panel A) or terminated (Panel B). A terminated cycle is one that was not renewed and whose grant was not revived later. The outcomes, as in Table 3, concern the five years after the cycle's last payment: the five-year citations of the publications that acknowledge the grant, on which the investigator is an author, published in those years (article citations); the patents that NIH RePORTER links to the grant and whose application was filed in those years (direct grant patents); the citations of those publications from US patents (patent citations), and from drug- and device-related patents among them (FDA patent citations); the private value of the distinct citing patents, from Kogan et al. (2017), in millions of 2023 dollars (KPSS patent value); and the mean commercial potential of the publications, from Masclans et al. (2025). For renewed cycles, these five years include work funded by the renewal; for terminated cycles, they follow the end of the grant's funding. Cycles: the cycles on which each outcome is defined; commercial potential is available for publications from 1990. The scores come from a separate model for each publication year, so they rank the publications of a year but are not comparable across years; the regressions of Table 3 include evaluation-year effects. The sources of the outcomes, and how each is matched to the cycles, are described in Appendix L.

Figure 1: Distribution of Grant Cycle Publications by Renewal Status



Notes: Histogram of the number of publications acknowledging the grant during the grant cycle, among cycles with at least one. The unit of analysis is the grant cycle. The 2,464 cycles with more than 25 publications (2.4% of the sample) are not shown. Each bar gives the share of the cycles of its group (renewed or not renewed), all cycles of the group counted.

Figure 2: Distribution of Risk-Taking Measures by Renewal Status



Notes: Histograms of the four measures among cycles with at least one publication acknowledging the grant, in 20 bins of equal width; each bar gives the share of the cycles of its group (renewed or not renewed). Extreme tail outcomes: the difference between the highest and the lowest citation percentile among the cycle's publications, in percentile points. Disruption: the highest percentile of the disruption index DI_1^{nok} among the cycle's publications. Pivoting: the fraction of MeSH term pairs in the cycle's publications that appear in none of the investigator's publications of the five years before the cycle. Standing out: the fraction of MeSH term pairs in the cycle's publications that appear in no other funded grant abstract of the same NIH institute, study section and evaluation year. The unit of analysis is the grant cycle.

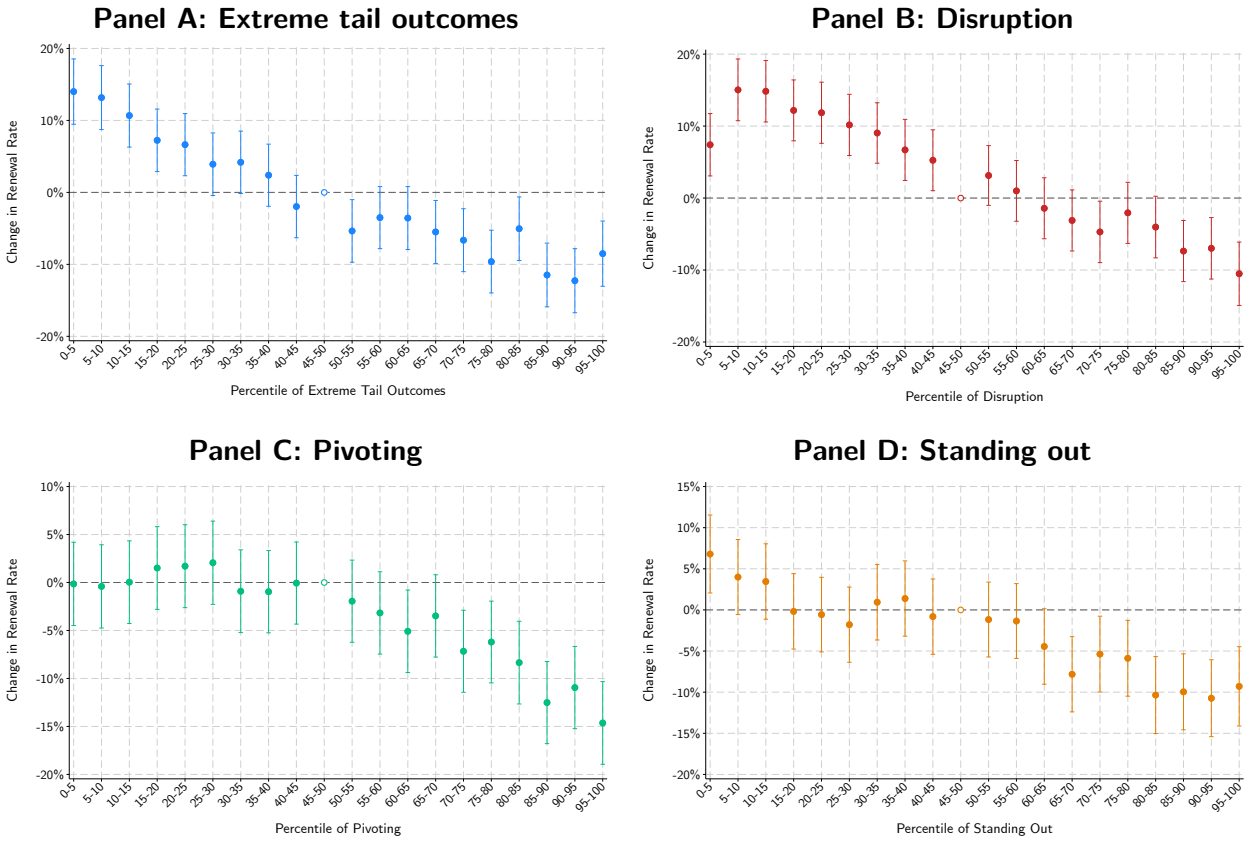
Table 2: Association of Risk Taking and Grant Renewal

	Extreme tail outcomes			Disruption			Pivoting			Standing out		
	(1) All	(2) First	(3) Later	(4) All	(5) First	(6) Later	(7) All	(8) First	(9) Later	(10) All	(11) First	(12) Later
Bottom decile ($<10^{th}$ percentile)	0.051**	0.054**	0.053**	0.023**	0.022**	0.032**	0.008	0.002	0.008	0.031**	0.025**	0.044**
	(0.006)	(0.007)	(0.011)	(0.005)	(0.006)	(0.010)	(0.005)	(0.007)	(0.008)	(0.006)	(0.007)	(0.010)
Top decile ($>90^{th}$ percentile)	-0.032**	-0.032**	-0.032**	-0.042**	-0.055**	-0.034**	-0.037**	-0.029**	-0.046**	-0.025**	-0.030**	-0.019*
	(0.006)	(0.009)	(0.008)	(0.005)	(0.007)	(0.008)	(0.005)	(0.006)	(0.012)	(0.006)	(0.007)	(0.010)
Mean of dependent variable	0.445	0.381	0.522	0.423	0.358	0.512	0.421	0.357	0.509	0.427	0.363	0.513
Effect of bottom decile, in s.d. units	0.103	0.112	0.105	0.046	0.045	0.065	0.015	0.003	0.016	0.062	0.051	0.089
Effect of top decile, in s.d. units	-0.064	-0.065	-0.064	-0.086	-0.116	-0.068	-0.076	-0.061	-0.091	-0.051	-0.062	-0.038
Nb. of investigators	34,792	30,115	17,682	38,791	34,966	18,465	39,337	35,638	18,621	35,396	31,468	16,625
Nb. of grant cycles	87,718	46,861	38,094	99,504	56,381	40,208	101,435	57,911	40,571	83,052	48,007	33,300
Adjusted R^2	0.1636	0.1485	0.1556	0.1699	0.1525	0.1596	0.1676	0.1486	0.1597	0.1662	0.1478	0.1592

Notes: Estimates stem from a linear probability model in which the unit of analysis is the grant cycle and the dependent variable is an indicator for whether the grant cycle was renewed, that is, whether the next competing award on the grant starts within three fiscal years of the cycle's last payment. Each column comes from a separate regression. For each measure, the first column (All) includes every cycle, the second (First) the first cycle of each investigator-grant, and the third (Later) its later cycles; cycles of grants first observed at a renewal have no cycle number and appear only in the first. The independent variables are indicators for the bottom and the top decile of the measure, computed within evaluation year on every cycle where the measure is defined; the middle eight deciles are the omitted category. Extreme tail outcomes: the difference between the highest and the lowest citation percentile among the publications that acknowledge the grant during the cycle; citation percentiles rank five-year citations within publication year. Disruption: the highest percentile of the disruption index DI_1^{nok}, which follows Bornmann et al. (2020) in excluding the N_k term, among those publications; percentiles rank it within publication year. Pivoting: the fraction of MeSH term pairs in those publications that appear in none of the investigator's publications of the five years before the cycle. Standing out: the fraction of MeSH term pairs in those publications that appear in no other funded grant abstract of the same NIH institute, study section and evaluation year. Every regression includes indicators for gender, degree, career age (years since degree, in five-year bins), the year of evaluation, the length of the cycle, and each pair of NIH institute and study section; the log of NIH funding to the investigator's institution, the log of the investigator's career citations, the log of the investigator's prior NIH funding, and an indicator for no prior NIH funding; and the log of the mean journal impact factor of the cycle's publications, indicators for a publication in *NEJM*, *Science*, *Cell* or *Nature* among the publications that acknowledge the grant and among the investigator's other publications, indicators for the number of publications in each group (0, 1, 2, 3, 4, 5, 6-7, 8-10, 11-15, 16-20, 21 or more), and an indicator for cycles whose window runs past 2018, where Author-ity ends. Effects in s.d. units: the coefficient divided by the standard deviation of the dependent variable.

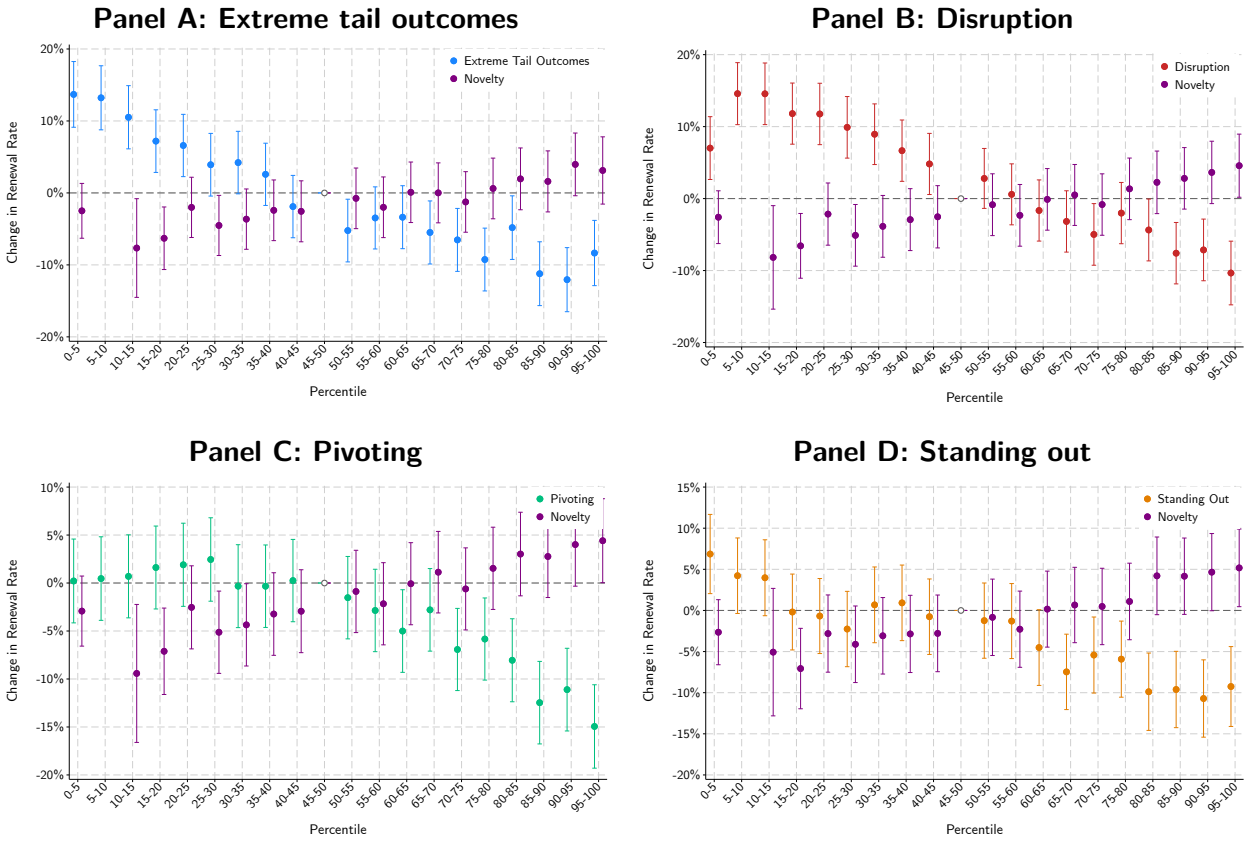
Standard errors in parentheses, clustered by investigator and by grant. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Figure 3: Association of Risk Taking and Grant Renewal



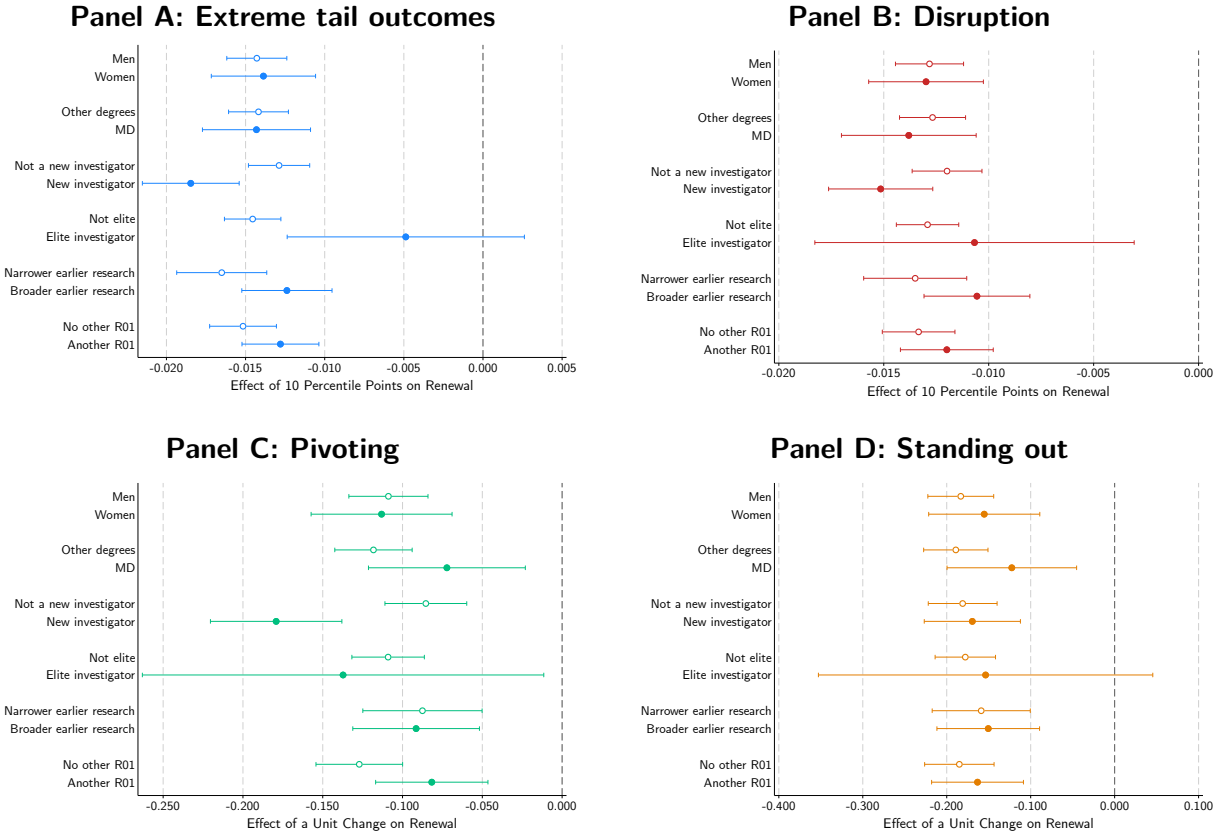
Notes: The plots present coefficient estimates, with 95% confidence intervals, from four linear probability models, one per measure, in which the unit of analysis is the grant cycle and the dependent variable is an indicator for whether the grant cycle was renewed, with the controls of Table 2. Each regression includes indicators for 19 of the 20 ventiles of the measure, computed within evaluation year; the omitted ventile is the 45th to 50th percentile (hollow marker). Coefficients and confidence intervals are divided by the mean renewal rate of the estimation sample, so that each point is a change in the renewal rate relative to its mean. Measures as in Table 2. Standard errors are clustered by investigator and by grant.

Figure 4: Association of Risk Taking and Novelty with Grant Renewal



Notes: The plots present coefficient estimates, with 95% confidence intervals, from a linear probability model in which the unit of analysis is the grant cycle and the dependent variable is an indicator for whether the grant cycle was renewed, with the controls of Table 2. Each regression includes indicators for 19 of the 20 ventiles of the risk measure and 19 of the 20 ventiles of novelty, all computed within evaluation year, the 45th to 50th percentile omitted (hollow marker); each panel shows both sets of coefficients from the same regression, the risk measure's in its color and novelty's in purple. Novelty is the share of the noun phrases in the cycle's publications that are new: first used by one of these publications in the corpus of Arts et al. (2025), and used again by a later publication that shares none of its authors (Appendix F); new noun phrases are summed over the cycle's publications and divided by all their noun phrases. 15% of cycles coin no new noun phrase and fall in its first ventile, so ventiles with fewer than 100 cycles are not plotted. Coefficients and confidence intervals are divided by the mean renewal rate of the estimation sample. Standard errors are clustered by investigator and by grant.

Figure 5: The Penalty for Risk Taking and the Traits of the Investigator



Notes: The plots present coefficient estimates, with 95% confidence intervals, from linear probability models in which the unit of analysis is the grant cycle and the dependent variable is an indicator for whether the grant cycle was renewed, with the controls of Table 2 (the regressions of Table I1). Each regression includes one risk measure as a continuous regressor (extreme tail outcomes and disruption in units of 10 percentile points, pivoting and standing out as fractions), an indicator for the trait noted in the figure, and their interaction; hollow markers show the measure's coefficient without the trait, solid markers its sum with the interaction. Traits, all fixed when the cycle starts: women; an MD alone, against PhDs, MD-PhDs and other degrees; new investigators, whose cycle starts in their first year of an R01-equivalent award; elite investigators, members of the National Academy of Sciences or of the National Academy of Medicine (formerly the Institute of Medicine), or investigators of the Howard Hughes Medical Institute, in the year before the cycle starts; breadth of earlier research above the median of the cycle's study section (cycles with at least five first- or last-author publications in the five years before the cycle); another R01-equivalent grant held in the cycle's first year. Standard errors are clustered by investigator and by grant.

Table 3: Risk Taking and the Downstream Outcomes of Renewed and Terminated Grant Cycles

	Extreme tail outcomes				Disruption			
	First cycle		Later cycles		First cycle		Later cycles	
	Renewed	Terminated	Renewed	Terminated	Renewed	Terminated	Renewed	Terminated
Article Citations	-0.185**	-0.130**	-0.096**	-0.070**	0.064**	-0.027	0.067**	0.009
	(0.018)	(0.019)	(0.024)	(0.025)	(0.018)	(0.018)	(0.024)	(0.027)
Direct Grant Patents	-0.023	0.090	-0.016	-0.162†	0.278**	0.037	0.274**	0.149†
	(0.069)	(0.071)	(0.071)	(0.083)	(0.070)	(0.065)	(0.069)	(0.082)
Patent Citations	-0.007	-0.023	0.012	-0.059	0.325**	0.170	0.181*	0.076
	(0.064)	(0.085)	(0.090)	(0.066)	(0.062)	(0.114)	(0.070)	(0.063)
FDA Patent Citations	0.061	0.081	0.161	0.081	0.197*	0.138	0.045	-0.093
	(0.100)	(0.203)	(0.115)	(0.194)	(0.098)	(0.182)	(0.125)	(0.180)
KPSS Patent Value	0.063	0.079	-0.004	-0.384**	0.368**	0.278*	0.318**	0.210
	(0.108)	(0.122)	(0.126)	(0.138)	(0.094)	(0.139)	(0.112)	(0.142)
Commercial Potential	-0.004	0.005	-0.005	0.001	0.013*	-0.001	0.009†	0.006
	(0.006)	(0.005)	(0.005)	(0.007)	(0.006)	(0.005)	(0.005)	(0.007)
Cycles	18,090	28,906	20,068	18,083	20,387	36,060	20,742	19,477

	Pivoting				Standing out			
	First cycle		Later cycles		First cycle		Later cycles	
	Renewed	Terminated	Renewed	Terminated	Renewed	Terminated	Renewed	Terminated
Article Citations	0.084**	-0.016	0.045*	0.032	-0.029	-0.063**	-0.006	-0.075**
	(0.017)	(0.018)	(0.023)	(0.023)	(0.018)	(0.019)	(0.022)	(0.027)
Direct Grant Patents	0.220**	0.212**	0.180**	0.156*	0.129†	0.077	0.132†	0.090
	(0.072)	(0.064)	(0.064)	(0.074)	(0.076)	(0.069)	(0.070)	(0.081)
Patent Citations	0.192**	0.076	0.220**	0.145*	0.049	0.073	0.121	0.066
	(0.054)	(0.078)	(0.072)	(0.071)	(0.068)	(0.088)	(0.085)	(0.080)
FDA Patent Citations	-0.007	-0.165	0.113	0.040	0.009	-0.166	0.136	0.279
	(0.099)	(0.132)	(0.105)	(0.194)	(0.116)	(0.166)	(0.111)	(0.179)
KPSS Patent Value	0.180*	-0.029	0.303**	0.190	-0.054	-0.005	0.246†	0.162
	(0.087)	(0.105)	(0.098)	(0.120)	(0.091)	(0.138)	(0.127)	(0.147)
Commercial Potential	0.048**	0.049**	0.039**	0.018**	0.024**	0.027**	0.025**	0.021**
	(0.006)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.008)
Cycles	20,882	37,105	20,838	19,741	17,446	29,998	17,107	15,757

Notes: Each coefficient comes from a separate Poisson regression in which the unit of analysis is the grant cycle; the regressor is an indicator for a cycle whose measure lies above its median, computed within evaluation year on every cycle where the measure is defined. The cycles are those evaluated by 2020 with at least one publication acknowledging the grant and on which the measure is defined: first cycles of a grant and later cycles, as in Table 2 (the cycles whose number within the grant is unknown are left out), renewed or terminated. A terminated cycle is one that was not renewed and whose grant was not revived later. The controls are those of Table 2. The outcomes concern the five years after the cycle's last payment and, except for direct grant patents, the publications that acknowledge the grant, on which the investigator is an author, published in those years. Article citations: the publications' five-year citations. Direct grant patents: the patents that NIH RePORTER links to the grant and whose application was filed in the five years. Patent citations: the publications' citations from US patents (Reliance on Science, 2026 release, match confidence 6 or more, front-page and in-text citations). FDA patent citations: the citations from drug- and device-related patents among them, that is, the Orange Book's patents, the patents on New Drug Applications in a list compiled by Bhaven Sampat, the patents on therapeutic biologics listed by Frakes and Wasserman (2026), and the patents whose term was extended for regulatory review. KPSS patent value: the private value of the distinct citing patents, from Kogan et al. (2017), for patents of listed firms granted through 2023, in millions of 2023 dollars. Commercial potential: the mean commercial potential of the publications, from Masclans et al. (2025), the predicted probability that the publication is cited by a patent that its owner keeps in force by paying the first maintenance fee. For renewed cycles, these five years include work funded by the renewal; for terminated cycles, they follow the end of the grant's funding. Descriptive statistics for the outcomes are presented in Table 1e. The sources of the outcomes, and how each is matched to the cycles, are described in Appendix L.

Standard errors in parentheses, clustered by investigator and by grant. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Appendix A: Sample Construction and Robustness

Our estimates rest on a sample built in several steps. We observe only funded grants. Each of our measures is defined only for the cycles whose publications allow it to be computed. A cycle's publications are those that acknowledge the grant and that Author-ity, which disambiguates the authors of *PubMed*, attributes to its principal investigator. A skeptic could object that the penalty reflects something other than the study section's decisions. If the riskiest projects produce no publication at all, they leave the sample before their risk can be measured. If publications are attributed to the wrong investigators, the measures contain errors that need not be random. If a grant that is not renewed was abandoned by an investigator whose interests lay elsewhere, non-renewal reflects her exit, not the reviewers' verdict.[1]

Which grants. Our sample is drawn from the R01-equivalent research grants first funded from 1980 to 2015. It includes the activity codes R01 and R37 (the MERIT awards, whose extension years continue a cycle without a competing review), and R29 and R23, the earlier forms of NIH's awards to new investigators. It excludes several mechanisms that NIH also counts as R01-equivalent but that may follow distinct dynamics around risk taking: DP1, DP2, and DP5, the awards of NIH's program for high-risk, high-reward research, which fund risk taking by design; R56 and RL1, which the investigator may not apply for as such; U01, cooperative agreements in which NIH staff take an active part in the research; and R35 and RF1, which provide longer-term support to investigators. It also excludes the cycles whose application was reviewed by a special emphasis panel, an ad hoc committee convened for a set of applications, which may not offer a path to renewal.

Which cycles. A grant's cycles are numbered from its first award, whoever leads them, and each cycle enters the sample or not on the strength of its own records. Whether a cycle enters therefore depends only on what happened during it, never on its outcome or on what happened to the grant afterward. A cycle enters if the competing award that starts it names a single principal investigator and that investigator holds the grant in every year of the cycle. A co-investigator added during the cycle does not exclude it; a principal investigator replaced before the cycle ends does. Our measures and controls are defined for one person: a cycle's publications are those that list its principal investigator among their authors, pivoting compares them with her earlier publications, and the controls describe her career. A cycle that starts with several principal investigators has no such person. The outcome is the grant's renewal, whoever applies for it. In 98 percent of the renewals in our sample, the principal investigator of the renewed cycle also leads the next one; another single investigator does so in 1 percent, and a team in 1 percent (6 percent among the cycles that start from 2008 onward).

We test three further filters the same way, on the cycle's own records. A cycle must have at least three budget years of its own. None of its records may be at an institution that has no code or is coded as a firm, a national laboratory, a society, or a foreign university, so that the sample consists of academic investigators in comparable settings. And the award that starts it must carry an application ID, which links the cycle to the study section that reviewed it.

Of the 138,115 cycles that start from 1980 to 2015, whose outcome we observe, and that pass our other filters on grants and cycles, 3,229 (2.3 percent) start with an award that names several principal investigators, and 3,041 (2.2 percent) lose their principal investigator before they end. Nearly all of the first group started after 2007: such awards start 10.0 percent of the cycles that begin from 2008 to 2015, against 0.04 percent of those that begin earlier. Both groups are renewed less often than the cycles we keep: from 2008 to 2015, 16.4 percent of the cycles that start with several principal investigators are renewed and 4.0 percent of those whose principal investigator is replaced, against 23.9 percent of the others. Our estimates therefore describe grants led by a single investigator, and we present no evidence on the growing share of grants led by teams.

Funded grants only. We observe a grant only once it has been funded, and NIH does not release the applications it turns down. If initial review treats risk differently from renewal review, the grants we observe are a selected set, and the direction of the resulting bias cannot be signed without the applications that failed: it depends on whether initial review screens out the riskiest proposals, the safest, or those in between. What we can say is that the first review is the more selective of the two. From 2015 to 2025, 18 percent of new applications for

[1]We thank an anonymous referee for pressing us on the first of these points.

R01-equivalent grants were funded, against 40 percent of competing renewals, and renewals were funded at two to three times the rate of new applications in every one of those years (NIH RePORT, Table 210).[2] The cycles we study have therefore already passed through the more selective review. Our estimates describe the second one, the renewal decision on grants that were funded. We do not know whether the first also penalizes risk; if it does, risky proposals are penalized at both stages, and we measure only the second.

Cycles without an acknowledging publication. Every one of our measures is computed from the publications that acknowledge the grant during the cycle. A cycle that produced no publications has no measure and is excluded from every estimation sample in the paper. A cycle with publications can also lack a measure when its publications do not meet that measure's requirements: extreme tail outcomes, for instance, requires two publications with a citation percentile (Section 3.3, footnote 7). Table A1 gives the number of cycles of each kind, for first and later cycles, with their renewal rates. Of the 123,018 cycles of the grants we study, 19,534 (15.9 percent) produced no acknowledging publication: 22 percent of first cycles and 6 percent of later ones. Another 14,288 (11.6 percent) produced exactly one. Our sample consists of the 103,484 cycles with at least one acknowledging publication. Cycles without a publication are renewed far less often than the others, 23.3 percent against 41.9 percent, with a gap of 15 percentage points for first cycles (20.0 against 35.5 percent) and 10 for later ones (40.9 against 50.6 percent). Cycles with exactly one publication are renewed about as rarely (26.7 percent). Not all of these cycles are failed projects, however: nearly a quarter of those without a publication were renewed. Their share has also fallen steadily, from 25 percent of the cycles that began in the 1980s to 18 percent in the 1990s, 10 percent in the 2000s, and 6 percent after 2010.

We cannot measure the risk of a cycle that produced nothing, but we can measure the risk of what its investigator did before it, as in Appendix G: for a grant's first cycle, each measure computed on her publications of the five years before the grant; for a later cycle, the measure of the grant's previous cycle. In Table A2, we compare this earlier record between cycles with and without an acknowledging publication. The controls are those of Table 2 except the grant output controls, which describe the cycle's own publications.

Cycles without a publication do not come from investigators with safer records. On extreme tail outcomes and disruption, their records are riskier, by 0.020 and 0.023 standard deviations (both significant at the 1 percent level); on standing out, by about as much, significant at the 10 percent level only; and on pivoting, not at all. On the first two measures, the difference comes from first cycles. It is also small: the 95 percent confidence intervals extend no further than about 0.04 standard deviations. By the estimates of Appendix G, a record riskier by 0.02 standard deviations is associated with a probability of renewal lower by about 0.03 percentage points, a negligible part of the 18.6-point gap in renewal between cycles with and without a publication.

If the risk of these cycles resembled their investigators' records, including them would strengthen the association between risk and renewal. We cannot say by how much: these cycles have no measure of their own, and the record before a cycle predicts the cycle's own risk only weakly (Appendix G).

A sample common to the four measures. Each measure is defined on its own set of cycles (Section 3), so the four panels of Table 2 rest on different samples. In column 2 of Table A3, we re-estimate each panel on the 69,826 cycles for which all four measures are defined. The penalty does not change. In the top decile it is 3.0 percentage points for extreme tail outcomes (3.2 in the full sample), 4.6 for disruption (4.2), 4.0 for pivoting (3.7), and 2.9 for standing out (2.5).

Misattributed publications. Our measures rely on Author-ity to attribute publications to investigators. Among NIH-funded scientists, its precision and recall both exceed 99 percent, and they vary little between common and uncommon last names (Lerchenmueller and Sorenson 2016). The errors that remain are nonetheless more likely for common names, so in columns 3 and 4 of Table A3, we split the sample at the median frequency of the investigator's last name among *PubMed* authors. The penalty is similar in both halves: in the top decile, 3.0 and 3.5 percentage points for extreme tail outcomes, 3.4 and 4.9 for disruption, 3.6 and 4.0 for pivoting, and 2.7 and 2.0 for standing out.

[2] Applications reviewed and funded, with first submissions and resubmissions counted separately, pooled over fiscal years 2015 to 2025. By year, the success rate of new applications ranged from 11 to 20 percent, that of competing renewals from 31 to 44 percent.

Rejection or abandonment. We do not observe whether a renewal application was submitted, so a grant that is not renewed may have been turned down by the study section or abandoned by its investigator. Abandonment is most plausible where the investigator's main work lies outside NIH's remit. A chemist, an engineer, or a social scientist with one biomedical project in a broader agenda may let the grant lapse and return to her own field, and her work, set against biomedical study sections and against her earlier publications, may well look risky by our measures. Part of the penalty would then reflect exit rather than review. An investigator who does not apply because she expects her risky work to be turned down is a different case: her exit is part of the penalty. In Table A4, we split the sample at the median share of the investigator's publications that acknowledge NIH funding, among those that appeared before the cycle began (74 percent). If exit produced the penalty, it would be larger among the investigators less attached to NIH. It is not. In the top decile, their penalty is 3.1 percentage points for extreme tail outcomes (2.9 among the others), 3.9 for disruption (4.6), 2.6 for pivoting (5.0), and 1.9 for standing out (3.4). Every measure is penalized in both halves, and for pivoting and standing out the penalty is about twice as large among the investigators most attached to NIH.

Table A1: Cycles without an Acknowledging Publication

	All cycles	First cycles	Later cycles
Nb. of cycles facing renewal	123,018	76,172	44,060
Nb. (share) with no acknowledging paper	19,534 (15.9%)	16,539 (21.7%)	2,591 (5.9%)
Nb. (share) with exactly one acknowledging paper	14,288 (11.6%)	11,492 (15.1%)	2,585 (5.9%)
Renewal rate, cycles with no acknowledging paper	0.233	0.200	0.409
Renewal rate, all other cycles	0.419	0.355	0.506
Renewal rate, cycles with exactly one acknowledging paper	0.267	0.250	0.332

Notes: The unit of analysis is the grant cycle. First cycles are the first cycle of their grant; later cycles, the second and beyond.

Table A2: Risk before the Cycle, for Cycles with and without an Acknowledging Publication

	(1) Extreme tail outcomes	(2) Disruption	(3) Pivoting	(4) Standing out
Cycle with no acknowledging publication	0.0502**	0.0500**	0.0008	0.0019†
	(0.0187)	(0.0166)	(0.0011)	(0.0010)
Mean of lagged risk, other cycles	4.603	6.847	0.783	0.804
Std. dev. of lagged risk, other cycles	2.486	2.198	0.132	0.097
Nb. of grant cycles	105,861	109,715	104,190	77,960
Nb. of cycles with no acknowledging publication	15,269	15,850	14,297	11,354

Notes: Estimates stem from a linear regression in which the unit of analysis is the grant cycle and the dependent variable is lagged risk, the measure computed before the cycle began. For a grant's first cycle, it is computed on the investigator's publications of the five years before the grant's first year; for a later cycle, it is the measure of the grant's previous cycle (Appendix G gives the details). The independent variable is an indicator for a cycle with no publication acknowledging the grant, whose own risk cannot be measured. The controls are those of Table 2 except the grant output controls, which describe the cycle's own publications, and with the investigator's career citations counted through the year before the cycle's first year instead of the year before its evaluation; every regression also includes an indicator for later cycles and indicators for the number of publications behind the lagged measure, separately for first and later cycles. Grants first observed at a renewal, whose cycle number is unknown, are excluded. Lagged risk is measured in units of 10 percentile points for extreme tail outcomes and disruption, and as a share of MeSH descriptor pairs for pivoting and standing out. A positive coefficient means that cycles without a publication come from riskier records.

Standard errors in parentheses, clustered by investigator and by grant. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Table A3: The Penalty on a Sample Common to the Four Measures, and by Frequency of the Investigator's Name

	Full sample	Uniform sample	Below-median name frequency	Above-median name frequency
Extreme tail outcomes				
<10%ile	0.051**	0.051**	0.049**	0.053**
	(0.006)	(0.006)	(0.008)	(0.008)
>90%ile	-0.032**	-0.030**	-0.030**	-0.035**
	(0.006)	(0.006)	(0.008)	(0.008)
Disruption				
<10%ile	0.023**	0.039**	0.023**	0.024**
	(0.005)	(0.007)	(0.007)	(0.007)
>90%ile	-0.042**	-0.046**	-0.034**	-0.049**
	(0.005)	(0.006)	(0.007)	(0.007)
Pivoting				
<10%ile	0.008	0.013†	0.001	0.015*
	(0.005)	(0.007)	(0.007)	(0.007)
>90%ile	-0.037**	-0.040**	-0.036**	-0.040**
	(0.005)	(0.007)	(0.007)	(0.007)
Standing out				
<10%ile	0.031**	0.038**	0.032**	0.030**
	(0.006)	(0.007)	(0.008)	(0.008)
>90%ile	-0.025**	-0.029**	-0.027**	-0.020*
	(0.006)	(0.007)	(0.008)	(0.008)

Notes: Estimates stem from a linear probability model in which the unit of analysis is the grant cycle and the dependent variable is an indicator for whether the grant cycle was renewed, with the controls of Table 2. Each column and each measure come from a separate regression. The independent variables are indicators for the bottom and the top decile of each measure, computed within evaluation year on every cycle where the measure is defined; the middle eight deciles are the omitted category. Column 1 is the full sample of Table 2. Column 2 restricts it to the 69,826 cycles for which all four measures are defined. Columns 3 and 4 restrict it to the cycles whose investigator's last name is less, or more, frequent among *PubMed* authors than the median, taken over the cycles with an acknowledging publication (41,408 to 50,593 cycles per column).

Standard errors in parentheses, clustered by investigator and by grant. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Table A4: The Penalty and the Investigator's Attachment to NIH Funding

	Full sample	Below-median share of NIH-funded papers	Above-median share of NIH-funded papers
Extreme tail outcomes			
<10%ile	0.051**	0.052**	0.050**
	(0.006)	(0.008)	(0.009)
>90%ile	-0.032**	-0.031**	-0.029**
	(0.006)	(0.009)	(0.008)
Disruption			
<10%ile	0.023**	0.020**	0.027**
	(0.005)	(0.007)	(0.008)
>90%ile	-0.042**	-0.039**	-0.046**
	(0.005)	(0.007)	(0.007)
Pivoting			
<10%ile	0.008	0.001	0.014†
	(0.005)	(0.007)	(0.007)
>90%ile	-0.037**	-0.026**	-0.050**
	(0.005)	(0.007)	(0.008)
Standing out			
<10%ile	0.031**	0.031**	0.032**
	(0.006)	(0.008)	(0.008)
>90%ile	-0.025**	-0.019*	-0.034**
	(0.006)	(0.008)	(0.009)

Notes: Estimates stem from a linear probability model in which the unit of analysis is the grant cycle and the dependent variable is an indicator for whether the grant cycle was renewed, with the controls of Table 2. Each column and each measure come from a separate regression. The independent variables are indicators for the bottom and the top decile of each measure, computed within evaluation year on every cycle where the measure is defined; the middle eight deciles are the omitted category. Column 1 is the full sample of Table 2. Columns 2 and 3 restrict it to the cycles whose investigator has a share of publications listing an NIH grant below, or above, the median of 74 percent, taken over the cycles with an acknowledging publication (40,332 to 49,708 cycles per column). The share is computed over the investigator's original research publications in English in Author-ity published from 1980 to the year before the cycle's first year, from the grants listed in their *PubMed* records; cycles whose investigator has no such publication are excluded from columns 2 and 3.

Standard errors in parentheses, clustered by investigator and by grant. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Appendix B: Dissecting the Extreme Tail Outcomes Measure

Our first measure of risk taking, extreme tail outcomes, is the difference between the highest and the lowest citation percentile among the publications that acknowledge a grant during a cycle. Its logic is that a project with a wide distribution of possible outcomes will, more often than a safe one, yield both a publication that is cited a great deal and one that is barely cited at all. That logic treats the publications of a grant cycle as independent draws from the distribution of outcomes of the work the grant funded. They are not. A single project usually yields several publications, and the way an investigator divides a body of work into papers can widen or narrow the range without any change in the risk of the underlying work.[1] This appendix takes the measure apart to learn how much of its association with renewal survives that objection.

The measure draws only on a record that reviewers can examine when they evaluate the renewal application. The publications are those that acknowledge the grant and appeared by the end of the cycle's last funded year; nothing published after the renewal decision enters. From this record, reviewers can judge which publications are notable and which are not, which lines of work open new avenues and which have reached a dead end. We measure the standing of each publication by the citations it receives in the year it appeared and the five that follow, ranked against all publications of the same year. For a publication that appears late in the cycle, most of these citations are made after the decision, so the percentile is best understood as a proxy for the reviewers' reading of the record: it records what the field came to make of the work that the reviewers had in front of them. Formally, let P_c denote the publications that acknowledge the grant and appeared by the end of cycle c's last funded year, k_p the number of citations that publication p, which appeared in year t, receives from *PubMed* publications of years t to $\min(t+5, 2025)$, and Y_t the *PubMed* publications of year t that OpenAlex covers. Then

$$q_p = 100 \cdot \frac{\#\{p' \in Y_t : k_{p'} < k_p\} + \frac{1}{2}\#\{p' \in Y_t : k_{p'} = k_p\}}{|Y_t|}, \qquad \text{Extreme tail outcomes}_c = \max_{p \in P_c} q_p - \min_{p \in P_c} q_p,$$

for cycles with at least two publications that have a percentile.

We proceed in two steps. We first separate the level of a cycle's citation outcomes from their spread, and the upper half of the spread from its lower half. We then group each cycle's publications into projects, the separate lines of work within the grant, and ask whether the weakest publication tends to be a minor piece of a project that did well or the remains of a line of work that did badly.

Level and spread. The range is the sum of two half-spreads: the highest percentile minus the median (the upper half) and the median minus the lowest percentile (the lower half). In Table B1, we enter these pieces in turn, with the controls of Table 2, in units of 10 percentile points. We compute them twice: on every publication that acknowledges the grant, and on those on which the investigator is first, last, or single author.[2] Both versions use the same 64,732 cycles: those with at least three publications on which the investigator is first, last, or single author. Three is the minimum for this exercise. With two publications, the median lies halfway between them, so the upper and the lower half-spreads are always equal, and their associations with renewal could not be told apart.

Columns 1 to 5 use every publication that acknowledges the grant. Entered alone, as in the main text, the range is associated with a lower probability of renewal (column 1, −0.0138). Holding the median fixed cuts that coefficient in half (column 2, −0.0069): a cycle with a wide range tends to have a weaker typical publication, and part of the measure's association with renewal reflects the level of citations rather than their dispersion. Columns 3 and 4 enter each half alone. The upper half appears heavily penalized (−0.0402), but only because it is strongly and negatively correlated with the median (−0.78): when the best publication lies far above the median, the median tends to be low. Column 5 enters the median and both halves together. Holding the typical publication fixed, the distance between the best publication and the median is unrelated to renewal (0.0042, not significant), while the distance between the median and the weakest publication is associated with a lower

[1] We are grateful to the Department Editor for pressing this point.

[2] The publications on which the investigator is first, last, or single author report the lines of work she leads; middle-author publications more often report work led by others.

probability of renewal (-0.0082). The association comes mostly from the lower tail. On all 74,914 cycles with at least three publications, the pattern of column 5 is the same (the two halves at -0.0012 and -0.0079).

Columns 6 to 10 repeat the exercise on the publications on which the investigator is first, last, or single author. The lower half behaves as before, the upper half does not. Entered alone, the range is again associated with lower renewal (-0.0095), but holding the median fixed leaves little of that association (-0.0019, significant at the 10 percent level only). With the median and both halves (column 10), the weakest publication's distance below the median is associated with lower renewal (-0.0039), and the best publication's distance above it with higher renewal (0.0124). Renewal does not fall with extraordinary success; among first- and last-author publications, it rises with it.

The table describes how renewal varies with the level and the dispersion of citation outcomes, but it cannot reveal how reviewers weigh a hit against a miss of the same size, because percentile points are not comparable across the scale. The best publication of a cycle averages the 93$^{\text{rd}}$ percentile and the weakest the 53$^{\text{rd}}$, and a percentile point near the top of the distribution represents many more citations than a point near its middle.

From publications to projects. The lower tail is also where the objection raised at the outset, that the publications of a grant cycle are not independent draws, would matter most. A publication far below the cycle's median could be the remains of a line of work that failed, which is what our interpretation of the measure requires, or a minor piece cut from a project that succeeded. The two stories make different predictions about where the weakest publication is found: in a project of its own in the first case, inside a successful project in the second. Telling them apart requires knowing which publications belong to the same project, and no source records it. We approximate projects in two ways that share no information.

The first relies on content. Publications from the same project share their specific MeSH descriptors (the protein, the disease, the method), while publications from different projects of the same laboratory share mostly broad ones (the organism, the field). We link two publications of a cycle when the cosine similarity of their descriptors, each weighted by its inverse document frequency over *PubMed*, is at least τ, and we define a project as a set of publications connected by a chain of links; a publication linked to no other publication, or without MeSH descriptors, is a project of its own. The second relies on people. In a biomedical laboratory, a line of work usually belongs to one trainee, who is first author on its publications, so we also group a cycle's publications by first author.

The threshold τ matters, and we chose it by examining what each value produces before estimating anything. At $\tau = 0.7$, 98.5 percent of projects are single publications; at $\tau = 0.5$, 90 percent are, and the weakest publication shares its project with another publication in only 18 percent of cycles. Thresholds that strict turn nearly every publication into a project of its own, so the decomposition below would classify almost every weak publication as a failed line of work by construction. At $\tau = 0.3$, projects average 2.1 publications, 14 percent of cycles consist of a single project, and the weakest publication shares its project with another in 55 percent of cycles. The first-author grouping yields projects of 1.46 publications on average, single-project cycles in 3 percent of cases, and a weakest publication that shares its project in 44 percent of cycles. Built from unrelated information, the two definitions produce projects of similar size and composition. We use $\tau = 0.3$ and the first-author grouping throughout.

Failed bets and slicing. Take the project that contains the cycle's weakest publication, and call that project's best publication its lead. The lead need not be the cycle's best publication, which enters only the upper half-spread. The lower half-spread then divides exactly into two parts:

$$\underbrace{\text{median} - \text{min}}_{\text{lower half-spread}} = \underbrace{(\text{median} - \text{lead})}_{\text{between projects}} + \underbrace{(\text{lead} - \text{min})}_{\text{within the project}} .$$

As in the decomposition of a variance into its between-group and within-group parts, the lead enters both parts with opposite signs, so the split is an identity; what gives each part its meaning is the reference point. The between part measures how far the weakest publication's whole line of work falls below the cycle's typical publication; it is large when a line of work failed. The within part measures how far the weakest publication falls below its own project's best; it is large when the weakest publication is a minor piece of a project that did well. We use the project's best publication rather than its average, because the average would be pulled down by the weakest

publication itself. Two cycles illustrate the split. In the first, project A produced two publications, at the 95^{th} and 50^{th} percentiles, and project B three, at the 80^{th}, 70^{th} and 40^{th}. The median of the five publications is 70. The weakest publication, at the 40^{th} percentile, belongs to project B, whose lead is the publication at the 80^{th}. Its lower half-spread, $70 - 40 = 30$, therefore splits into a between part of $70 - 80 = -10$ and a within part of $80 - 40 = 40$. The between part is negative because the weakest publication belongs to a line of work whose lead lies above the median; the whole downside comes from how far the weakest publication falls below that lead. This is slicing. The cycle's best publication, at the 95^{th} percentile, belongs to project A and plays no part in the split: it enters only the upper half-spread, $95 - 70 = 25$. In the second cycle, projects A (95^{th} and 85^{th} percentiles) and B (75^{th} and 70^{th}) are joined by a third project, C, whose only publication is at the 40^{th} percentile. The median of the five publications is now 75, and the weakest publication is C's only one, which is therefore its own lead. The lower half-spread, $75 - 40 = 35$, lies entirely between projects ($75 - 40 = 35$) and nothing lies within ($40 - 40 = 0$). This is a failed bet. When several publications tie for the lowest percentile, we use the one whose project has the strongest lead, which attributes as much of the downside as possible to slicing.

In Table B2, we replace the lower half-spread with its two parts, on the cycles with at least two projects, and add indicators for the number of projects, since a weak publication is more likely to stand alone when a cycle contains more of them. Per 10 percentile points, the between and the within parts are associated with similar reductions in the probability of renewal: -0.0047 and -0.0057 when projects are defined by content, -0.0080 and -0.0073 when they are defined by first author. Equality cannot be rejected under either definition ($p = 0.29$ and $p = 0.49$). For comparison, the last row gives the coefficient on the lower half-spread when it is not split: the specification of column 5 of Table B1, estimated on the same cycles with the same indicators (-0.0051 and -0.0077).

The two parts are also of similar size. On average, the between part amounts to 12 percentile points and the within part to 17 when projects are defined by content, and to 15 and 14 when they are defined by first author. Together with the coefficients, these magnitudes imply that roughly half of the measure's downside reflects how successful projects were published and half reflects lines of work that did badly. The between part, finally, carries its association with renewal mainly when the weakest publication is one on which the investigator is first or last author (-0.0055, against -0.0026 and not significant when the investigator is a middle author, with projects defined by content), which is what one would expect if it captures the failure of lines of work the investigator leads.

Interpretation. The objection raised at the outset, that the publications of a grant cycle are not independent draws, is right in an important respect: part of what the extreme tail measure records is how investigators publish the output of a successful project. That part is associated with lower renewal, and we do not attribute it to risk taking. But it is not the whole measure. Lines of work that did badly as a whole account for about as much of the measure's lower tail and are associated with renewal in about the same proportion, and most grant cycles contain several lines of work rather than one. Nor does the measure's association with renewal come from its upper tail: holding the typical publication fixed, extraordinary success is not associated with a lower probability of renewal. The measure therefore captures risk taking, though not risk taking alone. Two features of the data suggest that, if anything, it understates the failures that risk taking produces. The measure registers only the failures that reach publication, while many failed projects never leave the file drawer; and the cycles that produce no acknowledging publication at all come from investigators whose earlier record was slightly riskier on this measure.

Table B1: Level and Spread of Citation Outcomes within the Grant Cycle

	All publications					First- or last-author publications				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Max − Min	-0.0138**	-0.0069**				-0.0095**	-0.0019†			
	(0.0010)	(0.0011)				(0.0010)	(0.0011)			
Median Citation Percentile		0.0377**			0.0448**		0.0387**			0.0473**
		(0.0019)			(0.0029)		(0.0018)			(0.0025)
Max − Median			-0.0402**		0.0042			-0.0347**		0.0124**
			(0.0023)		(0.0036)			(0.0021)		(0.0033)
Median − Min				-0.0070**	-0.0082**				-0.0027*	-0.0039**
				(0.0012)	(0.0012)				(0.0011)	(0.0011)
H_0: max − median = median − min					$p = 0.001$					$p = 0.000$
Mean nb. of publications behind the measures	9.3	9.3	9.3	9.3	9.3	7.0	7.0	7.0	7.0	7.0
Mean of dependent variable	0.488	0.488	0.488	0.488	0.488	0.488	0.488	0.488	0.488	0.488
Nb. of investigators	27,466	27,466	27,466	27,466	27,466	27,466	27,466	27,466	27,466	27,466
Nb. of grant cycles	64,732	64,732	64,732	64,732	64,732	64,732	64,732	64,732	64,732	64,732

Notes: Estimates stem from a linear probability model in which the unit of analysis is the grant cycle and the dependent variable is an indicator for whether the grant cycle was renewed, with the controls of Table 2. The variables describe the citation percentiles (five-year citations, ranked within publication year) of the publications that acknowledge the grant during the cycle: the median, the highest minus the median, and the median minus the lowest, in units of 10 percentile points. The two half-spreads add up to the extreme tail outcomes measure, the highest minus the lowest percentile, which columns 1 and 6 enter alone, as in the main text, and columns 2 and 7 with the median. Columns 1 to 5 compute the variables on all these publications, columns 6 to 10 on those on which the investigator is first, last, or single author. Every column uses the same cycles: those with at least three first- or last-author publications with a citation percentile and a mean journal impact factor.

Standard errors in parentheses, clustered by investigator and by grant. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Table B2: Failed Bets and Slicing: The Lower Tail Between and Within Projects

	(1) Similar MeSH ($\tau = 0.3$)	(2) Same First Author
Median Citation Percentile	0.0364**	0.0381**
	(0.0028)	(0.0027)
Max − Median	0.0027	-0.0018
	(0.0035)	(0.0033)
Median − Min, Between Projects	-0.0047**	-0.0080**
	(0.0012)	(0.0012)
Median − Min, Within Projects	-0.0057**	-0.0073**
	(0.0013)	(0.0013)
H_0: between = within	$p = 0.292$	$p = 0.492$
Median − Min, Unsplit (Same Cycles)	-0.0051**	-0.0077**
	(0.0012)	(0.0011)
Mean of dependent variable	0.452	0.468
Nb. of investigators	28,080	29,835
Nb. of grant cycles	64,057	72,425

Notes: Estimates stem from a linear probability model in which the unit of analysis is the grant cycle and the dependent variable is an indicator for whether the grant cycle was renewed, with the controls of Table 2, on the cycles with at least three publications with a citation percentile and at least two projects. Projects group the cycle's publications with a citation percentile. Column 1: two publications are linked when the cosine similarity of their MeSH descriptors, each weighted by its inverse document frequency over *PubMed* 2025, is at least 0.3, and a project is a set of publications joined by a chain of links. Column 2: publications with the same first author. In both, a publication linked to no other is a project of its own. The lower half-spread, the median minus the lowest citation percentile, is split at the best publication (the lead) of the weakest publication's project: the part between projects is the median minus that lead, the part within the project is the lead minus the lowest percentile. When several publications tie for the lowest percentile, the project with the highest lead is used. Every regression includes indicators for the number of projects (2, 3, 4, 5 to 6, 7 or more). The last row comes from a separate regression, that of Table B1, column 5, with the same indicators and on the same cycles, in which the lower half-spread is not split. Units of 10 percentile points.

Standard errors in parentheses, clustered by investigator and by grant. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Appendix C: Dissecting the Disruption Measure

Our second measure of risk taking rests on a distinction between two kinds of contribution. Some research consolidates a line of work: it refines, extends, or confirms what came before, and those who build on it keep building on its predecessors as well. Other research breaks with the line of work it came from: those who build on it no longer need what preceded it. An investigator who sets out to overturn the prevailing account of a problem places a bet of the second kind, and the bet may or may not pay off. Funk and Owen-Smith (2017) proposed a way to tell the two kinds of contribution apart from citations, first for patents; Wu et al. (2019) brought it to scientific publications. Our measure of disruption for a grant cycle is built from this index: each publication that acknowledges the grant during the cycle receives a value, ranked against all *PubMed* publications of the same year, and the cycle's measure is the highest of these ranks.

This appendix explains how the index is computed, what a high value does and does not mean, and how a publication's index becomes a grant cycle's measure. It then takes up the measurement problems that the index raises, some known since it was proposed and some specific to how we use it. Each has a remedy in the form of an alternative version of the measure. We describe the version that addresses each problem, and then ask whether the association between disruption and renewal holds in every version.

The idea. Take a publication f together with the later publications that cite it, its *citers*, and the earlier publications that it cites, its *references*. The index rests on one question about each citer: does it also cite any of f's references? Figure C1 illustrates. A citer that cites f together with one or more of its references treats f as one more contribution to an existing line of work; in its author's eyes, f has not displaced its predecessors. A citer that cites f but none of its references treats f as a point of departure: for that citer, f has made what came before it unnecessary. Call N_i the number of citers of the second kind and N_j the number of the first. The index is

$$D = \frac{N_i - N_j}{N_i + N_j},$$

the share of citers of the second kind minus the share of the first. It equals $+1$ when no citer cites any of f's references, -1 when every citer does, and 0 when the two kinds are equally numerous. In Figure C1, two of the three citers ignore f's references and one does not, so that $D = (2-1)/(2+1) = 1/3$. A publication is called *disruptive* when its index is high and *consolidating* when it is low. The index depends on the number of citers only through the shares: a publication with ten citers, eight of which ignore its references, has the same index as one with a thousand citers, eight hundred of which do.

What a high value does and does not mean. Two publications illustrate. The first is BLAST, the method for searching sequence databases of Altschul et al. (1990). Of the 66,261 *PubMed* publications that cite it, 95 percent cite none of its ten references, and its index is 0.90, higher than that of nine in ten publications of 1990. Those who used BLAST to search sequence databases had little reason to cite the earlier work on which it built, which is the pattern the index was designed to detect. The second is the American Cancer Society's report of cancer statistics for 2010. Of its 10,455 citers, 99 percent cite none of its 14 references, and its index is 0.97, higher than that of 94 percent of the publications of 2010. An annual compilation of statistics is not disruptive in any sense that matters here. It is cited as a source of numbers, and those who cite it have no reason to cite the sources of its sources; Bornmann et al. (2020) found that these reports reach the maximum of all four versions of the index they examined. The index reflects how the field uses a publication, not what kind of contribution the publication makes. A publication that displaces its predecessors is used in this way, but so is a compilation or a database cited for the facts it contains. Like each of our measures, disruption is therefore a proxy, with noise as well as signal.

How well the index tracks displacement as experts would judge it is not settled. Wu et al. (2019) found that publications that won a Nobel Prize score high on it (as reported by Bornmann et al. 2020 and Lin et al. 2026). Bornmann et al. (2020) compared several versions of the index with the tags that the experts of the F1000Prime service attach to biomedical publications, such as "new finding" or "technical advance," and found weak agreement; since the tags capture novelty rather than disruption, they concluded that "we can't directly make claims on whether disruption indicators measure what they propose to measure." No version of the index

has been validated against an independent judgment of displacement. Our use of it rests on what it reflects, the behavior of the citers, and on the robustness of its association with renewal to the choices made in computing it.

The index in our data. We compute the index for every publication in *PubMed* from 1950 to 2025, 39 million in all, from the citation links between *PubMed* publications recorded in OpenAlex (snapshot of February 2026), about 822 million links. Only references and citers that are themselves in *PubMed* enter the calculation. The index is defined when a publication has at least one reference in *PubMed* and at least one citer, which is the case for 24.8 million publications. It is defined for 90 percent of the 860,249 publications that acknowledge a grant of our sample: 76 percent of those published in the 1980s, and 96 to 97 percent of those published from 2000 to 2019.

From a publication to a grant cycle. The level of the index drifts over time, for reasons that have little to do with how disruptive science is. Reference lists have lengthened, more of them are recorded in machine-readable form, and the volume of research has grown; each of these changes how likely a citer is to cite a publication's references. Park et al. (2023) document a long decline in the index; Petersen et al. (2025) attribute it to the growth of research and the lengthening of reference lists rather than to a decline in innovation. To compare publications of different years, we therefore rank each publication's index against those of all *PubMed* publications of the same year, as a percentile from 0 to 100.

A grant cycle's measure is the highest of these percentiles among the publications that acknowledge the grant during the cycle, and it is defined only for the cycles with at least one such publication for which the index can be computed. We take the highest because a grant that pursues a disruptive line of work need not produce a disruptive publication every time; one publication that displaces its predecessors is the mark of the bet. NIH-funded publications are well cited and have long reference lists, and they rank low on the index: the median acknowledging publication is at the 34$^{\text{th}}$ percentile of its year, and only one in ten is above the 73$^{\text{rd}}$. A cycle's highest percentile averages 0.60, on a scale from 0 to 1, with a standard deviation of 0.24.

The highest of several values rises with their number, so cycles with more publications score higher: in the top decile of the measure, cycles have 10.7 acknowledging publications with an index on average, against 2.5 in the bottom decile. Renewed cycles publish more (7.3 such publications, against 5.6), and in the raw data the two facts offset each other: the measure averages 0.60 in renewed and in non-renewed cycles alike. This is one reason why, throughout the paper, we compare cycles that produced the same number of publications, of similar quality (the controls of Table 2).

Measurement problems, and the versions of the measure that address them. We now take up six problems. Each corresponds to one or more columns of Tables C1 and C2, whose results we describe afterward. Every version follows the steps just described: an index for each acknowledging publication, its percentile among the *PubMed* publications of the same year for which that version is defined, and the highest percentile in the cycle.

Publications that cite the references but not the publication. The index as Funk and Owen-Smith (2017) and Wu et al. (2019) defined it has a third term in its denominator,

$$D^k = \frac{N_i - N_j}{N_i + N_j + N_k},$$

where N_k counts the publications that cite at least one of f's references but not f itself (publication X in Figure C1). The intent is to lower the score of a publication that leaves much of its field's attention on its predecessors. But N_k counts publications that are not citers of f, and it is large: for the median *PubMed* publication of 2010, it is 216 times the number of the publication's citers. The index with N_k equals the index without it divided by one plus the ratio $N_k/(N_i + N_j)$, a number in the hundreds. Lin et al. (2026) show that this ratio is governed by how many citations f's most-cited reference receives relative to f itself. The index with N_k therefore ranks a publication by two things at once: how far its citers ignore its references, and how well it competes for citations with its most famous reference. Lin, Li, and Wu read the combination as a virtue, since displacing a dominant idea is what a breakthrough does. For our purpose, it mixes what we want to keep apart. A measure of the risk an investigator took should reflect what the citers of her publications did; the volume of citations her publications received is a separate part of her record, which our controls and our measure of extreme

tail outcomes treat on their own terms. Bornmann et al. (2020) make the same point: N_k "does not refer to" the distinction between citers that do and do not cite the publication's references, and "captures the citation impact" of the publication instead. Wu and Yan (2019) proposed the version without N_k, Bu et al. (2021) argue for keeping disruption and citation impact separate, and the version without N_k is the one we use, because it reflects the behavior of a publication's citers and nothing else. Column 2 of Table C1 uses the index with N_k.

How much overlap makes a citer consolidating. A citer that shares a single reference with f may do so by coincidence, particularly when that reference is a well-known publication that much of the field cites. Bornmann et al. (2020) therefore also examine a stricter version, in which a citer counts as consolidating only if it cites at least five of f's references, while citers that cite one to four of them leave the index altogether. The stricter version has two costs. A publication with fewer than five references cannot have such a citer, so the version is defined only for publications with at least five references. And many more publications reach its maximum: 29 percent of the *PubMed* publications of 2010 score $+1$, against 6 percent with a threshold of one, so that a large block of publications of the same year cannot be ranked against each other, and the version's percentile cannot rise above about 85. Columns 3 and 4 of Table C1 use the stricter threshold, without and with N_k.

Which citers count. The index counts every citer recorded by early 2026, and four restrictions on the citers address four concerns. The first concerns timing. Most citers of a publication appear years after the renewal decision, so the index partly reflects what happened after it. Columns 5 and 6 count only the citers published within five and ten years of the publication; five years is also the window of our citation percentiles. The second concerns self-citation. Authors who cite their own earlier publication tend to cite its references as well, which makes that publication look consolidating. This matters here because renewal keeps a laboratory working on the same problem: a renewed grant's publications may receive more citations of this kind from the laboratory's later work, while a laboratory that loses its grant may abandon the line of work. Renewal itself, by adding such self-citations, would then make a cycle's publications look less disruptive. In column 7, we drop every citer that shares at least one author with f, as identified by OpenAlex. The third concerns reviews, which cite a publication together with its predecessors as a matter of course; in column 8, we keep only the citers that are original research. The fourth concerns references we cannot see. A citer is known to cite f only because its reference list is recorded, but the list may be recorded in part, and a reference missing from it cannot make the citer consolidating; incomplete lists therefore make a publication look more disruptive. In columns 9 to 11, we keep only the citers with at least 5, 10, and 20 references in *PubMed*, whose lists we observe well; these restrictions also select on field and era, so we report the three together.

Short reference lists. A publication with few references gives its citers few opportunities to cite one of them, and therefore tends to look disruptive; BLAST and the cancer statistics report have 10 and 14. Within a year, the index correlates -0.25 to -0.34 with the number of a publication's references (rank correlations), more strongly than with any other characteristic of the publication that we examined. In column 3 of Table C2, we rank each publication only against the *PubMed* publications of the same year with a similar number of references, before the cycle's highest is taken.

Few or weak citations. A publication with few citers can reach the ends of the index by chance: three citers that happen to ignore its references give it the maximum, $+1$, and among the *PubMed* publications of 2000 to 2010 at the maximum, half have three citers or fewer. Weakly cited publications are therefore overrepresented at the top of the index. Among the publications that acknowledge a grant of our sample, 33 percent of those in the bottom decile of five-year citations are in the top decile of disruption, against 0.4 percent of those in the top decile of citations. Since a cycle's measure is its highest percentile, a weakly cited publication can set it, and the measure would then reflect a weak citation record as well as disruption. Two versions hold citations fixed. In column 2 of Table C2, we rank each publication only against the publications of the same year with a similar number of citers, before the cycle's highest is taken. In column 4, we keep the measure of the main text and add to the regression the median five-year citation percentile of the same publications.

How a cycle's publications are summarized. The highest percentile is one summary among several. In column 5 of Table C2, we add indicators for the number of acknowledging publications with an index, among which the highest is taken, to Table 2's indicators for the number of all acknowledging publications. In column 6, we take the highest among the publications on which the investigator is first, last, or single author. In columns 7 to 9, we replace the highest percentile with the mean percentile, with the share of the publications whose percentile is

75 or more (a share does not rise with the number of publications), and with the highest value of the index itself rather than of its percentile.

Results. In Tables C1 and C2, we rerun the disruption regression of Table 2 with each version in place of the measure of the main text, with the same controls. Because the versions are distributed differently, each enters linearly and standardized on the regression's cycles, so that every coefficient is the change in the probability of renewal for one standard deviation of the version.[1] Every column of Table C1 uses the same 98,400 cycles, those on which all eleven versions are defined (99 percent of the cycles of Table 2); Table C2 uses the 99,504 cycles of Table 2.

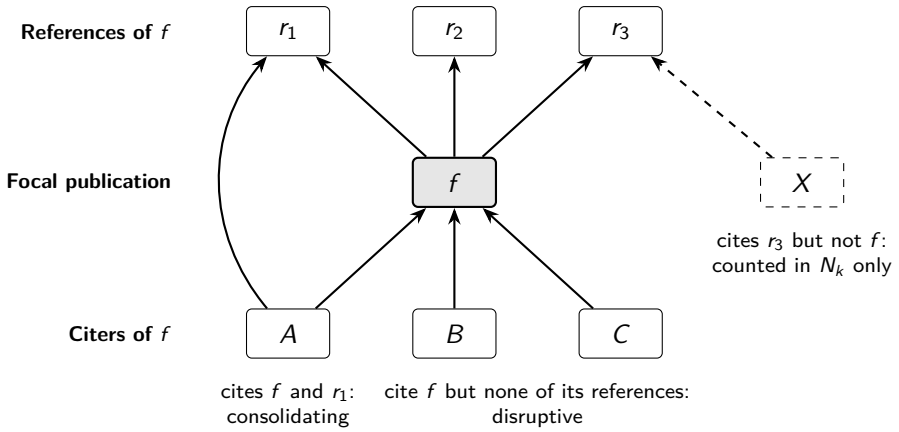
In Table C1, a one-standard-deviation increase in the measure of the main text is associated with a probability of renewal lower by 3.1 percentage points, 7 percent of the mean renewal rate (column 1). Every other version of the index yields a negative association, significant at the 1 percent level, of 2.5 to 3.5 points; the smallest of these comes from setting aside the citations a publication receives from its own authors (column 7). The one exception combines the stricter threshold with N_k (column 4), at 0.7 point; it is also the version least correlated with ours (0.66).

In Table C2, the measure of the main text yields 3.0 points on its cycles (column 1). Holding fixed the citations of the same publications reduces the association by a tenth to a fifth (columns 2 and 4), less than for extreme tail outcomes, half of whose association with renewal reflects the level of citations (Appendix B). Every other version yields a negative association, significant at the 1 percent level, of 1.7 to 3.2 points (columns 3 and 5 to 9).

Interpretation. The disruption index is an imperfect proxy for the displacement of earlier work. It reflects how a publication is used, which a compilation of statistics can share with a breakthrough; its value depends on how its terms are defined; and it is entangled with the number of a publication's references and citations. Taking it apart does not undo the association that Table 2 reports. In every version, more disruptive grant cycles are less likely to be renewed.

[1] Table 2 enters the measure as indicators for its bottom and top deciles. In that form, the top decile is penalized as much whichever citers are counted (coefficients between -0.042 and -0.046, against -0.043 for the measure of the main text). In the two versions with N_k, the top decile is not associated with renewal: by dividing the index by a number that falls as a publication's own citations rise, these versions place well-cited publications at the top of the ranking. In their top decile, the publication that sets a cycle's measure is at the 72$^{\text{nd}}$ percentile of five-year citations on average, against the 53$^{\text{rd}}$ for the measure of the main text. With the stricter threshold of five references, ties at the maximum leave too few cycles above the 90$^{\text{th}}$ percentile to form a top decile. Holding citations fixed also matters more for the top decile than for the linear form: with the median citation percentile of the same publications in the regression, the coefficient on the top decile falls from -0.042 to -0.022.

Figure C1: How the Disruption Index Is Computed



$$D = \frac{N_i - N_j}{N_i + N_j} = \frac{2-1}{2+1} = \frac{1}{3}$$

Notes: An arrow points from a citing publication to the publication it cites. The focal publication f cites three references, r_1 to r_3, and is cited by three later publications, A, B, and C. Citer A also cites one of f's references, r_1, and counts as consolidating ($N_j = 1$); citers B and C cite none of f's references and count as disruptive ($N_i = 2$). The index is the share of disruptive citers minus the share of consolidating ones. Publication X cites one of f's references but not f: it is not a citer of f, and it enters only the version of the index with the third term N_k in its denominator, which our measure leaves out.

Table C1: Versions of the Disruption Index and Grant Renewal

	The index				The citers						
	As in Table 2	With N_k	Five shared references	Five shared, with N_k	Citers within 5 years	Citers within 10 years	No self-citations	Original-research citers	Citers with 5+ references	Citers with 10+ references	Citers with 20+ references
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
One s.d. of the measure	-0.031**	-0.032**	-0.035**	-0.007**	-0.027**	-0.032**	-0.025**	-0.032**	-0.031**	-0.031**	-0.030**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Rank correlation with column 1	1.00	0.88	0.76	0.66	0.83	0.92	0.93	0.95	1.00	1.00	0.98
Mean of dependent variable	0.424	0.424	0.424	0.424	0.424	0.424	0.424	0.424	0.424	0.424	0.424
Nb. of investigators	38,358	38,358	38,358	38,358	38,358	38,358	38,358	38,358	38,358	38,358	38,358
Nb. of grant cycles	98,400	98,400	98,400	98,400	98,400	98,400	98,400	98,400	98,400	98,400	98,400
Adjusted R^2	0.1723	0.1724	0.1731	0.1697	0.1717	0.1725	0.1712	0.1724	0.1723	0.1723	0.1721

Notes: Estimates stem from a linear probability model in which the unit of analysis is the grant cycle and the dependent variable is an indicator for whether the grant cycle was renewed, with the controls of Table 2. Each column replaces the disruption measure of Table 2 with another version. In every version, each publication that acknowledges the grant during the cycle receives an index, ranked against all *PubMed* publications of the same year for which that version is defined, and the cycle's measure is the highest of these percentiles. The measure enters standardized on the column's cycles, so that its coefficient is the change in the probability of renewal for one standard deviation. Column 1 is the measure of Table 2: a citer counts as consolidating if it also cites at least one of the publication's references, and the index compares the numbers of consolidating and other citers (Figure C1). Column 2 adds to the denominator the publications that cite the publication's references but not the publication (N_k). Columns 3 and 4: a citer counts as consolidating only if it cites at least five of the publication's references, citers that cite one to four of them leave the index, and the index is defined for publications with at least five references; column 4 adds N_k. Columns 5 to 11 compute the index of column 1 on a subset of the citers: those published within five or ten years of the publication; those that share no author with it; those that are original research (no review or comment); and those with at least 5, 10, or 20 references in *PubMed*. Every column uses the same cycles, those of Table 2 on which every version is defined. The rank correlation (Spearman) is between the version and column 1 on these cycles.

Standard errors in parentheses, clustered by investigator and by grant. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Table C2: Other Versions of the Disruption Measure and Grant Renewal

	The highest percentile						Other summaries		
	As in Table 2	Ranked within citation bin	Ranked within reference bin	With median citation percentile	With publications with an index	First- or last-author publications	Mean percentile	Share in the top quartile	Highest raw index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
One s.d. of the measure	-0.030**	-0.028**	-0.030**	-0.024**	-0.032**	-0.024**	-0.029**	-0.017**	-0.029**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Median citation percentile, per 10 points				0.031**					
				(0.001)					
Rank correlation with column 1	1.00	0.97	0.91	1.00	1.00	0.90	0.73	0.79	0.99
Mean of dependent variable	0.423	0.423	0.423	0.423	0.423	0.431	0.423	0.423	0.423
Nb. of investigators	38,791	38,791	38,791	38,791	38,791	37,326	38,791	38,791	38,791
Nb. of grant cycles	99,504	99,504	99,504	99,504	99,504	94,793	99,504	99,504	99,504
Adjusted R^2	0.1717	0.1712	0.1717	0.1773	0.1720	0.1694	0.1718	0.1702	0.1716

Notes: Estimates stem from a linear probability model in which the unit of analysis is the grant cycle and the dependent variable is an indicator for whether the grant cycle was renewed, with the controls of Table 2. Each measure enters standardized on the column's cycles, as in Table C1. Column 1 is the measure of Table 2. Column 2 ranks each publication's index against the *PubMed* publications of the same year and the same number of citing publications in *PubMed* (0, 1, 2 to 4, 5 to 9, 10 to 19, 20 to 49, 50 to 99, 100 or more), column 3 against those of the same year and the same number of references in *PubMed* (1, 2, 3 to 4, 5 to 6, 7 to 9, 10 to 14, 15 to 19, 20 to 24, 25 to 29, 30 to 39, 40 to 49, 50 to 74, 75 to 99, 100 or more); in both, the cycle's measure is the highest of these percentiles. Column 4 adds to column 1 the median five-year citation percentile of the acknowledging publications with an index, in units of 10 percentile points. Column 5 adds indicators for the number of acknowledging publications with an index (1 to 5, 6 to 7, 8 to 10, 11 to 15, 16 to 20, 21 or more). Column 6 is the highest percentile among the publications on which the investigator is first, last, or single author. Column 7 is the mean percentile of the acknowledging publications, column 8 the share of them whose percentile is 75 or more, and column 9 the highest value of the index itself rather than of its percentile. Every column uses the cycles of Table 2, except column 6, which uses those of them with a first- or last-author publication with an index.

Standard errors in parentheses, clustered by investigator and by grant. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Appendix D: Dissecting the Pivoting Measure

Our third measure of risk taking, pivoting, captures the intellectual distance between the work a grant supports and the investigator's earlier work. It reflects the view that part of the risk in a research program comes from who conducts it. An investigator who takes up topics, techniques, or model organisms she has not worked with before leaves the area in which her expertise and her judgment are best calibrated; such shifts can yield discoveries, but they can also devalue her prior expertise and invite skepticism from reviewers (Arts and Fleming 2018, Hill et al. 2025). To measure pivoting, we rely on the Medical Subject Headings (MeSH), the controlled vocabulary with which the indexers of the National Library of Medicine describe the content of each publication in *PubMed*. For each grant cycle, we list the pairs of MeSH descriptors that appear together on the publications that acknowledge the grant during the cycle, and we compare them with the pairs that appear on the investigator's publications of the five years before the cycle starts (original research in English, in any authorship position). Pivoting is the share of the cycle's descriptor pairs that appear on none of those earlier publications, with a pair counted once for each of the cycle's publications on which it appears. On the 101,435 cycles of the pivoting regressions of Table 2, three quarters of a cycle's descriptor pairs are new in this sense, on average, and cycles in the top decile of the measure are 3.7 percentage points less likely to be renewed than cycles in the eight middle deciles.

Its construction involves design choices and simplifications that one ought to question.[1] Four of them stand out. The measure counts a new pair the same way whether it recombines two descriptors the investigator has long used or brings in a domain she has never touched. It treats all MeSH descriptors alike, although some designate a narrow object, such as a single protein, and others a broad domain. It treats every departure from earlier work as a bet, whereas an investigator may leave a field because the field itself is drying up, a prudent choice rather than a risky one. And it records the pivot the investigator carried out, without regard to whether she proposed it in the application that a study section reviewed and NIH funded.

We examine these choices in four steps. We first split the measure into near pivots, which recombine familiar descriptors, and far pivots, which bring in new ones. We then measure, on a continuous scale, how distant a cycle's new descriptors are from the investigator's earlier work, using the hierarchical structure of MeSH. We then ask whether pivots away from fields that were fading are penalized less, and, last, whether pivots proposed in the application are.

Near and far pivots. We call a new pair *near* when both of its descriptors already appear in the investigator's publications of the five preceding years: the pair is new, but it recombines familiar material. We call it *far* when at least one of its descriptors does not. Formally, let P_c denote the publications that acknowledge the grant during cycle c, D_p the descriptors of publication p, and Π_p the pairs of distinct descriptors on p, of which there are $|D_p|(|D_p|-1)/2$; let K_c and H_c denote the descriptors and the pairs that appear on the investigator's publications of the five years before the cycle. Then

$$\text{Pivoting}_c = \frac{\sum_{p\in P_c} |\Pi_p \setminus H_c|}{\sum_{p\in P_c} |\Pi_p|},$$

$$\text{Near}_c = \frac{\sum_{p\in P_c} \left|\{\pi \in \Pi_p \setminus H_c : \pi \subseteq K_c\}\right|}{\sum_{p\in P_c} |\Pi_p|}, \qquad \text{Far}_c = \frac{\sum_{p\in P_c} \left|\{\pi \in \Pi_p : \pi \not\subseteq K_c\}\right|}{\sum_{p\in P_c} |\Pi_p|}.$$

The shares of near and far pairs add up to the measure, since a pair with a descriptor outside K_c cannot appear on the earlier publications. Their descriptive statistics are presented in Table D1. Near pairs average 0.10 of a cycle's descriptor pairs and far pairs 0.64, so that 87 percent of new pairs bring in at least one unfamiliar descriptor. The share of far pairs is nearly the share of the cycle's descriptors that are new to the investigator (correlation 0.97), and the split therefore separates pivots that bring in new descriptors from recombinations of familiar ones. Because the split is defined against the investigator's own record, it depends on how much she published before the cycle: an investigator with few earlier publications has few familiar descriptors, and more of her new pairs are far. New investigators, whose records are thinner, have a share of far pairs of 0.72, against 0.61 for other investigators. Figure D1 shows the distributions of the two shares for renewed and non-renewed cycles. In raw

[1] We thank the Department Editor and two anonymous referees for raising these points.

form they differ little: near pairs average 0.102 of a renewed cycle's pairs and 0.098 of a non-renewed cycle's, and far pairs 0.633 and 0.644. The regressions below compare cycles that are alike in the controls of Table 2.

Table D2 enters the two shares in place of the measure, with the controls of Table 2 and indicators for the number of the investigator's earlier publications, which absorb the dependence just described. On all cycles, far pairs are associated with a lower probability of renewal (-0.113), near pairs are not (-0.026, with a standard error of 0.034), and the two coefficients differ ($p = 0.002$). At the same amount of new work, shifting 10 percent of a cycle's pairs from near to far is associated with a probability of renewal lower by 0.9 percentage points. The pattern is the same among investigators who already held an R01-equivalent grant (columns 5 and 6: -0.101 and 0.002, $p < 0.001$). Among new investigators (columns 3 and 4), both coefficients are negative and cannot be told apart (-0.140 and -0.077, $p = 0.31$); with nine of every ten new pairs far, the share of near pairs varies too little among new investigators to be estimated precisely.

How far: a graded distance. The near/far split distinguishes only two degrees of distance. To grade distance more finely, we use the hierarchical structure of MeSH. Each descriptor occupies one or more positions in a tree whose sixteen branches correspond to broad categories (Anatomy, Organisms, Diseases, Chemicals and Drugs, and so on), and each position is nested within broader ones: Mice is nested within Murinae, Murinae within Muridae, and so on up to Animals. Two descriptors are related through the most specific position under which both are found. For Mice and Rats, that position is Murinae; for Humans and the rhesus macaque, it is Catarrhini, the Old World primates.

The number of steps that separate two descriptors in the tree is a poor measure of their distance, because the branches of MeSH are not equally detailed. Mice and Rats meet at Murinae, and Humans and the rhesus macaque at Catarrhini, both nine levels below the top of the tree, yet the two meeting points differ greatly. Murinae covers 11 percent of the *PubMed* publications that carry a MeSH descriptor; Catarrhini covers 71 percent, because it includes Humans. Knowing that a publication falls under Catarrhini says little about its subject; knowing that it falls under Murinae says much more. We therefore measure how specific a position is by the share of *PubMed* publications indexed with it or with a descriptor below it, expressed as an information content: minus the natural logarithm of that share (Resnik 1995, Lin 1998).[2] A position that covers half of *PubMed* has an information content of 0.7, and one that covers one publication in a thousand an information content of 6.9. Catarrhini's information content is 0.35, Murinae's 2.23.

The distance of a new descriptor from the investigator's earlier work compares two information contents: that of the descriptor itself, and that of the most specific position it shares with any descriptor of her earlier publications. Their ratio measures how much of the new descriptor's specificity her earlier work already covers, and the distance is one minus that ratio. The upper panel of Figure D2 gives two examples. For an investigator who has published on mice, Rats shares Murinae with her earlier work, a position nearly as specific as Rats itself (2.23 against 2.93), and lies at a distance of $1 - 2.23/2.93 = 0.24$. For an investigator who has published on human subjects, the rhesus macaque shares only Catarrhini with her earlier work, a position that says little about either (0.35 against 6.59), and lies at a distance of $1 - 0.35/6.59 = 0.95$. The distance is 0 for a descriptor the investigator has used, and also for a broader term whose most specific position is above one she has used; it is 1 when her earlier work shares no position with it, not even one of the sixteen broad categories.

The lower panel of Figure D2 gives three examples among diseases, for an investigator who has published on obesity. Morbid obesity, a narrower form of obesity, lies at a distance of 0.32; type 2 diabetes, which shares only the branch of nutritional and metabolic diseases with obesity, at 0.41; HIV infections, which share nothing with obesity below the broad category of diseases, at 0.87. The examples also show two limits of the distance. The tree classifies what a descriptor is, not what it is studied with, so obesity and type 2 diabetes belong to different branches, although their literatures overlap a great deal. A narrower term below a familiar one, such as morbid obesity, still lies about a third of the way out, because much of its information is not contained in the broader term.

[2] The information content of an event of probability p, $-\ln p$, is the standard measure of how much its occurrence tells us. Two of its properties are useful here. The shares of *PubMed* covered by MeSH positions span several orders of magnitude, from 83 percent for Mammals to a few thousandths of a percent for some compounds, and the logarithm places them on a scale on which each tenfold narrowing adds the same amount, 2.3. And a position is never less informative than a broader position above it, so that the ratio defined below lies between 0 and 1.

The distance of a new pair is the larger distance of its two descriptors, so that near pairs, whose two descriptors are familiar, have a distance of 0. The distance of a cycle's pivot is the average distance of its new pairs, each pair counted once for each of the cycle's publications on which it appears, as in pivoting. It runs from 0 to 1 and does not grow with the number of new pairs: pivoting records how much of a cycle's work is new, the distance how different the new work is from the old.

Formally, let $\mathcal{T}(d)$ denote the positions of descriptor d in the tree, and $\mathcal{A}(d)$ these positions together with all the positions above them; let n_t denote the number of *PubMed* publications indexed with a descriptor at or below position t, and N the number of *PubMed* publications with any descriptor. The information content of a position, that of a descriptor, and that of the most specific position a descriptor shares with the investigator's earlier descriptors K_c are

$$I(t) = -\ln \frac{n_t}{N}, \qquad I(d) = \max_{t \in \mathcal{T}(d)} I(t), \qquad S_c(d) = \max \left\{ I(t) : t \in \mathcal{A}(d) \cap \bigcup\nolimits_{k \in K_c} \mathcal{A}(k) \right\},$$

with $S_c(d) = 0$ when they share no position. The distance of a new descriptor, of a new pair, and of a cycle's pivot are

$$\delta_c(d) = 1 - \frac{S_c(d)}{I(d)} \quad \text{for } d \notin K_c \quad (\text{and } 0 \text{ for } d \in K_c), \qquad \delta_c(\{a, b\}) = \max\{\delta_c(a), \delta_c(b)\},$$

$$\text{Distance}_c = \frac{\sum_{p \in P_c} \sum_{\pi \in \Pi_p \setminus H_c} \delta_c(\pi)}{\sum_{p \in P_c} |\Pi_p \setminus H_c|}.$$

Descriptive statistics for the distance are presented in Table D1. It averages 0.44, with a standard deviation of 0.14, and its distribution is nearly the same for renewed and non-renewed cycles (means of 0.442 and 0.446; Figure D1, panel C). It is related to the variables of Table D2 without duplicating them. Cycles in which a larger share of the pairs are far, that is, involve at least one descriptor new to the investigator, also tend to have new pairs more distant from her earlier work (correlation 0.73). Cycles in which a larger share of the work is new, as measured by pivoting, tend to have more distant new pairs as well (correlation 0.59). Neither correlation is close to one, so the distance records information that neither the share of far pairs nor the pivoting measure records. The distance is higher for new investigators (0.51, against 0.42) and falls as the investigator's earlier record grows, since a longer record contains more descriptors close to any new one.

Table D3 relates renewal to the distance, alone and together with pivoting, with the same controls. Alone, one standard deviation of the distance is associated with a probability of renewal lower by 1.5 percentage points (-0.107). With pivoting held fixed (that is, among cycles in which the same share of the work is new), a pivot one standard deviation more distant is associated with a probability of renewal lower by 0.75 percentage points (-0.053), and pivoting itself keeps most of its coefficient (-0.103). Both how much of the work is new and how distant it is from her earlier work are associated with a lower likelihood of renewal. The gradient is steepest for new investigators: 1.6 percentage points per standard deviation with pivoting held fixed (-0.108), against 0.4 points for other investigators (-0.031, significant at the 10 percent level).[3] The two tables are consistent once one recalls which variation each one uses. For an investigator who already holds a grant, the main contrast is between recombining familiar descriptors and bringing in new ones, which Table D2 captures; grading the far pairs adds less. For a new investigator, nearly every new pair is far, and what varies is how distant those pairs are from her earlier work, which is what the distance records.

[3] About forty MeSH descriptors are check tags, which indexers apply to nearly every publication to record its subjects: Humans, Animals, sex, age groups, pregnancy, the common laboratory animals, and a few others. They record whom or what a study examined, and they appear in clusters: an investigator who moves from animal to human subjects acquires Adult, Middle Aged, and Aged at once. Computed on the new pairs that involve no check tag, the distance correlates 0.98 with the one used here, and its coefficients with pivoting held fixed are slightly larger: -0.063, -0.115, and -0.041 for all, new, and other investigators. The list of check tags is that of the National Library of Medicine (Le and Mork 2023).

Leaving a fading field. A pivot may also be a retreat. An investigator whose field is drying up may move on because staying has become the riskier choice. If such pivots made up much of the measure, and if reviewers recognized them, pivots out of fading fields would carry a smaller penalty.

To measure whether a field was fading, we need to identify the publications that are intellectually close to the investigator's earlier work. We rely on the *PubMed* Related Citations Algorithm (PMRA), which produces the list of similar articles that *PubMed* displays alongside each record (Lin and Wilbur 2007). PMRA scores the similarity of two publications from the words of their titles and abstracts and from their MeSH descriptors, weighting each term by its rarity in *PubMed* and by its prominence in each publication, and it returns for each publication its closest neighbors, about a hundred of them (129 at the median in our data). Nothing else enters the calculation: not the authors, their institutions, or their collaborators, not the journals, and not the references the publications cite. The relatedness that PMRA records is thus intellectual, stripped as far as possible of the social ties through which proximity between publications is often measured. Because the MeSH descriptors are assigned by the Library's indexers rather than by the authors, it also depends less on how authors choose to present their work than a measure built from their words alone would. The parameters of the algorithm were calibrated against human judgments of relevance, and about a fifth of non-trivial *PubMed* search sessions make use of it (Lin and Wilbur 2007). Azoulay et al. (2019a, Appendix C) describe the algorithm in detail and use it to delineate scientific fields.

For each of the investigator's publications of the five years before the cycle, we set aside the related articles she wrote herself, since her own output would otherwise make her field look alive when she stays in it and dead when she leaves, and we compute the share published in the five years before the cycle among those published in the ten years before it, for publications with at least five such articles. A share above one half means that work close to the publication had become more abundant. Related articles cluster around the date of the publication itself (Azoulay et al. 2019a, Appendix C), so this share mechanically rises with the publication's own year: it averages 0.73 for publications of the year before the cycle and 0.61 for those of five years before. We therefore compare each publication only with publications of the same year in cycles that start the same year, and average the differences over the investigator's publications. We call this average the *vibrancy* of the field, and a field is fading when its vibrancy falls below the median among the cycles that start the same year. The measure looks only at the years before the cycle, the field as the investigator could see it when choosing her direction, and it draws on information, the dates of intellectually related publications, that pivoting does not use.

Table D4 adds the indicator of a fading field to the pivoting regression, with the same controls, first alone and then interacted with pivoting. Vibrancy is defined for 98 percent of the cycles, and restricting the sample to them leaves the coefficient on pivoting essentially unchanged (-0.121, against -0.124). Investigators in fading fields are renewed at the same rate as others (0.001, column 3). The interaction is close to zero (0.008, with a standard error of 0.021, column 4): pivots out of fading fields carry a penalty of -0.117 (0.015), against -0.126 for the others. Even at the upper end of the interaction's confidence interval (0.048), pivots out of fading fields would keep more than half of the penalty.

A comparison of the two halves of the distribution could hide differences at its ends, where the classification is the least ambiguous. We therefore divide the cycles into ten groups by the vibrancy of the field, within the year the cycle starts, and estimate in one regression a separate effect of pivoting in each group, with indicators for the groups and the controls common to all of them (Figure D3). The ten effects lie between -0.093 and -0.147, and they cannot be told apart ($p = 0.95$). The penalty in the most fading tenth of cycles (-0.130) is nearly as large as in the most thriving tenth (-0.141).

The motive for leaving a field, as far as we can measure it, does not change how reviewers judge the pivot. The result may seem surprising: a grant system that allocated attention nimbly across topics might reward investigators who leave a declining field, or at least spare them any penalty. But science is subject to strong inertial forces. The eminent scientists who dominate a field can slow the entry of newcomers and new ideas into it (Azoulay et al. 2019a), and at NIH some of this inertia is ratified by the partition of intellectual space into study sections, whose boundaries the Center for Scientific Review redraws only slowly and which therefore lag the shifting allocation of scientific attention across topics. However surprising, the absence of a difference fits what is known about the ossification of the NIH grant system, of which this paper studies one facet. Our measure records the vibrancy of the literature around the investigator's earlier work, not her reasons for moving, and a different measure of fading

fields might draw a different line; on this one, the penalty on pivoting is the same at every level of vibrancy of the field left behind.

A pivot proposed in the application. A pivot need not stray from the plan. An investigator may propose new territory in the application that opens the cycle, and a study section may review the proposal and recommend it for funding. If the penalty at renewal fell only on pivots that departed from the approved plan, it would reflect reviewers enforcing the terms of the award, and would say little about how they judge risk. To tell the two apart, we measure the pivot in the application itself. The National Library of Medicine's Medical Text Indexer assigns MeSH descriptors to grant abstracts, and for each cycle we compute the share of the starting application's descriptors that appear on none of the investigator's publications of the five years before the cycle, the same publications against which pivoting is measured. We count single descriptors, since machine indexing of a short abstract rarely reproduces the pairs that indexers assign to publications. The measure is defined for three quarters of the cycles of Table 2, and it correlates 0.39 with the pivoting of the cycle's publications: applications that propose more new territory lead to more of it, though far from one for one. We call a pivot proposed when the share lies above its median among the cycles that start the same year. In the median application, 57 percent of the descriptors are new to the investigator; proposed pivots average 70 percent, the others 44 percent.

Table D5 follows the design of Table D4. On the cycles with an application measure, pivoting carries the same penalty as in Table 2 (-0.128, against -0.124). Holding pivoting fixed, cycles whose application proposed a pivot are renewed slightly less often, by 1.0 percentage point (significant at the 1 percent level, columns 3 and 4). The interaction is small and, if anything, negative (-0.019, with a standard error of 0.025): proposed pivots carry a penalty of -0.131 (0.019), against -0.112 for the others. A pivot that a study section read in the proposal and recommended for funding is penalized at renewal at least as much as one the investigator undertook along the way.

Interpretation. The pivoting measure is a simplification. It counts new pairs of descriptors without regard to their distance from the investigator's earlier work, and it knows nothing of why the investigator moved. Taking it apart sharpens our reading of it without overturning it. Reviewers do not appear to penalize investigators for recombining what they already know; pivots into unfamiliar territory are penalized, and the more so the more distant the new descriptors are from her earlier work, holding fixed how much of the work is new. That territory is new to her study section as well: pivots into combinations of descriptors that some other funded project of her study section contains carry no penalty, and only those that no such project contains are penalized (Appendix E, Table E3). It is not concentrated among investigators leaving fields that were fading, and it is not a penalty for departing from the approved plan: pivots proposed in the application that won the grant are penalized as much as the others. This is the gradient that Hill et al. (2025) document for the impact of research: the further scientists move from their prior work, the lower the impact of what they produce. The measure is therefore a useful summary of a behavior whose components this appendix separates.

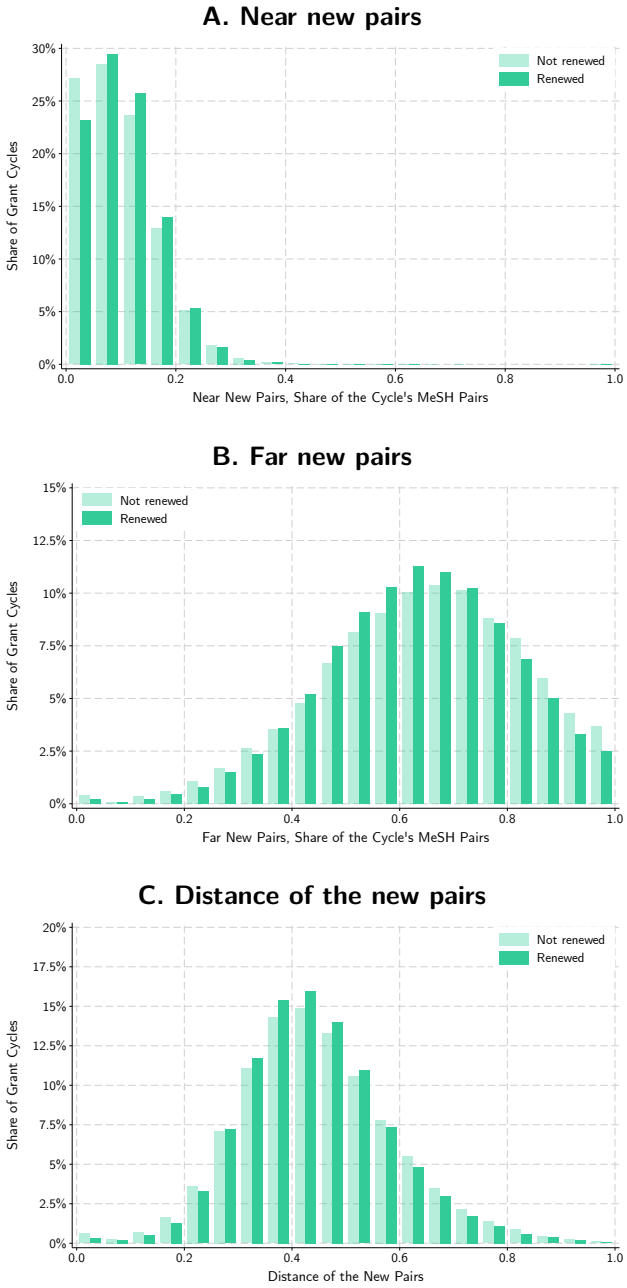
Table D1: The Parts of the Pivoting Measure, and the Distance of the New Pairs

	Mean	Std. dev.	Median	Share = 0	Cycles	Mean: New inv.	Mean: Others
A. The parts of the pivoting measure							
Pivoting, all new pairs	0.739	0.142	0.753	0.000	101,435	0.790	0.722
Near new pairs	0.100	0.068	0.092	0.058	101,435	0.069	0.110
Far new pairs	0.639	0.181	0.647	0.003	101,435	0.720	0.612
Among far pairs, share with both descriptors new	0.281	0.145	0.259	0.009	101,134	0.342	0.260
Distance of the new pairs	0.444	0.141	0.434	0.004	101,404	0.512	0.422
Cycles						25,624	75,811

B. Correlations	All new pairs	Near	Far	Distance
Share of new single descriptors	0.910	-0.687	0.971	0.743
Cosine dissimilarity from prior work	0.584	-0.298	0.570	0.364
Number of baseline papers	-0.367	0.470	-0.464	-0.490
Number of baseline papers (log)	-0.428	0.565	-0.547	-0.623
Near new pairs			-0.701	
Distance of the new pairs	0.588	-0.725	0.733	
Cycles		101,404		

Notes: The unit of analysis is the grant cycle, on the cycles of the pivoting regressions of Table 2. Pivoting is the share of the MeSH descriptor pairs on the cycle's acknowledging publications that appear on none of the investigator's English original research publications of the five years before the cycle starts, with a pair counted once for each of the cycle's publications on which it appears. A new pair is near when both of its descriptors appear on those earlier publications, far when at least one does not; the near and far shares add up to pivoting. The share of far pairs with both descriptors new is defined for the cycles with at least one far pair. The distance of the new pairs is their average distance from the investigator's earlier publications, each weighted as pivoting weights it; a pair's distance is the larger distance of its two descriptors, and a descriptor's distance is one minus the ratio of the information content of the most informative MeSH tree position it shares with a descriptor of the earlier publications to its own information content. The information content of a position is $\ln(N/c)$, where N is the number of *PubMed* publications with a MeSH descriptor (33.0 million) and c the number carrying a descriptor at or below the position. New investigators are those whose cycle starts in their first year of an R01-equivalent award. Panel B gives Pearson correlations. The share of new single descriptors counts each of the cycle's descriptors once for each of its publications indexed with it. The cosine dissimilarity is one minus the cosine similarity of the cycle's and the earlier publications' descriptor profiles, each descriptor weighted by the number of publications carrying it times its inverse document frequency. The number of baseline papers counts the earlier publications with a MeSH descriptor.

Figure D1: Near and Far Pivots and the Distance of the New Pairs, by Renewal Status



Notes: The cycles of the pivoting regressions of Table 2, by renewal status, in 20 bins of equal width; each bar gives the share of the cycles of its group (renewed or not renewed). Near and far new pairs are shares of the cycle's MeSH descriptor pairs, and they add up to pivoting; the distance of the new pairs runs from 0 to 1 (Table D1). Means for renewed and non-renewed cycles: near pairs 0.102 and 0.098, far pairs 0.633 and 0.644, distance 0.442 and 0.446.

Table D2: Near and Far Pivots and Grant Renewal

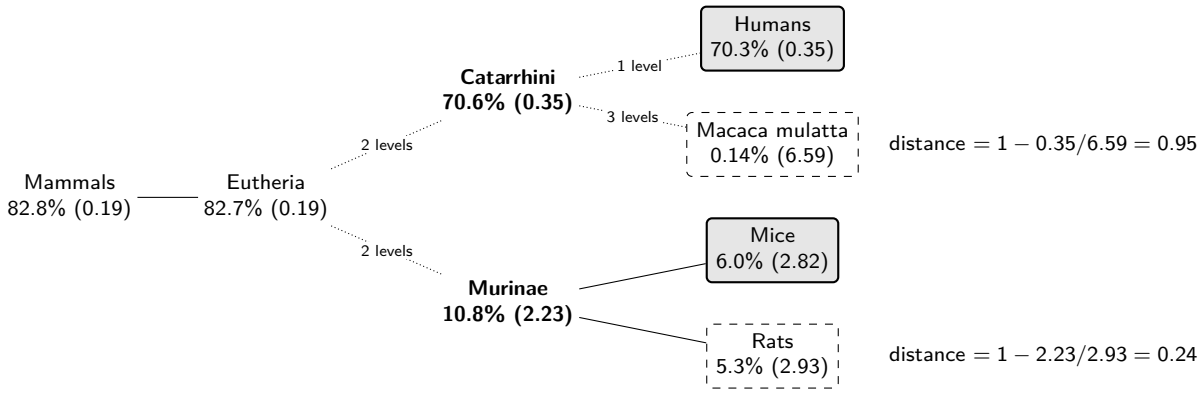
	All cycles		New investigators		Other investigators	
	(1)	(2)	(3)	(4)	(5)	(6)
Pivoting, all new pairs	-0.124**		-0.149**		-0.114**	
	(0.012)		(0.024)		(0.014)	
Near new pairs		-0.026		-0.077		0.002
		(0.034)		(0.075)		(0.038)
Far new pairs		-0.113**		-0.140**		-0.101**
		(0.012)		(0.026)		(0.014)
p-value, near = far		0.002		0.311		<0.001
Mean of dependent variable	0.421	0.421	0.372	0.372	0.438	0.438
Nb. of grant cycles	101,435	101,435	25,153	25,153	75,630	75,630
Nb. of investigators	39,337	39,337	24,812	24,812	27,680	27,680
Adjusted R^2	0.1682	0.1683	0.1434	0.1434	0.1771	0.1772

Notes: Estimates stem from a linear probability model in which the unit of analysis is the grant cycle and the dependent variable is an indicator for whether the grant cycle was renewed, with the controls of Table 2 and indicators for the number of the investigator's publications with a MeSH descriptor in the five years before the cycle (1, 2, 3-4, 5-9, 10-19, 20-49, 50 or more). Pivoting and its near and far parts (Table D1) enter as shares of the cycle's MeSH descriptor pairs. Columns 1 and 2 include the cycles of the pivoting regressions of Table 2; columns 3 and 4 the cycles of new investigators, whose cycle starts in their first year of an R01-equivalent award; columns 5 and 6 the cycles of other investigators. Within a group, both columns use the same cycles. The p-value is that of the test that the near and far coefficients are equal.

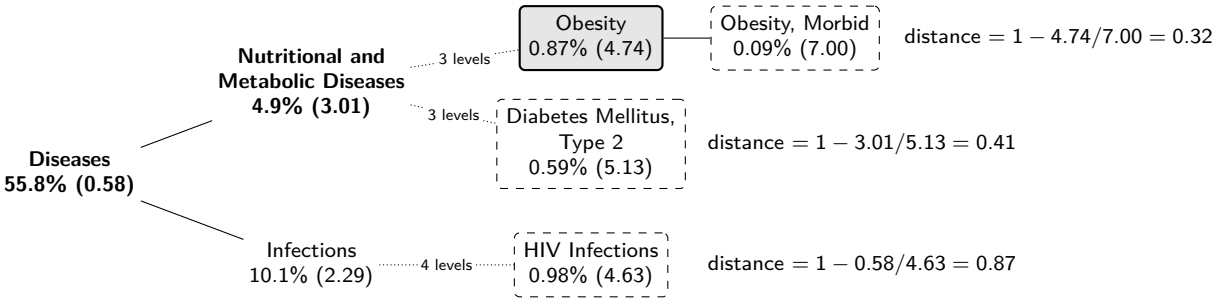
Standard errors in parentheses, clustered by investigator and by grant. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Figure D2: The Distance of a New Descriptor: Five Examples from the MeSH Tree

A. Organisms: an investigator who has published on human subjects and on mice



B. Diseases: an investigator who has published on obesity



Notes: Each panel shows part of the MeSH tree. Shaded boxes are descriptors that appear on the investigator's publications of the five years before the cycle; dashed boxes are descriptors new to her that appear on the cycle's publications. Under each position, the share of the 33 million *PubMed* publications with a MeSH descriptor that are indexed with it or with a descriptor below it, and in parentheses its information content, minus the natural logarithm of that share. Positions in bold are the most specific positions that a new descriptor shares with the investigator's earlier work: Murinae for Rats, Catarrhini for the rhesus macaque, the branch of nutritional and metabolic diseases for type 2 diabetes, and the broad category of diseases for HIV infections; for morbid obesity, it is Obesity itself. The distance of a new descriptor is one minus the ratio of that position's information content to the descriptor's own. Dotted lines skip intermediate positions, whose number is shown. A descriptor may occupy several positions in the tree; the figure shows one, and its numbers are those of the descriptor's most specific position. Catarrhini and Murinae both lie nine levels below the top of the tree.

Table D3: How Much and How Far: The Distance of a Pivot and Grant Renewal

	All cycles		New investigators		Other investigators	
	(1)	(2)	(3)	(4)	(5)	(6)
Pivoting, all new pairs		-0.103**		-0.100**		-0.102**
		(0.014)		(0.028)		(0.016)
Distance of the new pairs	-0.107**	-0.053**	-0.157**	-0.108**	-0.087**	-0.031†
	(0.013)	(0.015)	(0.025)	(0.028)	(0.016)	(0.018)
One s.d. of distance, in points of renewal	-1.52	-0.75	-2.31	-1.58	-1.15	-0.41
Mean of dependent variable	0.421	0.421	0.373	0.373	0.438	0.438
Nb. of grant cycles	101,404	101,404	25,146	25,146	75,606	75,606
Nb. of investigators	39,328	39,328	24,805	24,805	27,673	27,673
Adjusted R^2	0.1678	0.1683	0.1435	0.1439	0.1767	0.1771

Notes: Estimates stem from a linear probability model in which the unit of analysis is the grant cycle and the dependent variable is an indicator for whether the grant cycle was renewed, with the controls of Table 2 and the indicators for the number of earlier publications of Table D2. The distance of the new pairs (Table D1) runs from 0 to 1; pivoting enters as a share of the cycle's MeSH descriptor pairs. The groups are those of Table D2, restricted to the cycles with at least one new pair; within a group, both columns use the same cycles. The effect of one standard deviation of the distance is computed on each column's cycles.

Standard errors in parentheses, clustered by investigator and by grant. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

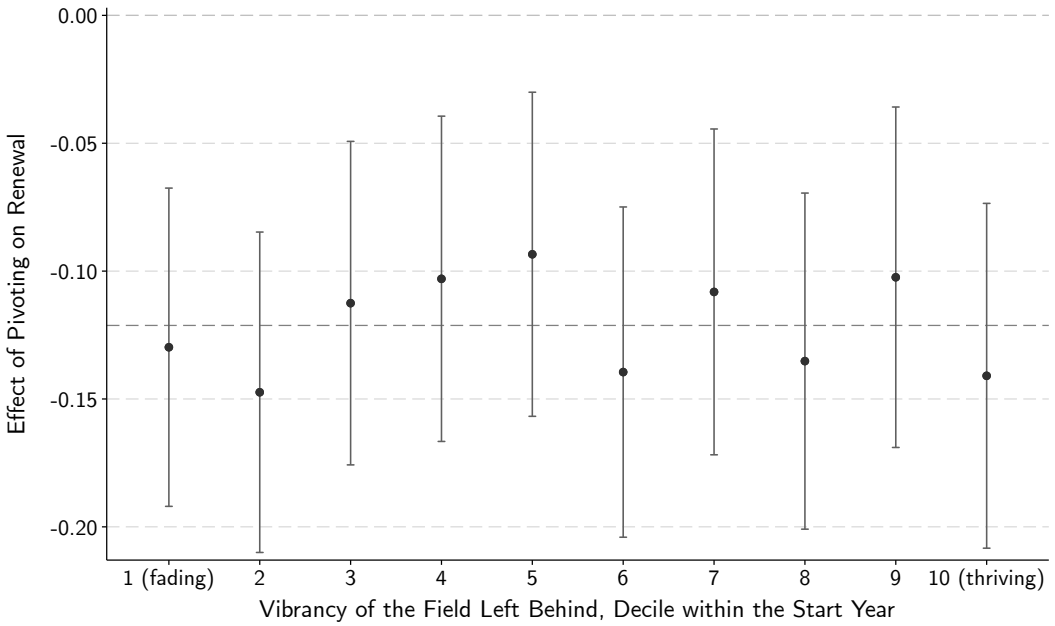
Table D4: Pivoting and the Vibrancy of the Field Left Behind

	(1)	(2)	(3)	(4)
Pivoting	-0.124**	-0.121**	-0.121**	-0.126**
	(0.012)	(0.012)	(0.012)	(0.016)
Fading field			0.001	0.001
			(0.003)	(0.003)
Pivoting × fading field				0.008
				(0.021)
Sample	Table 2	Vibrancy defined	Vibrancy defined	Vibrancy defined
Mean of dependent variable	0.421	0.419	0.419	0.419
Nb. of grant cycles	101,435	99,605	99,605	99,605
Nb. of investigators	39,337	39,216	39,216	39,216
Adjusted R^2	0.1682	0.1674	0.1674	0.1674

Notes: Estimates stem from a linear probability model in which the unit of analysis is the grant cycle and the dependent variable is an indicator for whether the grant cycle was renewed, with the controls of Table 2 and the indicators for the number of earlier publications of Table D2. Pivoting enters as a share of the cycle's MeSH descriptor pairs. Column 1 includes the cycles of the pivoting regressions of Table 2; columns 2 to 4, those among them for which the vibrancy of the field is defined. For a cycle starting in year s, the investigator's earlier publications are her English original research of $s-5$ to $s-1$, in any authorship position, except those acknowledging the grant under review. For each of them, we take its *PubMed* Related Articles, set aside those the investigator wrote, and compute the share published in $s-5$ to $s-1$ among those published in $s-10$ to $s-1$, for publications with at least five such articles; we subtract the mean share of the publications of the same year in cycles starting in year s. Vibrancy is the mean over the cycle's earlier publications, and a field is fading when vibrancy falls below its median among the cycles of column 2 with the same s. In column 4, pivoting is centered at its mean on the cycles of column 2, so that the coefficient on the fading field is the difference at average pivoting.

Standard errors in parentheses, clustered by investigator and by grant. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Figure D3: The Penalty on Pivoting, by Vibrancy of the Field Left Behind



Notes: Estimates from one linear probability model of renewal, on the cycles of Table D4, column 2. The cycles are divided into ten groups of equal size by the vibrancy of the field (Table D4), within the year the cycle starts; group 1 holds the most fading fields and group 10 the most thriving. The model contains a separate effect of pivoting (a fraction) for each group, indicators for the groups, and the controls of Table D4, which are common to all groups. Each point is the effect of pivoting within its group, with its 95 percent confidence interval. The dashed line is the effect for all these cycles together (Table D4, column 2). The test that the ten effects are equal has a p-value of 0.95. Standard errors clustered by investigator and by grant.

Table D5: Pivots Proposed in the Application

	(1)	(2)	(3)	(4)
Pivoting	-0.124**	-0.128**	-0.121**	-0.112**
	(0.012)	(0.014)	(0.014)	(0.019)
Pivot proposed in the application			-0.010**	-0.010**
			(0.004)	(0.004)
Pivoting × pivot proposed in the application				-0.019
				(0.025)
Sample	Table 2	Application measured	Application measured	Application measured
Mean of dependent variable	0.421	0.432	0.432	0.432
Nb. of grant cycles	101,435	77,153	77,153	77,153
Nb. of investigators	39,337	34,057	34,057	34,057
Adjusted R^2	0.1682	0.1682	0.1682	0.1682

Notes: Estimates stem from a linear probability model in which the unit of analysis is the grant cycle and the dependent variable is an indicator for whether the grant cycle was renewed, with the controls of Table 2 and the indicators for the number of earlier publications of Table D2. Pivoting enters as a share of the cycle's MeSH descriptor pairs. Column 1 includes the cycles of the pivoting regressions of Table 2; columns 2 to 4, those among them whose starting application can be compared with the investigator's earlier work: its abstract has at least two MeSH descriptors assigned by the National Library of Medicine's Medical Text Indexer (its own, or those of an identical abstract), and the investigator has a publication with MeSH descriptors in the five years before the cycle. The application measure is the share of the application's descriptors that appear on none of the investigator's English original research publications of those five years; a pivot is proposed when the measure lies above its median among the cycles of column 2 that start the same year. In column 4, pivoting is centered at its mean on the cycles of column 2, so that the coefficient on the proposed pivot is the difference at average pivoting.

Standard errors in parentheses, clustered by investigator and by grant. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Appendix E: Dissecting the Standing Out Measure

Our fourth measure of risk taking, standing out, captures how far the work a grant supports lies from the work of the other projects that its study section reviewed and its institute funded. A study section judges an application against what its members know, and the projects it has recommended for funding record what the scientific community it serves works on. An investigator whose publications combine topics that none of those projects combines accepts the chance that the reviewers of her renewal application will lack a frame of reference for her work. This appendix defines the measure in full and then takes it apart.

Two objections deserve an answer.[1] The first is that the measure resembles a measure of novelty computed against a narrower group. A measure of novelty counts the combinations of descriptors that no earlier publication contains; standing out counts those that no funded project of the study section contains. If standing out recorded novelty relative to a study section, its association with renewal would tell us nothing about risk taking that a measure of novelty does not. The second objection is that a combination can be absent from the funded projects of a study section because its descriptors are specific, whether or not the investigator took a risk in pursuing it.

The measure. For each grant cycle, we list the pairs of MeSH descriptors that appear together on the publications that acknowledge the grant during the cycle, counting a pair once for each publication on which it appears: a publication indexed with n descriptors contributes $n(n-1)/2$ pairs. We compare these pairs with the cycle's *pool*: every grant application funded by the cycle's institute in the cycle's evaluation year, the last year of its funding, and reviewed by the study section that reviewed the application that started the cycle, whatever the mechanism, new awards and continuing ones alike, except the applications on which the investigator is a principal investigator. The pool is thus the portfolio of projects that the institute supported, out of the study section's reviews, in the year the cycle ended. A pair is *shared* when at least one application of the pool is indexed with both of its descriptors, and *unique* otherwise. Standing out is the share of the cycle's pairs that are unique. For a cycle c whose acknowledging publications $p \in P_c$ are indexed with sets of descriptors D_p, and whose pool consists of applications $j \in A_c$ indexed with sets T_j,

$$\text{Standing out}_c \;=\; \frac{\sum_{p\in P_c} \#\big\{\{a,b\} \subseteq D_p :\ \{a,b\} \not\subseteq T_j \text{ for every } j \in A_c\big\}}{\sum_{p\in P_c} |D_p|\,(|D_p|-1)/2}.$$

The National Library of Medicine indexes publications but not grant applications. The descriptors of the pool's applications were assigned to their abstracts by the Library's Medical Text Indexer (MTI), the software that has suggested descriptors to its indexers since 2002 (Mork et al. 2017). In the pools of our cycles, 95 percent of the applications have descriptors of their own; the others take those of an application whose abstract is identical, since the continuing years of an award repeat the abstract of its competing year. An application has 17 descriptors at the median. Applications without an abstract or without descriptors leave the pool. On both sides, we keep only the 29,053 descriptors that the two vocabularies share, which account for 99.3 percent of the cycles' pairs, so that a pair cannot be unique because one of its descriptors was unknown on the other side.

We compute the measure when the pool holds at least 20 applications with descriptors besides the investigator's own, and when the cycle's publications yield more than five pairs. The first condition excludes 19 percent of the analysis cycles with a pair, the second 52 cycles. Standing out is defined for 83,098 cycles, 80 percent of the analysis cycles with an acknowledging publication: between 83 and 90 percent in each evaluation year until 2005, between 67 and 81 percent afterwards, when pools are smaller (the median pool holds 159 applications for the cycles evaluated in the 1980s, 108 for those of the 2010s), and 9 percent in 1996, a year for which RePORTER provides the abstracts of few applications.

Standing out averages 0.80, with a standard deviation of 0.11 (Table E1): four-fifths of a cycle's pairs appear together on no application of its pool. Many of the shared pairs involve *check tags*, about forty descriptors that the Library's indexers consider for every publication, whatever its topic, to record whom or what a study examined: Humans or Animals, common laboratory species such as Mice and Rats, sex (Female, Male), age groups such as Adult and Aged, Pregnancy, and a few others (Le and Mork 2023). Check tags appear on many publications and

[1] We thank the Department Editor and two anonymous referees for pressing us on these points.

many applications: Humans, for instance, on 48 percent of the publications that acknowledge our grants and on 69 percent of the applications that MTI indexed. A pair that includes one is therefore shared more often than others: 61 percent of the shared pairs include a check tag, against 30 percent of all pairs. This raises a concern. Check tags say little about a study's topic, and they arrive in clusters (a study of patients is often indexed with Female, Male, and several age groups at once), so a measure built partly from them could record the kind of study, on human subjects or animals, on adults or children, as much as its topic. We return to this concern below by computing the measure without them. The distribution of the measure is nearly the same for renewed and non-renewed cycles (Figure 2; means of 0.797 and 0.800), and every comparison below holds fixed the controls of Table 2.

An example. A publication of 1984 on the cross-linking of DNA by psoralens, the only publication acknowledging its grant in that cycle, is indexed with eight descriptors: Base Sequence, Binding Sites, Cross-Linking Reagents, DNA Restriction Enzymes, Deoxyribonucleases, Type II Site-Specific, Furocoumarins, Kinetics, and Trioxsalen. Of its 28 pairs, four appear together on at least one of the 199 applications of the pool, such as Base Sequence with Kinetics and Binding Sites with DNA Restriction Enzymes. The other 24 are unique, and the cycle's standing out is $24/28 = 0.86$.

Two indexers. The two sides of the comparison are indexed differently. The Library's indexers describe a publication from its full text, with the vocabulary in use when they index it, and the Library does not add a descriptor to publications indexed before the descriptor existed.[2] MTI described each abstract from its text alone, with a recent vocabulary. A pair can therefore be unique because the two indexers describe similar work in different terms. Both differences affect alike all the cycles evaluated in the same year, and every comparison below holds the evaluation year fixed. The difference in vocabularies shrinks over the period, from about four decades for publications of the 1980s to a few years for the most recent, while the penalty on standing out shows no trend over the period (Appendix K).

Where the unique pairs are found. If standing out recorded novelty, its association with renewal would come from pairs that are new to science. We divide each cycle's unique pairs into three groups that add up to the measure. A unique pair may appear on one of the investigator's own publications of the five years before the cycle, the publications against which pivoting is measured: she has worked on the combination before, and the funded projects of her study section have not. Otherwise the pair is new to her. It is then either *established in PubMed*, if a publication first used it more than three years before the cycle's publication, or *new to PubMed*, if it was first used within those three years. We use three years because it is the threshold of the share of new MeSH pairs, one of the measures of novelty of Appendix F. In the example, 13 of the 24 unique pairs appear on the investigator's earlier publications, such as Cross-Linking Reagents with Furocoumarins; 7 are new to her but had appeared in *PubMed* by 1980, such as DNA Restriction Enzymes with Furocoumarins; and the last 4, all of them with Deoxyribonucleases, Type II Site-Specific, first appeared in 1981 or later. Across the cycles of Table E2, the three groups average 16, 56, and 7.5 percent of a cycle's pairs.

In Table E2, we enter the three groups in place of the measure, with the controls of Table 2 and indicators for the number of the investigator's earlier publications, on which the first group depends. One standard deviation of standing out is associated with a probability of renewal lower by 1.9 percentage points (column 1). Almost all of that association comes from unique pairs new to the investigator and established in *PubMed* (-0.220, or 2.8 percentage points per standard deviation, column 2). Pairs new to *PubMed* carry a penalty a third as large per unit (-0.074; the two coefficients differ at $p < 0.0001$), 0.4 point per standard deviation, and unique pairs from the investigator's own earlier publications with a small one (-0.051, significant at the 5 percent level), 0.5 point. The conclusion does not depend on the threshold of three years. Counting as new to *PubMed* every pair first used within ten years, a fifth of a cycle's pairs, the coefficient is -0.220 for the established pairs and -0.152 for the new ones.

Unique pairs also differ in their distance from the projects of the study section. A unique pair may combine descriptors that some project of the section uses, though never together, or bring in a descriptor that no project of

[2] "NLM never retrospectively indexes MEDLINE citations with new MeSH Headings" (Willis 1999). The main exception is a chemical whose supplementary concept record is promoted to a descriptor: since 2001, the descriptor is added to the publications indexed with the supplementary concept (Knecht et al. 2001).

the section uses; the 13 pairs of the example that involve Trioxsalen or Deoxyribonucleases, Type II Site-Specific are of the second kind. Those pairs carry most of the penalty (−0.157, against −0.080 for the first kind, or 2.7 and 0.8 percentage points per standard deviation). We do not lean on this contrast: a descriptor that no application of the pool uses may be one that MTI rarely assigns. Time Factors, for instance, appears on 5.3 percent of the publications that acknowledge our grants and on 0.07 percent of the applications that MTI indexed, so that a pair with Time Factors is far in most pools.

Standing out and pivoting. The group that carries the penalty, pairs new to the investigator and absent from the funded projects of her study section, also belongs to pivoting, which counts every pair new to the investigator, whether or not those projects contain it. The two measures correlate 0.25. In Table E3, we divide the cycle's pairs both ways, new to the investigator or not, and shared with the study section or not. Pivoting, computed on the same pairs (it correlates 0.9996 with the measure of Table 2), is associated with a probability of renewal lower by 1.8 percentage points per standard deviation (column 1). In column 2, we split it by the study section's portfolio. Pairs new to the investigator that some project of the study section contains are associated with a slightly higher probability of renewal (0.070); pairs new to her that no project contains, with a lower one (−0.161, or 2.2 points per standard deviation). Column 3 adds the first group of Table E2, combinations familiar to the investigator and absent from her study section, so that every group is compared with the combinations familiar to both. For that group, the difference is close to zero (−0.001). The pattern holds among investigators who already held an R01-equivalent grant (0.118 and −0.155 for the two parts of pivoting), and among new investigators the pairs her study section contains carry no penalty (0.000, against −0.165).

On either measure, then, the cycles renewed less often are those whose work is new to the investigator and absent from the funded projects of her study section. A pivot into combinations those projects contain carries no penalty, and neither does a line of work that the investigator has pursued before and her study section does not share. For pivoting, the penalty is concentrated on departures from the investigator's expertise that also leave her study section's portfolio; for standing out, on departures from that portfolio that are also new to her, more than on a lasting difference between her work and her study section's. The penalties that Table 2 reports for the two measures therefore overlap: both fall on departures from the investigator's earlier work that are also departures from her study section's portfolio. They do not coincide: the two measures correlate 0.25, and with all four measures in one regression, each keeps most of its coefficient (Appendix H, Table H1).

A placebo study section. A pair can also be unique because it is specific: combinations of specific descriptors are rare everywhere, and work described in specific terms would stand out against any study section. To set this apart from the comparison with the investigator's own study section, we compute standing out against placebo study sections: the same share, computed against the pools of five other study sections of the same evaluation year, drawn at random among those with at least 20 applications, and averaged over the five. Against the portfolio of a study section from another field, 95 percent of a cycle's pairs are unique on average, and what varies across cycles is the small remainder that is shared. That remainder consists mostly of pairs that involve check tags: standing out against the placebo study sections correlates −0.51 with the share of a cycle's pairs that involve one. The measure thus ranks cycles by how generically their publications are indexed. At its low end are cycles whose publications are indexed largely with check tags, as studies of patients often are (sex, age groups); at its high end, cycles whose publications are indexed with specific molecules and methods.

In Table E4, standing out against the placebo study sections is associated with a higher probability of renewal (0.214, or 0.7 percentage point per standard deviation, column 2): within a study section and evaluation year, and at equal output, cycles whose work is indexed in specific terms are renewed somewhat more often. The association may reflect differences between kinds of research that our controls do not capture, and we do not interpret it further. It matters here for two reasons. First, being distinctive against unrelated study sections is not penalized, so the specificity of the work does not account for the penalty on standing out. Second, the measure against the placebo study sections correlates 0.36 with the measure against the actual one. Holding it fixed removes from standing out a component associated with higher renewal, and the penalty becomes larger (−0.226 against −0.177, or 2.4 percentage points per standard deviation against 1.9, column 3).

The two rows of Table E4 hold the same measure, computed against two comparison groups, and their coefficients have opposite signs. The comparison group determines what the measure records. Against the investigator's own study section, standing out records work that the funded projects of her study section do not share, and it is

associated with a lower probability of renewal; against unrelated study sections, it records only how specific the work is, and it is associated with a higher one.

The pool and the check tags. A larger pool shares more pairs by construction, and pools differ in size across sections and years. With indicators for the size of the pool (20 to 49, 50 to 99, 100 to 149, 150 to 199, 200 to 299, and 300 or more applications), the coefficient on standing out is -0.181; on the cycles whose pool holds at least 50 or at least 100 applications, -0.171 and -0.168; on all the cycles of Table 2, -0.177. To address the concern raised by the check tags, we compute standing out on the pairs that involve none of them. Its coefficient is then -0.273, or 2.4 percentage points per standard deviation, against 1.9 for the measure on the same cycles: without the check tags, the association with renewal is slightly stronger.

Interpretation. Standing out is not a measure of novelty computed against a narrower group, and it is not a measure of specific work. Its association with renewal comes from combinations that science had used for years and that the projects of the investigator's study section did not contain, and it grows when the specificity of the work is held fixed. Taking it apart also shows how it relates to pivoting: both measures are associated with lower renewal when the investigator enters territory that is new to her and that the funded projects of her study section do not cover, and neither is when only one of the two holds. What it records is mostly a departure that takes the investigator outside both her own expertise and her study section's portfolio.

Table E1: The Parts of the Standing Out Measure

				Mean		
	Mean	Std. dev.	Median	Not renewed	Renewed	Cycles
A. Standing out and its parts						
Standing out	0.798	0.107	0.813	0.800	0.797	82,291
Unique, on her earlier publications	0.161	0.105	0.145	0.158	0.164	82,291
Unique, new to her, established in *PubMed*	0.562	0.127	0.568	0.571	0.551	82,291
Unique, new to *PubMed*	0.075	0.059	0.063	0.071	0.082	82,291
Unique and near	0.194	0.100	0.183	0.185	0.205	82,291
Unique and far	0.605	0.171	0.615	0.615	0.591	82,291
Standing out (placebo study section)	0.950	0.031	0.955	0.949	0.953	82,291
Standing out without check tags	0.887	0.086	0.905	0.892	0.879	82,195
Other tagged applications in the pool	151.7	117.6	129.0	143.0	163.3	82,291

B. Correlations

	Standing out	Pivoting	New MeSH pairs	New noun phrases	Pool size
Standing out		0.250	0.150	-0.035	-0.310
Unique, on her earlier publications	0.164	-0.873	-0.252	-0.074	-0.136
Unique, new to her, established in *PubMed*	0.601	0.769	-0.121	-0.052	-0.174
Unique, new to *PubMed*	0.202	0.347	0.981	0.181	0.062
Unique and near	-0.382	-0.028	0.008	0.084	0.472
Unique and far	0.842	0.172	0.088	-0.070	-0.468
Standing out (placebo study section)	0.357	0.113	0.095	-0.008	0.068
Standing out without check tags	0.853	0.175	0.064	-0.064	-0.319
Cycles			80,579		

Notes: The unit of analysis is the grant cycle, on the cycles of Table E2. A pair is two MeSH descriptors of the vocabulary common to the National Library of Medicine's indexing and to MTI that appear together on one of the cycle's acknowledging publications, counted once for each publication on which they appear. A pair is unique when no application of the cycle's pool is indexed with both of its descriptors; standing out is the share of the cycle's pairs that are unique. The unique pairs on her earlier publications appear on one of the investigator's English original research publications of the five years before the cycle starts, in any authorship position; the others are new to her and established in *PubMed* (first used more than three years before the cycle's publication appeared) or new to *PubMed* (first used at most three years before). A unique pair is near when each of its descriptors appears on some application of the pool, far otherwise. Standing out (placebo study section) is the mean, over five study sections of the cycle's evaluation year drawn at random among those with at least 20 applications with descriptors (the cycle's own study section excluded, the investigator's applications removed), of the share of the cycle's pairs that no application of that study section's pool has. Standing out without check tags is computed on the pairs that involve none of the 40 check tags of the National Library of Medicine (Le and Mork 2023), when more than five remain. New MeSH pairs is the share of the cycle's pairs first used at most three years before the publication; new noun phrases is the measure of novelty of Figure 4. The pool's size counts its applications with descriptors, the investigator's own excluded. Panel B gives Pearson correlations.

Table E2: Where the Unique Pairs Are Found, and Grant Renewal

	(1)	(2)
Standing out	-0.182**	
	(0.018)	
	[-0.0193]	
Unique, on her earlier publications		-0.051*
		(0.025)
		[-0.0053]
Unique, new to her, established in *PubMed*		-0.220**
		(0.019)
		[-0.0280]
Unique, new to *PubMed*		-0.074*
		(0.034)
		[-0.0043]
p-value, equal coefficients		<0.001
Mean of dependent variable	0.427	0.427
Nb. of grant cycles	82,291	82,291
Nb. of investigators	35,068	35,068
Adjusted R^2	0.1674	0.1683

Notes: Estimates stem from a linear probability model in which the unit of analysis is the grant cycle and the dependent variable is an indicator for whether the grant cycle was renewed, with the controls of Table 2 and indicators for the number of the investigator's publications with a MeSH descriptor in the five years before the cycle (1, 2, 3-4, 5-9, 10-19, 20-49, 50 or more). The cycles are those of the standing-out regressions of Table 2 with at least one such publication; both columns use the same cycles. Standing out and its three parts (Table E1) enter as shares of the cycle's MeSH descriptor pairs; the three parts add up to standing out. Under each coefficient, its standard error and, in brackets, the coefficient times the regressor's standard deviation on these cycles. The p-value is that of the test that the three coefficients of column 2 are equal.

Standard errors in parentheses, clustered by investigator and by grant. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Table E3: Pivoting and the Section's Portfolio

	(1)	(2)	(3)
Pivoting	-0.130**		
	(0.013)		
	[-0.0184]		
New to her, shared with the section		0.070**	0.070
		(0.027)	(0.043)
		[0.0051]	[0.0051]
New to her, absent from the section		-0.161**	-0.161**
		(0.014)	(0.028)
		[-0.0219]	[-0.0220]
Not new to her, absent from the section			-0.001
			(0.039)
			[-0.0001]
p-value, shared = absent		<0.001	<0.001
Mean of dependent variable	0.427	0.427	0.427
Nb. of grant cycles	82,291	82,291	82,291
Nb. of investigators	35,068	35,068	35,068
Adjusted R^2	0.1674	0.1681	0.1681

Notes: Estimates stem from a linear probability model in which the unit of analysis is the grant cycle and the dependent variable is an indicator for whether the grant cycle was renewed, with the controls and indicators of Table E2, on its cycles; every column uses the same cycles. The cycle's pairs (Table E1) fall into four groups: new to the investigator or not (on none of her publications of the five years before the cycle, or on at least one), and shared with the section or absent from it (on some application of the cycle's pool, or on none). Pivoting, computed on these pairs, is the share new to her; it correlates 0.9996 with the pivoting measure of Table 2. Column 2 splits it into the pairs shared with the section and those absent from it; the latter equal the last two parts of standing out in Table E2. Column 3 adds the pairs not new to her and absent from the section, the first part of standing out, so that the omitted group is the pairs not new to her and shared with the section. Every part enters as a share of the cycle's MeSH descriptor pairs; in brackets, the coefficient times the regressor's standard deviation on these cycles. The p-value is that of the test that the shared and absent parts of pivoting have equal coefficients.

Standard errors in parentheses, clustered by investigator and by grant. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Table E4: Standing Out against a Placebo Study Section

	(1)	(2)	(3)
Standing out	-0.177**		-0.226**
(actual study section)	(0.018)		(0.019)
	[-0.0189]		[-0.0241]
Standing out		0.214**	0.447**
(placebo study section)		(0.055)	(0.059)
		[0.0067]	[0.0140]
Mean of dependent variable	0.427	0.427	0.427
Nb. of grant cycles	83,052	83,052	83,052
Nb. of investigators	35,396	35,396	35,396
Adjusted R^2	0.1667	0.1659	0.1673

Notes: Estimates stem from a linear probability model in which the unit of analysis is the grant cycle and the dependent variable is an indicator for whether the grant cycle was renewed, with the controls of Table 2, on the cycles of the standing-out regressions of Table 2; every column uses the same cycles. Standing out (actual study section) is the measure of Table 2. Standing out (placebo study section) is the mean, over five study sections of the cycle's evaluation year drawn at random among those with at least 20 applications with descriptors (the cycle's own study section excluded under every institute, the investigator's applications removed), of the share of the cycle's pairs that no application of that study section's pool has. Both measures enter as shares of the cycle's MeSH descriptor pairs; in brackets, the coefficient times the regressor's standard deviation on these cycles.

Standard errors in parentheses, clustered by investigator and by grant. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Appendix F: Novelty, Risk Taking, and Grant Renewal

Section 2 distinguishes risk taking from novelty: an investigator can pursue novel work without taking large risks, and take large risks on work that is not novel. This appendix describes how we measure novelty, from the words of the titles and abstracts of a cycle's publications, and three alternative measures; it relates each measure to renewal; and it estimates the penalty on risk taking with novelty held fixed, and separately for more and less novel work.[1]

Measuring novelty. Our headline measure of novelty, the only one in the main text (Tables 1c and 1d, Figure 4), is the share of new noun phrases in a cycle's publications, from Arts et al. (2025). A noun phrase is a noun together with the words that specify what it refers to: in "vascular endothelial growth factor," the noun is "factor," and the three words before it say which factor. Noun phrases name the objects, methods, and concepts of a field, and a new one often marks a new idea. "Polymerase chain reaction" first appeared in 1986, in Kary Mullis's description of the method, and more than 116,000 publications have used it since. Most of the noun phrases of a publication are familiar ones, such as "cell proliferation" or "randomized controlled trial." Arts, Melluso, and Veugelers extract the noun phrases of the titles and abstracts of some 75 million publications indexed in OpenAlex and record the first publication to use each. In their public data, a phrase is new only in the publication that first uses it, and only if at least one later publication uses it as well; phrases that appear in a single publication are discarded. We add one condition: the later publication must share no author with the first. Otherwise an investigator's own later publications, which a renewal makes possible, could make a phrase new. For each grant cycle, we add up the new noun phrases of the publications that acknowledge the grant and divide by the number of their noun phrases. Formally, with P_c the publications that acknowledge the grant during cycle c, m_p the number of distinct noun phrases in the title and abstract of publication p, and m_p^{new} the number of those that p is the first to use and that a later publication sharing no author with p uses as well,

$$\text{Share of new noun phrases}_c = \frac{\sum_{p \in P_c} m_p^{\text{new}}}{\sum_{p \in P_c} m_p}.$$

In the median cycle, 3.2 percent of noun phrases are new; 15 percent of cycles introduce none. Because a phrase must be used again, by other authors, to count as new, publications near the end of the corpus have had less time to see their phrases taken up, and the measure declines over time; every comparison below is among cycles evaluated in the same year.

We chose this measure as our headline for two reasons. The first is that it is simple and transparent: a reader can see at once what it counts, and whether a given phrase is new can be checked against the record of publications. The second is that it is built without MeSH descriptors. Two of our four measures of risk taking, pivoting and standing out, are built from the pairs of MeSH descriptors of the cycle's publications. A measure of novelty built from the same pairs, the share of pairs that are new to *PubMed*, would differ from standing out mainly by the group against which a combination of descriptors is judged new: the funded applications of one study section in one case, all of *PubMed* in the other. Measured from text, novelty is nearly uncorrelated with pivoting and standing out (0.08 and −0.05, Table 1d), against 0.32 and 0.15 for the share of new MeSH pairs. Its correlations with extreme tail outcomes and disruption are −0.17 and −0.07.

Like our measures of risk taking, each measure of novelty is flawed in its own way, which is why we also report three alternatives. Authors choose their words, and they may choose them to make their work look newer than it is; the share of new noun phrases is exposed to that choice. Requiring a later use by other authors makes the measure depend in part on how other researchers received the work, as citations do. And OpenAlex may assign one author several identifiers, so that some later uses by the same authors count as uses by others. The *share of new MeSH pairs* follows Boudreau et al. (2016): the share of the pairs of MeSH descriptors of the cycle's publications, counted once for each publication, that first appeared in *PubMed* at most three years before the publication. Its descriptors are assigned by the indexers of the National Library of Medicine, who describe the content of a publication independently of how its authors chose to present it. But the indexers work from

[1]We thank the Department Editor, the Associate Editor, and the referees for pressing us on the distinction between risk and novelty.

a controlled vocabulary, which may recognize a new line of research only some time after it appears, and the measure shares its raw material with pivoting and standing out.

Semantic distance, from the same release as our headline measure, compares each publication with everything published in the previous five years. SPECTER, a language model trained on scientific publications (Cohan et al. 2020), represents the title and abstract of each publication by a vector, trained so that a publication lies closer to the publications it cites than to those it does not. A publication's distance is one minus the cosine similarity between its vector and that of the most similar publication of the previous five years, and the measure of a cycle is the average distance of its publications. The measure is hard to explain, and it treats citation as a sign of intellectual proximity. That assumption has merit, but publications cite one another for many reasons, acquaintance and strategy among them, so the measure mixes intellectual proximity with social proximity. The *share of new noun-phrase combinations* (Arts et al. 2025) counts the pairs of noun phrases that appear together for the first time, summed over the cycle's publications and divided by the number of possible pairs. It treats every pair of phrases in a title and abstract alike, related or not, so it can count as new a juxtaposition that holds no new idea. Arts, Melluso, and Veugelers also count new words and new combinations of words; most publications introduce no new word, and combinations of words overlap with noun phrases, so we leave both aside. Each of the four measures is defined for 98 to 100 percent of the cycles of Table 2. They are weakly correlated with one another: the share of new noun phrases correlates 0.28 with new noun-phrase combinations and 0.19 with new MeSH pairs, and semantic distance is nearly unrelated to the other three (correlations between −0.08 and 0.05).

Novelty and renewal. Figure F1 shows the distribution of each measure for renewed and non-renewed cycles. On three of the four measures, renewed cycles are more novel: the share of new noun phrases averages 0.044 for renewed cycles and 0.035 for the others, the share of new MeSH pairs 0.090 and 0.076, the share of new noun-phrase combinations 0.121 and 0.105. Semantic distance barely differs (0.124 and 0.126). These raw comparisons mix in the passage of time. The renewal rate fell from 50 percent among cycles evaluated before 1996 to 27 percent among those evaluated after 2008, and the three measures that count new terms or combinations fell too: the share of new noun phrases from 0.053 to 0.023.

Figure F2 therefore compares cycles evaluated in the same year, with the controls of Table 2. Each panel comes from one regression of renewal on indicators for the ventiles of one measure of novelty, computed within evaluation year. The share of new noun phrases rises steadily with renewal: relative to the middle of the distribution, the probability of renewal is lower by 7 to 9 percent of the mean rate between the 10^{th} and 20^{th} percentiles, and higher by 3 to 5 percent in the three most novel ventiles. New noun-phrase combinations follow a similar pattern, most of it among the least novel cycles (−7 to −8 percent in the three bottom ventiles). New MeSH pairs show no gradient: apart from the least novel ventile, renewed 8 percent more often than the middle, renewal varies within a few percent across the distribution. Under semantic distance, renewal declines with novelty: below the median, renewal is up to 5 percent above the middle, and above the median up to 5 percent below it, significantly so in two of the ten upper ventiles.

Whether novelty is rewarded thus depends on how it is measured: the reward appears with the two measures built from noun phrases, and not with new MeSH pairs or semantic distance. NIH's renewal decisions appear favorable to novel work under the two measures built from noun phrases, neutral under new MeSH pairs, and mildly unfavorable under semantic distance. Boudreau et al. (2016) reached a different conclusion in another setting: when they randomly assigned evaluators to the proposals of a grant competition at a research university, evaluators gave lower scores to more novel proposals, measured, as in Panel B of Figures F1 and F2, by new combinations of MeSH descriptors.

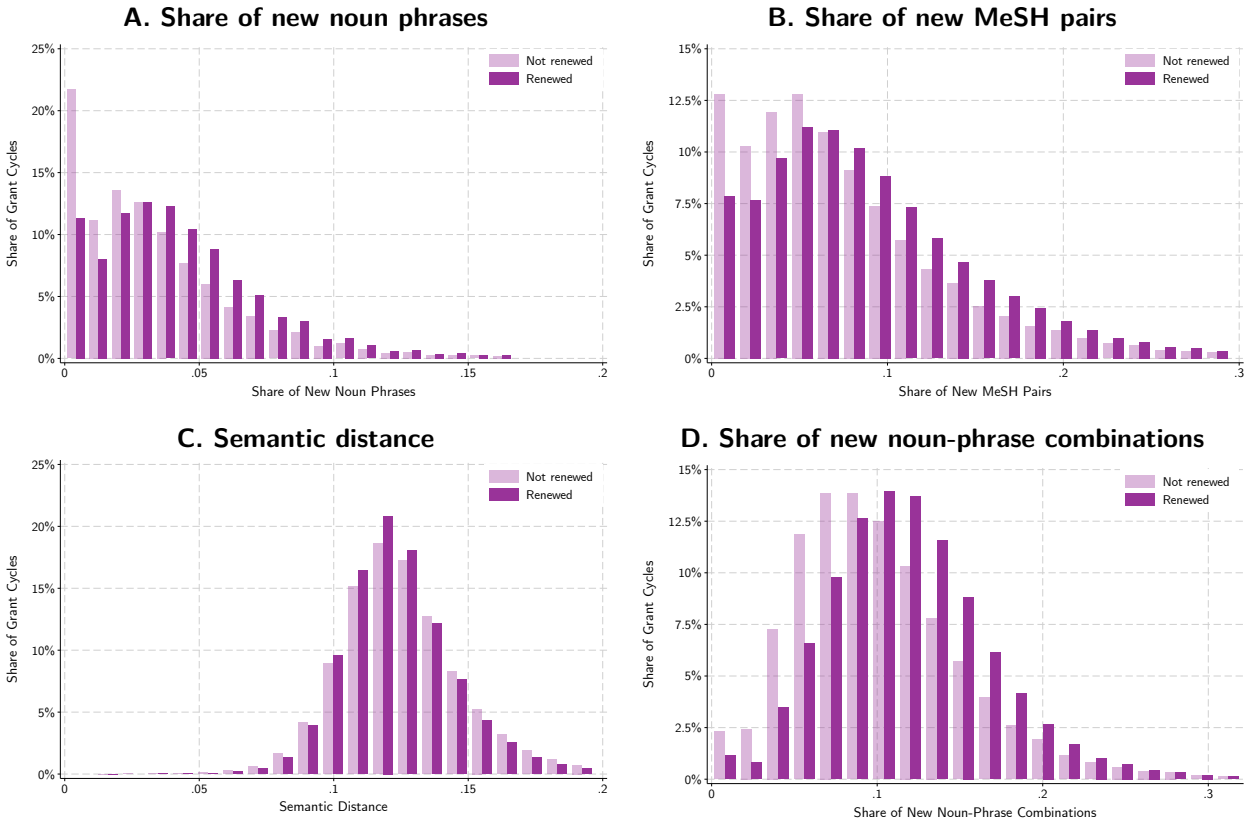
Novelty and the penalty on risk taking. Table F1 relates renewal to an indicator for high novelty, a share of new noun phrases above the median of the cycles evaluated the same year, and to each of our four measures, on the cycles of Table 2 for which the share is defined. The first column of each measure includes the indicator alone. Cycles of high novelty are renewed 1.9 to 2.0 percentage points more often than the others, an increase of 4.2 to 4.6 percent relative to the mean renewal rate. The second column adds the risk measure. The coefficient on high novelty falls by an eighth with extreme tail outcomes and by a twentieth with disruption, the two measures with which novelty is negatively correlated, rises by a sixteenth with pivoting, and does not change with standing out. The coefficient on each risk measure changes by about 3 percent at most relative to the same regression without novelty: from −0.0143 to −0.0141 for extreme tail outcomes, −0.0129 to −0.0128 for disruption, −0.1111 to

-0.1140 for pivoting, and -0.1800 to -0.1809 for standing out. Figure 4 makes the same point with the ventiles of both measures: each risk measure's gradient changes little when novelty is held fixed, and novelty's gradient is close to the one in Panel A of Figure F2.

The third column adds the interaction of the risk measure with high novelty, the measure centered at its mean, so that the coefficient on the risk measure is its association with renewal among cycles below the median of novelty and the coefficient on high novelty is unchanged. Among cycles above the median, the penalty is steeper on all four measures: by 15 percent for extreme tail outcomes (an interaction of -0.0020), 22 percent for disruption (-0.0025, significant at the 5 percent level), 9 percent for pivoting (-0.009), and 55 percent for standing out (-0.079, significant at the 1 percent level). Only the differences for disruption and standing out are statistically significant, but on none of the four measures does the 95 percent confidence interval of the penalty above the median, as a share of the penalty below it, admit a penalty more than 30 percent milder (delta method). In percentage points of renewal per standard deviation of the measure, the penalty is 2.9 below the median of novelty and 3.4 above it for extreme tail outcomes, 2.7 and 3.3 for disruption, 1.5 and 1.7 for pivoting, and 1.5 and 2.4 for standing out.

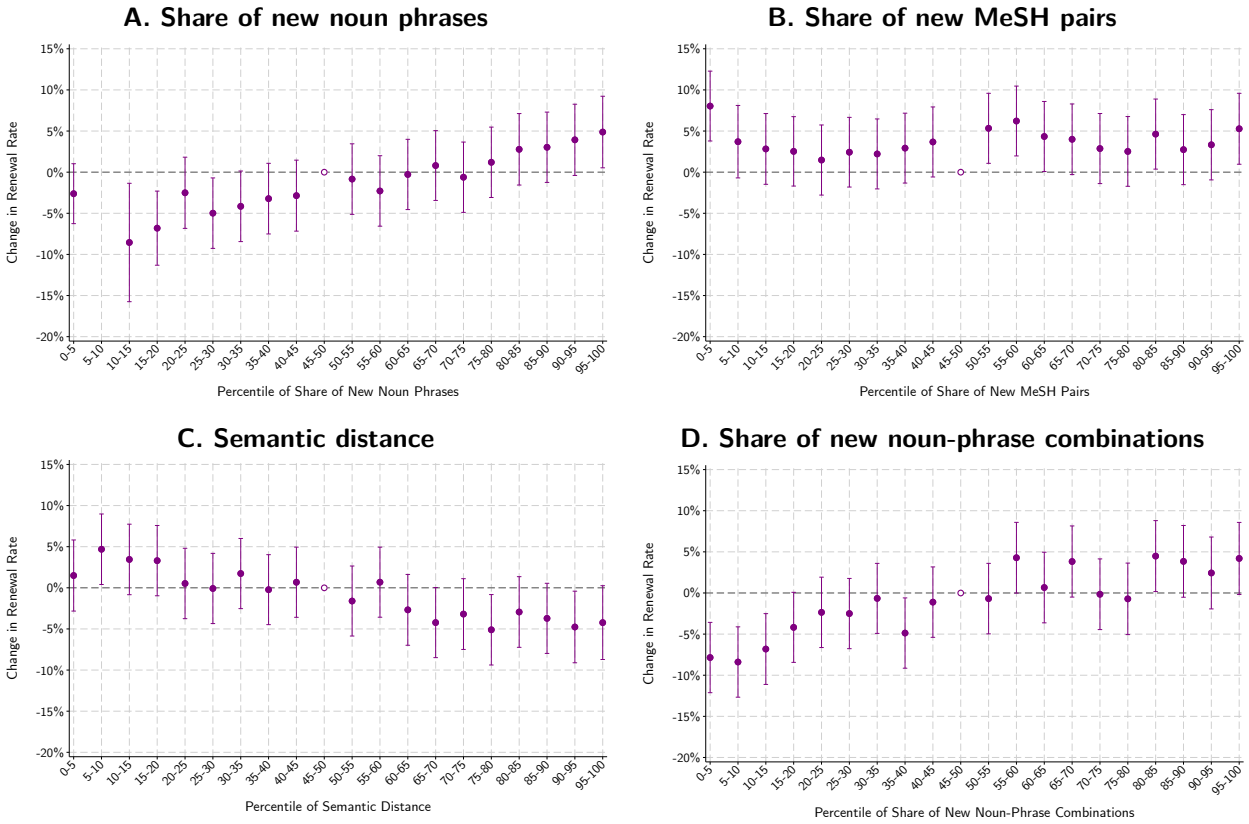
Interpretation. Novel work is not penalized the way risky work is. Measured from the words of titles and abstracts, novelty is associated with a higher probability of renewal, a gap of 4 to 5 percent of the mean rate between cycles above and below the median; measured otherwise, it is associated with renewal rates close to those of less novel work. The association of novelty with renewal and the penalty on risk taking are nearly independent of one another: holding either fixed leaves the other nearly unchanged, as the weak correlations between our measures of risk taking and of novelty would lead one to expect. Among cycles more novel than the median, the penalty on risk taking is no milder on any of our four measures and steeper in every point estimate, though the difference is statistically significant only for disruption and standing out.

Figure F1: Four Measures of Novelty, by Renewal Status



Notes: Histograms by renewal status, among cycles with at least one publication acknowledging the grant, in 20 bins of equal width from the minimum to the 99th percentile of each measure. Each bar gives the share of the cycles of its group (renewed or not renewed) that fall in the bin. Panel A: the new noun phrases of the cycle's publications, as a share of all their noun phrases (Arts et al. 2025), our headline measure, in which a phrase is new only if a later publication that shares none of its authors uses it as well. Panel B: the share of the MeSH descriptor pairs of the cycle's publications, one instance per publication and pair, that first appeared in *PubMed* at most three years before the publication. Panel C: the mean over the cycle's publications of one minus the highest SPECTER cosine similarity to a publication of the previous five years. Panel D: the pairs of noun phrases that appear together for the first time, summed over the cycle's publications and divided by the possible pairs of their noun phrases.

Figure F2: Novelty and Grant Renewal



Notes: Coefficient estimates, with 95 percent confidence intervals, from linear probability models in which the unit of analysis is the grant cycle and the dependent variable is an indicator for whether the grant cycle was renewed, with the controls of Table 2. Each panel comes from one regression with indicators for 19 of the 20 ventiles of one measure of novelty, computed within evaluation year, the 45th to 50th percentile omitted (hollow marker). Ventiles with fewer than 100 cycles are not plotted: in Panel A, the 15 percent of cycles that introduce no new noun phrase all fall in the first ventile, which leaves the second nearly empty. Coefficients and confidence intervals are divided by the mean renewal rate of the estimation sample; one vertical scale for the four panels. Measures as in Figure F1. Standard errors are clustered by investigator and by grant.

Table F1: Novelty, Risk Taking, and Grant Renewal

	Extreme tail outcomes			Disruption			Pivoting			Standing out		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
High novelty	0.0186**	0.0164**	0.0164**	0.0190**	0.0180**	0.0179**	0.0186**	0.0197**	0.0197**	0.0199**	0.0202**	0.0202**
	(0.0033)	(0.0033)	(0.0033)	(0.0031)	(0.0031)	(0.0031)	(0.0030)	(0.0030)	(0.0030)	(0.0033)	(0.0033)	(0.0033)
Risk measure		-0.0141**	-0.0130**		-0.0128**	-0.0116**		-0.1140**	-0.1096**		-0.1809**	-0.1447**
		(0.0009)	(0.0012)		(0.0008)	(0.0010)		(0.0117)	(0.0151)		(0.0187)	(0.0232)
		[-0.0318]	[-0.0294]		[-0.0301]	[-0.0272]		[-0.0161]	[-0.0155]		[-0.0191]	[-0.0152]
Risk measure × high novelty			-0.0020			-0.0025*			-0.0094			-0.0793**
			(0.0014)			(0.0012)			(0.0206)			(0.0306)
			[-0.0046]			[-0.0060]			[-0.0013]			[-0.0084]
Mean of dependent variable	0.446	0.446	0.446	0.424	0.424	0.424	0.423	0.423	0.423	0.429	0.429	0.429
Share of cycles of high novelty	0.514	0.514	0.514	0.501	0.501	0.501	0.499	0.499	0.499	0.500	0.500	0.500
Nb. of investigators	34,635	34,635	34,635	38,416	38,416	38,416	38,713	38,713	38,713	34,780	34,780	34,780
Nb. of grant cycles	87,185	87,185	87,185	98,288	98,288	98,288	99,453	99,453	99,453	81,353	81,353	81,353
Adjusted R^2	0.1632	0.1657	0.1657	0.1699	0.1724	0.1725	0.1682	0.1690	0.1690	0.1667	0.1677	0.1677

Notes: Estimates stem from a linear probability model in which the unit of analysis is the grant cycle and the dependent variable is an indicator for whether the grant cycle was renewed, with the controls of Table 2. High novelty: a share of new noun phrases (Figure F1, Panel A) above its median, computed within evaluation year on every cycle where it is defined. Each column is a separate regression; the three columns of a measure use the same cycles, those of Table 2 for which the risk measure and the share of new noun phrases are defined. The first column of each measure includes the indicator for high novelty alone. The second adds the risk measure, continuous (extreme tail outcomes and disruption in units of 10 percentile points, pivoting and standing out as shares of MeSH descriptor pairs). The third adds their interaction, with the measure centered at its mean on these cycles, so that the coefficient on high novelty is the gap in renewal at the measure's mean and the coefficient on the measure is its association with renewal among cycles below the median of novelty. Under the measure and the interaction, in brackets, the coefficient times the standard deviation of the measure on these cycles.

Standard errors in parentheses, clustered by investigator and by grant. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Appendix G: The Timing and Persistence of Risk Taking

Our measures of risk taking are built from the publications that acknowledge a grant during the cycle under review. Risk taking, as we define it in Section 2, is a choice the investigator makes when she commits to a line of work, before any of these publications exist. A skeptic could object that projects resolve their uncertainty as they unfold, so that the risk visible in a cycle's output need not be the risk the investigator took on at its start, and that our measures might record how the cycle turned out rather than what was attempted.[1] In one version of this objection, an investigator whose project goes badly takes a larger gamble near the end of the cycle; grants that are failing would then look risky because they are failing, and the penalty we estimate would reflect the order of events.

This appendix sets out what our interpretation of the results requires, and brings three kinds of evidence to bear on it.

What the interpretation requires. Risk taking need not stay constant within a cycle, and our interpretation does not require it to. Two of our measures, extreme tail outcomes and disruption, are built from the citations that the cycle's publications go on to receive, and so reflect how the uncertainty of the work was resolved. Changes in an investigator's risk taking during the cycle that are unrelated to the grant's prospects add noise to the measures, and noise, if anything, weakens the association we estimate. What the interpretation requires is narrower: that risk taking not change during the cycle in a way that tracks the prospect of renewal. Investigators who take larger gambles as their projects falter would produce a penalty where there is none; investigators who retreat to safer work would mask one.

Nor does the interpretation require the work proposed for the next cycle, which we do not observe, to be as risky as the cycle's. Our claim concerns the record that reviewers evaluate. When a grant comes up for renewal, NIH instructs the study section to consider the progress made in the last funding period, and much of that progress is reported in the publications that acknowledge the grant. Work about to appear typically figures in the renewal application itself, alongside the more far-reaching work that the new funds are meant to support. We do not know whether reviewers also penalize risk in what is proposed for the next cycle. The penalty on the record is also what bears on the incentive to take risk in the first place: an investigator deciding how ambitious to be anticipates how the record of her cycle will be judged when the grant comes up for renewal, which is the sense in which tolerance for early failure shapes exploration (Manso 2011).

The institutional setting also limits how far an investigator can stray from the original plan within a cycle. An R01 funds a set of specific aims, and a change in the direction, aims, or objectives of the approved project requires the prior approval of the funding institute (NIH Grants Policy Statement, section 8.1.2.5). The renewal is then judged against those aims.

The early and the late years of the cycle. If the measures recorded how a cycle turned out rather than what was attempted, their association with renewal should be concentrated in the publications that appear late, when the grant's prospects are clearer. Table G1 computes each measure three times: on all the publications of the cycle, on those of its first two years, and on those of its last two years, the evaluation year and the year before. The first two years are the closest to what the investigator originally proposed. The sample is the cycles that last at least four years, so that the two windows do not overlap, and on which the measure is defined in all three versions: from 34,701 cycles for extreme tail outcomes to 53,985 for pivoting. The controls are the same in all three columns: those of Table 2, with the output controls computed separately for the first two years, the last two years, and the years between. The only difference across the columns is therefore the set of publications behind the measure.

Every measure is associated with a lower probability of renewal in all three versions, at the 1 percent level. A measure computed on two years of publications is distributed differently from one computed on the whole cycle, so we compare the effects of a one-standard-deviation increase. The first two years carry between 56 percent (extreme tail outcomes) and 75 percent (disruption) of the effect of the full-cycle measure; the last two years

[1] We are grateful to the Associate Editor and to an anonymous referee for pressing this point.

carry between 79 percent (standing out) and 112 percent (disruption). The penalty is present in the publications of the early years, well before the renewal, and it is not concentrated at the end of the cycle.

Two cautions apply. Each window rests on few publications, three to six on average, so the early and late versions of a measure are only moderately correlated, from 0.25 for extreme tail outcomes to 0.57 for standing out; as argued above, variation of this kind does not by itself bias the estimates. And the cycles with publications in both windows are productive ones, with renewal rates of 47 to 51 percent against 42 percent in our sample.

Risk measured before the cycle began. A more direct test uses risk measured on publications that appeared before the cycle began. Nothing that happened during the cycle, whether the resolution of its uncertainty, a competitor's discovery, or a late gamble, can have shaped them. For a grant's first cycle, we compute each measure on the investigator's original research publications of the five years before the grant's first year, in any author position, most of which have nothing to do with the grant. For a later cycle, we use the measure of the grant's previous cycle. Table G2 reports, for each measure, three regressions on the same cycles: the cycle's own risk alone, risk before the cycle alone, and the two together, with the controls of Table 2 and indicators for the number of publications behind the earlier measure.

Risk measured before the cycle began predicts that the grant will not be renewed. One standard deviation of the investigator's record before the grant is associated with a probability of renewal lower by 1.5 percentage points for extreme tail outcomes, 1.2 for disruption, 0.5 for pivoting, and 1.3 for standing out; one standard deviation of the grant's previous cycle, by 1.2, 1.8, 0.5, and 0.7 points. These compare with 3.4, 3.1, 1.3, and 2.1 points for the cycle's own risk. All are statistically significant at the 5 percent level.

This association does not arise because the earlier record predicts the cycle's own risk. That prediction is weak: the investigator's record before the grant barely predicts the extreme tail outcomes of the grant's first cycle (a slope of 0.04, with the controls of Table 2). When both enter together, each keeps most of its association with renewal. The cycle's own risk keeps 93 to 98 percent of its coefficient for every measure, and the earlier record keeps 70 to 90 percent of its own for extreme tail outcomes and disruption, and about half for standing out's record before the grant. For pivoting, and for standing out in the previous cycle, much of the earlier record's association runs through the cycle's own risk, and what remains is not significant.

Like every estimate in the paper, these are conditional correlations, and our data do not allow us to distinguish between two interpretations. Reviewers may react to the investigator's record as a whole, which they see in the biographical sketch that accompanies every application, or the earlier record may carry traits of the investigator that our controls do not capture. In both cases, the association between risk and renewal is not confined to measures built from the realized outcomes of the cycle under review, and the cycle's own risk is penalized beyond what the investigator brought to it.

Persistence across cycles. Table G3 describes how risk taking persists across the successive cycles of a grant. Each cell is the correlation between a measure in one cycle and the same measure k cycles later, after removing the evaluation-year effects alone or all the controls of Table 2. With year effects only, the correlation between consecutive cycles ranges from 0.24 (extreme tail outcomes) to 0.43 (standing out), and it fades slowly: four cycles apart, it still ranges from about 0.14 to 0.23, far more than one step of persistence repeated four times would produce. Risk taking thus has a durable component, consistent with an investigator's ambition being part of how she works, not a draw that varies from one cycle to the next. With all the controls, persistence between consecutive cycles falls to between 0.09 and 0.29. The output controls remove much of the persistence of extreme tail outcomes and disruption, both of which rise with the number of publications, and little of that of pivoting and standing out. Persistence is nearly the same when we keep only the pairs of cycles that start at the grant's first cycle, so the first cycle is not special.

Two features limit what the table can show. A pair of cycles k apart exists only if the grant was renewed k times in a row, so persistence is measured on grants that survived repeated renewals. And persistence is partial, which is why the analysis does not rely on risk carrying over from one cycle to the next.

Risk taking across the investigator's work. Persistence across cycles follows one grant over time. The same question can be asked of the investigator's work at a single moment.[2] An investigator with several projects might balance them, keeping one safe while another takes a long shot, as an investor diversifies a portfolio. If investigators hedge in this way, the risk of one grant is a poor guide to how much risk the investigator takes overall, and a penalty on a risky grant costs her less than it seems, since her safer work continues. Our measures are computed grant by grant, so an investigator who balanced a risky grant with a safe one would appear in our data as one risky grant and one safe grant, and the penalty would concern the renewal of the risky one. Whether investigators hedge at all can be examined directly. For each cycle, we compute the measures a second time on the investigator's other work of the same years, in two versions. The first uses all her other publications: her original research publications in English published during the cycle, in any author position, that do not acknowledge the grant, twelve per cycle on average. The second keeps only those that acknowledge at least one of her other NIH grants, all these grants taken together. Hedging concerns the projects an investigator runs, and the publications acknowledging her other grants are the output of her other funded projects. All her other publications are a broader record, which also includes papers on which she is a middle author on a collaborator's project, work funded by other sources or by none, and papers whose link to one of her grants is missing from NIH's records. If investigators hedge, the offsetting should be clearest in the second version, which is available for 37 to 44 percent of the cycles of Table 2, depending on the measure. Pivoting on the other work is measured against the grant's own baseline, the investigator's publications of the five years before the cycle. Standing out is left out, since it compares a grant's work with the funded applications of its own study section, and the investigator's other work has no study section of its own. Hedging predicts that the risk of the grant's work and the risk of the other work go in opposite directions; a consistent style of research predicts that they go together.

They go together. In Table G4, we regress the grant's measure on the same measure computed on the other work, with the controls described in the table's notes, among them indicators for the number of publications behind each measure, since the spread of citation percentiles and the highest disruption percentile rise with the number of publications. For each measure, the other work is all of it in the first two columns and the work of her other grants in the last two, and the second column of each pair includes grant fixed effects, so that the comparison is among the successive cycles of the same grant. Across investigators, with all the other work (columns 1, 5, and 9), one standard deviation more risk in the other work is associated with 0.09 standard deviation more risk in the grant's work for extreme tail outcomes, 0.15 for disruption, and 0.40 for pivoting. With the work of her other grants alone (columns 3, 7, and 11), the associations are about as large: 0.10, 0.15, and 0.34. A positive association across investigators would arise if some investigators took more risk in all their work, whether or not each of them also balanced her projects. For a sharper test, we compare the cycles of the same grant: when a grant's work is riskier than in its other cycles, is the investigator's other work safer? It is not. With all the other work (columns 2, 6, and 10), the association stays positive, at 0.025 standard deviation for extreme tail outcomes, 0.048 for disruption, and 0.266 for pivoting, all significant at the 1 percent level. With the work of her other grants (columns 4, 8, and 12), the version closest to hedging across her funded projects, it is 0.031 (significant at the 5 percent level), 0.042, and 0.191 (both at the 1 percent level).[3] Within a grant, the risk of the investigator's other work thus barely moves with the grant's for extreme tail outcomes and disruption, moves with it for pivoting, and in no case moves against it. We find no evidence that investigators offset a risky grant with safer work elsewhere. The positive association may reflect a style of research, as the persistence of risk across cycles suggests, but also what an investigator's projects share: topics, collaborators, and publications from one project that acknowledge another grant. For pivoting, part of it comes from the baseline the two measures share, since an investigator whose earlier record is thin or narrow appears to pivot in all her work. Whatever its source, the association is the opposite of what hedging predicts.

Whether reviewers penalize the risk of the project or the risk of the person is a separate question, and it bears on the interpretation of Table G2. Study sections see the investigator's other work in her biographical sketch. In Table G5, we enter the risk of the grant's work and the risk of her other work in the same renewal regression, in

[2] We are grateful to an anonymous referee for raising this question.

[3] We do not use investigator fixed effects. When two of an investigator's grants are reviewed in overlapping cycles, the other work of each includes the publications of the other, so that within the investigator the two cycles are mirror images, the riskier grant having the safer one in its other work, and a comparison within investigators shows a negative association whatever she does. With investigator fixed effects, the coefficient on her other grants' work is indeed negative for extreme tail outcomes and disruption (-0.10 and -0.09), and 84 to 87 percent of those cycles have a concurrent cycle of hers on another grant; on the cycles without one, no estimate is negative and statistically significant.

the layout of Table G2. For extreme tail outcomes and disruption, the grant's risk keeps 99 and 94 percent of its coefficient when the risk of the other work is held fixed, much as it kept 93 to 98 percent against the record before the cycle, and the risk of the other work is associated with a much smaller decline in renewal: 0.4 and 1.4 percentage points per standard deviation, against 3.2 and 2.8 for the grant's own risk. For pivoting, the grant's risk keeps two thirds of its coefficient (69 percent with indicators for the number of baseline publications), and pivoting in the other work is associated with lower renewal about as strongly as pivoting in the grant's work, 1.3 against 1.0 percentage points per standard deviation. For extreme tail outcomes and disruption, then, the association with renewal concerns the grant's work; for pivoting, it extends to a move away from the investigator's earlier work wherever that move appears.

Late gambles. We also looked directly for late gambles. If investigators whose grants are failing take larger risks near the end of the cycle, the change in a measure between the first two years and the last two should be larger in the cycles that are not renewed, at the same output. We find no support for this. For pivoting and standing out, the change does not differ between renewed and non-renewed cycles. For extreme tail outcomes and disruption, the measures rise more in the cycles that are not renewed, but because their late publications are weaker: the citation percentiles of their median and weakest publications fall more from the early years to the late ones, while that of their best publication does not rise. A falling floor widens the range of citation percentiles, and weakly cited publications more often have a high disruption percentile, so both measures rise without any gamble. The estimates are available from the authors.

Interpretation. No analysis of published output can recover the intentions behind a grant application, and we do not claim to. What the evidence shows is that the association between risk and renewal does not depend on the publications that appear late in the cycle, that it extends to risk measured before the cycle began, and that it shows no sign of having been produced by gambles taken as grants fail. Nor do investigators appear to offset a risky grant with safer work elsewhere: the risk of a grant's work rises with the risk of the investigator's other work of the same years. We therefore read our measures as records of the risk that a cycle's work carried, which reviewers have before them when they decide. This reading requires that the record reflect the choices investigators made during the cycle rather than their reactions to how it was going, and nothing in the evidence assembled here suggests otherwise.

Table G1: Risk Taking Early and Late in the Grant Cycle

	Full cycle (1)	First two years (2)	Last two years (3)
A. Extreme tail outcomes			
Risk measure	-0.0126**	-0.0069**	-0.0099**
	(0.0014)	(0.0014)	(0.0014)
Effect of one SD	-0.0263	-0.0148	-0.0214
Std. dev. of the measure	2.093	2.153	2.161
Mean papers behind the measure	12.3	4.7	5.5
Mean of dependent variable	0.514	0.514	0.514
Nb. of investigators	17,260	17,260	17,260
Nb. of grant cycles	34,701	34,701	34,701
B. Disruption			
Risk measure	-0.0161**	-0.0104**	-0.0157**
	(0.0012)	(0.0010)	(0.0010)
Effect of one SD	-0.0334	-0.0251	-0.0373
Std. dev. of the measure	2.069	2.425	2.377
Mean papers behind the measure	9.3	3.4	4.3
Mean of dependent variable	0.469	0.469	0.469
Nb. of investigators	23,711	23,711	23,711
Nb. of grant cycles	52,043	52,043	52,043
C. Pivoting			
Risk measure	-0.1987**	-0.0976**	-0.1647**
	(0.0195)	(0.0141)	(0.0164)
Effect of one SD	-0.0248	-0.0155	-0.0231
Std. dev. of the measure	0.125	0.159	0.140
Mean papers behind the measure	9.7	3.5	4.4
Mean of dependent variable	0.471	0.471	0.471
Nb. of investigators	24,312	24,312	24,312
Nb. of grant cycles	53,985	53,985	53,985
D. Standing out			
Risk measure	-0.2896**	-0.1360**	-0.1972**
	(0.0327)	(0.0234)	(0.0259)
Effect of one SD	-0.0253	-0.0150	-0.0200
Std. dev. of the measure	0.087	0.110	0.101
Mean papers behind the measure	9.4	3.5	4.4
Mean of dependent variable	0.475	0.475	0.475
Nb. of investigators	21,578	21,578	21,578
Nb. of grant cycles	43,844	43,844	43,844

Notes: Estimates stem from a linear probability model in which the unit of analysis is the grant cycle and the dependent variable is an indicator for whether the grant cycle was renewed. Each column of each panel is a separate regression, with one version of the measure: computed from all the publications that acknowledge the grant during the cycle (column 1), from those of the first two years of the cycle (column 2), or from those of its last two years, the evaluation year and the year before (column 3). The first two years start with the cycle's first year of payments, or the year after when that year also carries the previous cycle's last payment. Cycles of at least four years, so that the two windows do not overlap, on which the measure is defined in all three versions; the three columns of a panel use the same cycles. Extreme tail outcomes and disruption in units of 10 percentile points, pivoting and standing out as shares of MeSH descriptor pairs. The controls are the same in every column: those of Table 2, with the grant output controls (indicators for the number of acknowledging and of other publications, the log of their mean journal impact factor with an indicator for none, indicators for a publication in *NEJM*, *Science*, *Cell*, or *Nature*) computed separately for the first two years, the last two years, and the years between. The effect of one standard deviation is the coefficient times the standard deviation of the column's measure on the panel's cycles. Mean papers behind the measure counts the acknowledging publications with a citation percentile (panel A), a disruption index (panel B), or a pair of MeSH terms (panels C and D). Pivoting is measured in every column against the investigator's publications of the five years before the cycle, so late-window pivoting also reflects movement accumulated during the cycle.

Standard errors in parentheses, clustered by investigator and by grant. $\dagger p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Table G2: Renewal, the Cycle's Risk, and Risk Measured before the Cycle

	Extreme tail outcomes			Disruption			Pivoting			Standing out		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Risk in the cycle	-0.0151**		-0.0147**	-0.0132**		-0.0126**	-0.0901**		-0.0862**	-0.1925**		-0.1786**
	(0.0010)		(0.0010)	(0.0008)		(0.0008)	(0.0126)		(0.0129)	(0.0210)		(0.0221)
	[-0.0342]		[-0.0335]	[-0.0311]		[-0.0297]	[-0.0126]		[-0.0121]	[-0.0206]		[-0.0191]
Risk before the grant (first cycles)		-0.0061**	-0.0055**		-0.0064**	-0.0045**		-0.0444*	-0.0276		-0.1426**	-0.0702*
		(0.0011)	(0.0011)		(0.0012)	(0.0012)		(0.0203)	(0.0204)		(0.0300)	(0.0312)
		[-0.0147]	[-0.0132]		[-0.0123]	[-0.0086]		[-0.0053]	[-0.0033]		[-0.0129]	[-0.0063]
Risk in the previous cycle (later cycles)		-0.0054**	-0.0039**		-0.0077**	-0.0057**		-0.0405*	-0.0181		-0.0652*	-0.0105
		(0.0014)	(0.0014)		(0.0013)	(0.0013)		(0.0195)	(0.0198)		(0.0321)	(0.0328)
		[-0.0120]	[-0.0088]		[-0.0177]	[-0.0130]		[-0.0055]	[-0.0024]		[-0.0069]	[-0.0011]
Mean of dependent variable	0.440	0.440	0.440	0.418	0.418	0.418	0.419	0.419	0.419	0.429	0.429	0.429
Nb. of investigators	32,096	32,096	32,096	36,719	36,719	36,719	35,897	35,897	35,897	28,793	28,793	28,793
Nb. of grant cycles	77,775	77,775	77,775	91,127	91,127	91,127	89,554	89,554	89,554	60,214	60,214	60,214

Notes: Estimates stem from a linear probability model in which the unit of analysis is the grant cycle and the dependent variable is an indicator for whether the grant cycle was renewed. Each column is a separate regression; all the columns of a measure use the same cycles, those with the measure and its value before the cycle. For each measure, the first column includes the cycle's risk measure, the second the same measure computed before the cycle began, and the third both. Before the cycle means, for a grant's first cycle, the investigator's original research publications in English, in any author position, published in the five years before the grant's first year; for pivoting, measured against her publications of the five years before those, and for standing out, against the funded applications of the grant's institute and study section in its first year. For a later cycle, it means the measure of the grant's previous cycle. Each has its own coefficient. Every regression includes an indicator for later cycles, indicators for the number of publications behind the earlier measure (1, 2, 3 to 4, 5 to 9, 10 to 19, 20 to 49, 50 or more) separately for first and later cycles, and the controls of Table 2. Grants first observed at a renewal, whose cycle number is unknown, are excluded. Extreme tail outcomes and disruption in units of 10 percentile points, pivoting and standing out as shares of MeSH descriptor pairs. Under each coefficient, in brackets, the effect of a one-standard-deviation increase: the coefficient times the standard deviation of the variable on these cycles (for risk before the cycle, among first cycles and among later cycles).

Standard errors in parentheses, clustered by investigator and by grant. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Table G3: Persistence of Risk Taking across the Cycles of a Grant

	Cycles apart			
	1	2	3	4
Extreme tail outcomes				
Year effects only	0.241**	0.184**	0.146**	0.143**
	(0.006)	(0.009)	(0.013)	(0.021)
All controls of Table 2	0.085**	0.056**	0.036**	0.044*
	(0.006)	(0.009)	(0.013)	(0.020)
Pairs of cycles	25,885	11,620	5,436	2,344
Disruption				
Year effects only	0.318**	0.254**	0.239**	0.229**
	(0.006)	(0.009)	(0.013)	(0.019)
All controls of Table 2	0.140**	0.078**	0.060**	0.044*
	(0.006)	(0.009)	(0.012)	(0.019)
Pairs of cycles	28,883	12,958	6,056	2,628
Pivoting				
Year effects only	0.348**	0.236**	0.198**	0.167**
	(0.006)	(0.009)	(0.013)	(0.018)
All controls of Table 2	0.217**	0.135**	0.098**	0.074**
	(0.006)	(0.009)	(0.013)	(0.017)
Pairs of cycles	29,619	13,223	6,182	2,683
Standing out				
Year effects only	0.431**	0.333**	0.255**	0.213**
	(0.007)	(0.010)	(0.013)	(0.017)
All controls of Table 2	0.290**	0.187**	0.126**	0.089**
	(0.007)	(0.009)	(0.012)	(0.016)
Pairs of cycles	21,975	9,736	4,525	2,001

Notes: Each cell is the coefficient of a regression of a measure at cycle $c + k$ on the same measure at cycle c of the same grant and investigator, using every pair of cycles k apart. Both are residualized, on the evaluation-year fixed effects only or on all the controls of Table 2, then standardized to mean 0 and standard deviation 1: a coefficient of 0.24 means that a cycle one standard deviation above the mean is followed, k cycles later, by a cycle 0.24 standard deviations above it on average. A pair k cycles apart exists only if the grant was renewed k times in a row. Cycles of investigator-grants with at least two cycles where the measure and the mean journal impact factor are defined; cycles with several principal investigators and grants first observed at a renewal are excluded. Pairs that start at the grant's first cycle give nearly the same coefficients. The pairs reported are those with all controls; the version with year effects only has slightly more, since the residualization on the controls drops cycles alone in their fixed-effect cells.

Standard errors in parentheses, clustered by grant. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Table G4: The Risk of the Grant's Work and of the Investigator's Other Work

	Extreme tail outcomes				Disruption				Pivoting			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Risk of the other work	0.081**	0.023**	0.102**	0.032*	0.155**	0.047**	0.142**	0.040**	0.442**	0.280**	0.314**	0.173**
	(0.004)	(0.006)	(0.007)	(0.013)	(0.004)	(0.007)	(0.006)	(0.011)	(0.004)	(0.007)	(0.005)	(0.010)
	[0.088]	[0.025]	[0.101]	[0.031]	[0.151]	[0.048]	[0.148]	[0.042]	[0.403]	[0.266]	[0.336]	[0.191]
Other work	All	All	Other grants	Other grants	All	All	Other grants	Other grants	All	All	Other grants	Other grants
Grant fixed effects	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Mean of the grant's measure	3.708	3.903	3.808	3.968	5.990	6.142	6.072	6.214	0.737	0.724	0.717	0.706
Nb. of investigators	31,142	10,933	12,069	3,462	36,616	13,159	15,418	4,666	37,611	13,640	15,906	4,882
Nb. of grant cycles	75,642	36,941	31,960	13,111	91,692	44,992	42,797	17,874	94,840	46,818	44,534	18,774
Adjusted R^2	0.388	0.449	0.391	0.433	0.334	0.428	0.354	0.435	0.338	0.441	0.294	0.427

Notes: Estimates stem from a linear regression in which the unit of analysis is the grant cycle and the dependent variable is the risk measure of the grant's work, computed on the publications that acknowledge the grant during the cycle, as in Table 2. The regressor is the same measure computed on the investigator's other work: her original research publications in English published during the cycle, in any author position, that do not acknowledge the grant. "All" uses all of them; "Other grants" uses those that acknowledge another of her NIH grants. Pivoting on the other work is measured against the same baseline as the grant's, the investigator's publications of the five years before the cycle starts. Extreme tail outcomes and disruption in units of 10 percentile points, pivoting as a share of MeSH descriptor pairs, on both sides. Every regression includes the fixed effects of Table 2 (gender, degree, career age, evaluation year, cycle length, institute × study section), institutional funding, the investigator's prior NIH funding with an indicator for none, her career citations through the year before the cycle starts, the indicators of Table 2 for the number of publications acknowledging the grant, indicators for the number of publications behind the measure on the other work (1, 2, 3 to 4, 5 to 9, 10 to 19, 20 to 49, 50 or more), and an indicator for cycles whose window runs past 2018, where Author-ity ends. Columns 2, 4, 6, 8, 10 and 12 add grant fixed effects, so that the comparison is among the cycles of the same grant and investigator; grants observed for a single cycle leave those columns. In brackets, the coefficient times the standard deviation of the other work's measure on the column's cycles, divided by the standard deviation of the grant's measure on the same cycles.

Standard errors in parentheses, clustered by investigator and by grant. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Table G5: Renewal, the Risk of the Grant's Work, and the Risk of the Investigator's Other Work

	Extreme tail outcomes			Disruption			Pivoting		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Risk of the grant's work	-0.0141**		-0.0139**	-0.0129**		-0.0122**	-0.1032**		-0.0676**
	(0.0010)		(0.0010)	(0.0008)		(0.0008)	(0.0119)		(0.0128)
	[-0.0320]		[-0.0317]	[-0.0303]		[-0.0285]	[-0.0147]		[-0.0096]
Risk of the other work		-0.0026**	-0.0017†		-0.0081**	-0.0063**		-0.1293**	-0.0998**
		(0.0009)	(0.0009)		(0.0009)	(0.0009)		(0.0135)	(0.0146)
		[-0.0065]	[-0.0042]		[-0.0186]	[-0.0144]		[-0.0167]	[-0.0129]
Mean of dependent variable	0.443	0.443	0.443	0.422	0.422	0.422	0.420	0.420	0.420
Nb. of investigators	31,140	31,140	31,140	36,614	36,614	36,614	37,603	37,603	37,603
Nb. of grant cycles	75,640	75,640	75,640	91,687	91,687	91,687	94,824	94,824	94,824
Adjusted R^2	0.1640	0.1616	0.1641	0.1716	0.1698	0.1720	0.1668	0.1670	0.1672

Notes: Estimates stem from a linear probability model in which the unit of analysis is the grant cycle and the dependent variable is an indicator for whether the grant cycle was renewed. Each column is a separate regression; all the columns of a measure use the same cycles, those of Table 2 for which the measure is defined on the grant's work and on the investigator's other work. For each measure, the first column includes the measure of the grant's work, the second the same measure computed on all the investigator's other work (Table G4), and the third both. Every regression includes the controls of Table 2, with the investigator's career citations counted through the year before the cycle starts rather than the year before the evaluation, and the indicators for the number of publications behind the measure on the other work of Table G4. Extreme tail outcomes and disruption in units of 10 percentile points, pivoting as a share of MeSH descriptor pairs. Under each coefficient, in brackets, the effect of a one-standard-deviation increase: the coefficient times the standard deviation of the variable on these cycles.

Standard errors in parentheses, clustered by investigator and by grant. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Appendix H: The Four Measures Together and Omitted Variables

In the paper, we estimate the penalty one measure of risk taking at a time. A skeptic could object that the association between a measure and renewal reflects something else that is correlated with the measure. It could be another of our measures. The four are correlated in pairs (Table 1d): extreme tail outcomes with disruption, both built from the citations that a cycle's publications receive, and pivoting with standing out, both built from pairs of the MeSH descriptors assigned to those publications. The penalty on one measure could then be the penalty on its partner. It could also be a confounder that we do not observe, a trait of the investigator or of the project that is associated both with risk taking and with a lower probability of renewal. To answer the first objection, we include the four measures in one regression, as continuous variables. We cannot hold fixed a confounder that is not in the data, but Oster (2019) proposes a way to gauge how strongly such a confounder would have to be related to a measure to account for its association with renewal.[1]

The four measures in one regression. We estimate Table H1 on the 69,826 cycles for which all four measures are defined, 67 percent of the 103,484 cycles in our sample and between 69 and 84 percent of the cycles on which each measure is estimated in Table 2. In columns 1 to 4, each measure enters alone; in column 5, all four enter together. Every column includes the controls of Table 2 and uses the same cycles, so that a difference between a measure's coefficient alone and its coefficient in column 5 reflects the other three measures, not a change of sample. The measures are continuous: extreme tail outcomes and disruption in units of 10 percentile points, pivoting and standing out as shares of MeSH descriptor pairs. Since their coefficients are therefore in different units, we give under each, in brackets, the effect of a one-standard-deviation increase in the measure, which puts the four on a common scale.

On these cycles, the correlation is 0.39 between extreme tail outcomes and disruption and 0.24 between pivoting and standing out; across the two pairs, none reaches 0.10 in absolute value. Each measure keeps a negative association with renewal when the other three are held fixed, significant at the 1 percent level, with t-statistics from -5.1 for pivoting to -13.9 for disruption. Holding the other three fixed reduces the coefficient of disruption by 15 percent, that of extreme tail outcomes by 19 percent, that of standing out by 19 percent, and that of pivoting by 40 percent. In comparable units, a one-standard-deviation increase in each measure is associated with a probability of renewal lower by 2.9 percentage points for disruption, 2.5 for extreme tail outcomes, 1.8 for standing out, and 1.1 for pivoting, against a mean renewal rate of 45.1 percent. Compared with each measure alone on all the cycles where it is defined (Table H2, column 1), the coefficients of column 5 are 74 to 105 percent as large, since, entered alone, three of the four measures have larger coefficients on the common sample than on all the cycles where each is defined.

One measure showing through another. Had a measure owed its association with renewal to its correlation with another of our measures, it would have lost that association once the other was held fixed. Each keeps most of it: holding the other three fixed takes at most 40 percent off its coefficient. The penalty on extreme tail outcomes is therefore not the penalty on disruption, as we measure it, showing through, nor is the penalty on pivoting that on standing out. We can conclude no more than this. Each measure reflects what it is meant to capture with error, and holding the measure fixed does not hold fixed what it is meant to capture. Four noisy measures of a single dimension would each keep a coefficient when entered together, since each adds information about that dimension that the other three lack, and with a joint regression we cannot distinguish one dimension measured four times from four distinct dimensions. What each measure captures is the subject of the appendices that take each one apart. Nor do we hold fixed, in the joint regression, any trait that we do not observe: were there a trait of the investigator that raised all four measures and lowered the probability of renewal, we would estimate four negative coefficients just as well.

Unobserved confounders. Oster (2019) proposes to extrapolate from the controls we observe to those we do not. Following her, we compare the coefficient and the R^2 of a regression without controls with those of a regression that includes them all, and ask how the coefficient would change if we could also include the unobserved confounders.

[1] We are grateful to the Associate Editor for suggesting the regression with the four measures together, and to an anonymous referee for stressing the limits of Oster's test.

Her statistic, δ, is the strength of the relationship between the measure and the unobserved confounders, relative to its relationship with the observed controls, at which the coefficient would reach zero. Computing it requires a value for the R^2 that a regression including every confounder would reach. We set it at 1.3 times the R^2 with all controls, the value Oster proposes. A δ above 1 is usually taken as a sign of robustness: the unobserved confounders would have to matter more than the observed controls to bring the coefficient to zero.

In Table H2, we apply the test to each measure entered alone, as a continuous variable, on the cycles of Table 2 and with all its controls. With no control at all, the coefficient is -0.003 for extreme tail outcomes, $+0.0008$ for disruption, -0.075 for pivoting, and -0.064 for standing out; with the controls of Table 2, it is -0.014, -0.013, -0.110, and -0.177. A positive δ arises when adding controls brings a coefficient closer to zero. Here, adding them takes every coefficient further from zero, and δ is negative for all four measures: -5.2 for extreme tail outcomes, -2.9 for disruption, -30.1 for pivoting, and -7.1 for standing out. Leaving out the observed controls thus makes the penalty look smaller than it is with them. If the unobserved confounders resembled the observed controls, adding them would enlarge the penalty further. For the penalty to vanish, leaving out the unobserved confounders would have to make it look larger, the opposite of what leaving out the observed controls does, and their relationship with the measure would have to be three to 30 times as strong as that of the observed controls.

The limits of the test. The test rests on two assumptions, neither of which can be checked. The first is proportional selection: the unobserved confounders relate to the measure as the observed controls do, up to the factor δ. We observe the investigator's gender, degree, career, and funding, the institute, study section, and year of the review, and the publications of the cycle. Some of the confounders we cannot observe are of another kind, such as the reputation of the investigator among the members of her study section, and nothing guarantees that they resemble the controls. For extreme tail outcomes and disruption, the assumption is especially hard to defend. The change in their coefficients comes from the controls for the cycle's output, which are tied to how the two measures are built: both rise with the number of publications, and renewed cycles publish more, so that without these controls the penalty and the advantage of publishing more offset each other (Section 4; for disruption, Appendix C). From a change of this kind, we learn little about confounders unrelated to output. The second assumption is the value we give the maximum R^2, between 0.23 and 0.25. We do not know what share of the variation in renewal a regression with every confounder would explain. The higher that share, the smaller the deltas in absolute value, and the weaker the confounders able to bring the penalty to zero.[2] We therefore treat the deltas of Table H2 as a description of how the coefficients change when the controls we observe are added, not as a bound on the penalty.

Interpretation. Like every estimate in the paper, those of this appendix are conditional correlations, and neither analysis identifies a cause. With the joint regression, we can exclude one interpretation of the penalty: that the association between a measure and renewal is another of our measures, as we observe it, showing through. From Oster's test, we learn how the coefficients change when the controls we observe are added: they grow, so that confounders resembling those controls would enlarge the penalty, although we do not know whether the confounders we cannot observe resemble them.

[2]With a maximum R^2 of 1, the deltas are -0.37 for extreme tail outcomes, -0.21 for disruption, -2.10 for pivoting, and -0.46 for standing out: confounders working against the observed controls, from a fifth as strongly for disruption to twice as strongly for pivoting, would then bring the penalty to zero. A maximum R^2 of 1 would mean that the measure, the controls, and the confounders together determine renewal, leaving no part of the decision to chance.

Table H1: The Four Measures of Risk Taking in One Regression

	Each measure alone				All four	S.D. of measure
	(1)	(2)	(3)	(4)	(5)	
Extreme tail outcomes	-0.0135**				-0.0109**	2.257
	(0.0010)				(0.0010)	
	[-0.0304]				[-0.0247]	
Disruption		-0.0157**			-0.0133**	2.212
		(0.0009)			(0.0010)	
		[-0.0347]			[-0.0294]	
Pivoting			-0.1353**		-0.0816**	0.132
			(0.0155)		(0.0161)	
			[-0.0179]		[-0.0108]	
Standing out				-0.2286**	-0.1855**	0.096
				(0.0233)	(0.0241)	
				[-0.0220]	[-0.0178]	
Mean of dependent variable	0.451	0.451	0.451	0.451	0.451	
Nb. of investigators	30,492	30,492	30,492	30,492	30,492	
Nb. of grants	47,986	47,986	47,986	47,986	47,986	
Nb. of grant cycles	69,826	69,826	69,826	69,826	69,826	

Notes: Estimates stem from a linear probability model in which the unit of analysis is the grant cycle and the dependent variable is an indicator for whether the grant cycle was renewed. Each column is a separate regression: columns 1 to 4 include one measure, column 5 all four. Every column uses the same cycles, those for which all four measures and a mean journal impact factor are defined. The measures are continuous: extreme tail outcomes and disruption in units of 10 percentile points, pivoting and standing out as shares of MeSH descriptor pairs. Every regression includes the controls of Table 2. Under each coefficient, in brackets, the effect of a one-standard-deviation increase in the measure, that is, the coefficient times the measure's standard deviation on these cycles, which puts the four measures in the same units. The last column gives the standard deviation of each measure on these cycles, in the units of its coefficient (22.6 and 22.1 percentile points for extreme tail outcomes and disruption).

Standard errors in parentheses, clustered by investigator and by grant. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Table H2: Unobserved Confounders: Oster's δ

	Coefficient estimate	Oster's δ
Extreme tail outcomes	-0.0142**	-5.187
	(0.0009)	
Disruption	-0.0129**	-2.894
	(0.0008)	
Pivoting	-0.1097**	-30.074
	(0.0115)	
Standing out	-0.1773**	-7.060
	(0.0182)	

Notes: Each coefficient in column 1 comes from a separate linear probability model in which the unit of analysis is the grant cycle and the dependent variable is an indicator for whether the grant cycle was renewed, with the controls of Table 2 and the measure entered as a single continuous regressor: extreme tail outcomes and disruption in units of 10 percentile points, pivoting and standing out as shares of MeSH descriptor pairs. Each regression uses the cycles of Table 2 for its measure: 87,718 for extreme tail outcomes, 99,504 for disruption, 101,435 for pivoting, and 83,052 for standing out. Column 2 reports the δ of Oster (2019), the degree of selection on unobserved variables, relative to the observed controls, at which the coefficient would be zero. It is computed from the regression of column 1 and a regression of renewal on the measure alone, without any control or fixed effect, with the maximum R^2 set at 1.3 times the R^2 of the regression of column 1 (0.18 to 0.19, for a maximum R^2 of 0.23 to 0.25), using the Stata command `psacalc` after a regression with the same controls, institute $\times$ study section absorbed.

Standard errors in parentheses, clustered by investigator and by grant. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Appendix I: The Penalty and the Traits of the Investigator

The penalty in Table 2 is an average over investigators who differ in many ways. A skeptic could object that it reflects how reviewers treat some kinds of investigators, such as newcomers, women, physicians, or scientists whose work ranges across fields, more than how they treat risk. This appendix estimates the penalty separately for investigators with and without six traits, each fixed by the time the cycle starts: gender, an MD degree, being a new investigator (one whose cycle starts in the first year of her first R01-equivalent award), membership in the biomedical research elite (defined by election to prestigious societies, made precise below), the breadth of earlier research, and holding a second, concurrent R01 at the start of the cycle.[1]

The specification. Each regression adds to the specification of Table 2 an indicator for the trait and its interaction with the measure, which enters as a continuous variable, as in Figure 5. The fixed effects of Table 2 include gender and degree, which absorb these two traits' own coefficients. Every trait is measured before the cycle's work could shape it, so that none can reflect how the cycle turned out; the case of other R01 grants, below, shows why this matters.

Gender. We take each investigator's gender from, in order, our own coding of the investigators we identified by hand, our records for prominent life scientists, the gender that Genni (Smith et al. 2013) associates with the investigator's first name in the Author-ity database (Torvik and Smalheiser 2009), and, for a few investigators, a probability that the first name is a woman's, when it is below 5 percent or above 95 percent; women account for about a fifth of the cycles.

Degree. The indicator marks investigators who hold an MD and no PhD, 18 percent of the cycles, against holders of a PhD, of both degrees, or of another doctorate, as recorded in our file of NIH grants.

New investigators. Following NIH's category, a new investigator's cycle starts in the investigator's first year of any R01-equivalent award: an R01, R37, R29, R23, R56, R35, RF1, RL1, DP1, DP2, DP5, or U01. About a quarter of the cycles are new investigators'.

Elite investigators. An elite investigator had been elected to the National Academy of Sciences or to the National Academy of Medicine (formerly the Institute of Medicine), or held an appointment at the Howard Hughes Medical Institute, in the year before the cycle starts. The year of election counts, and the one-year lag covers the delay to induction and to the start of an appointment. Testing status at the end of the cycle instead would let investigators become elite on the strength of the work under review. Three percent of the cycles are elite investigators'.

Breadth of earlier research. Some scientists spread their work across many topics and others pursue one, Berlin's (1953) foxes and hedgehogs. A fox has a thinner record in any one area, which could leave reviewers less to go on, the logic the paper applies to new investigators. We measure breadth on the investigator's first- or last-author publications of the five years before the cycle, which describe her own agenda, excluding any publication that acknowledges the grant under review. For every pair of these publications, we compute one minus the cosine similarity of their MeSH descriptors, weighted by how rare each descriptor is in *PubMed* that year, and average over the pairs. The index answers a simple question: how different are two of her publications, drawn at random? Because it averages over pairs, it does not rise mechanically with the number of publications. On cycles with at least twenty publications, the index computed on random subsets of 5, 10, and 20 publications has nearly the same mean as on all of them (0.864 in every case). Across investigators, breadth rises mildly with the size of the record (rank correlation 0.21), so the regressions also include indicators for the number of publications behind the index. We call an investigator a fox when her breadth is above the median of the cycles reviewed by the same study section. The comparison is made within the study section because fields differ in how they index their publications: chemists' publications, for instance, carry very specific descriptors and share few of them, so that even a chemist with a focused agenda looks broad, and a single cutoff for all fields would make foxes of whole

[1] We are grateful to the Associate Editor and to two anonymous referees for suggesting several of these comparisons.

fields. The index needs at least five publications, so it exists for 62 percent of the cycles in Table 2, and for 49 percent of new investigators' cycles.

Another R01 at the start of the cycle. The indicator marks investigators who held another R01, R37, R29, or R23 grant in the first year of the cycle, 39 percent of the cycles. It speaks to whether reviewers treat risk differently when the investigator has other funding. The timing matters. Counted in the cycle's last year instead, another R01 is associated with a markedly smaller penalty on every measure but pivoting: the interaction is 0.0039 (standard error 0.0014) for extreme tail outcomes, 0.0024 (0.0013) for disruption, 0.0022 (0.0211) for pivoting, and 0.1004 (0.0316) for standing out, whose penalty is nearly halved. That difference comes from grants won during the cycle, which change the count for about one cycle in five. A grant won during the cycle may be won because the cycle's work went well, so the interaction would credit the cycle's success twice. Counted at the start of the cycle, as in Table I1, the interaction is smaller for extreme tail outcomes, disruption, and standing out (0.0024, 0.0013, and 0.0217), and larger for pivoting (0.0454, standard error 0.0216).

What the table shows. Table I1 reports, for each trait, the measure's coefficient among investigators without the trait and its interaction with the trait. Among investigators without each trait, every measure's coefficient is negative and statistically significant at the 1 percent level.

- *New investigators* pay a steeper penalty: by 43 percent for extreme tail outcomes and 110 percent for pivoting, both significant at the 1 percent level, and 26 percent for disruption, significant at the 5 percent level. For standing out, their penalty is 6 percent smaller, and the difference is not significant.
- *Women and men*: the interactions are close to zero for every measure (for extreme tail outcomes, 0.0004 against a penalty of −0.0143).
- *MDs*: no difference for extreme tail outcomes and disruption; for pivoting, a penalty smaller by about two-fifths, significant at the 10 percent level only, and for standing out a smaller one that is not significant.
- *Elite investigators* pay about a third of the others' penalty on extreme tail outcomes (5 percent level); on the other measures the estimates are imprecise, since 3 percent of the cycles are elite investigators'.
- *Broad earlier research* is associated with a penalty smaller by 25 percent on extreme tail outcomes (5 percent level) and 22 percent on disruption (10 percent level), and no reliable difference on pivoting and standing out.
- *Another R01 at the start of the cycle* is associated with a penalty smaller by 36 percent on pivoting (5 percent level), and no reliable difference on the other three measures.

Pervasive, with one exception. With six traits and four measures, the table holds 24 interactions; at the 5 percent level, about one would be significant by chance. Six are, three of them for new investigators. The other three each concern a single measure, and all point to a somewhat milder penalty for investigators with a longer or broader record, or with other funding. None is strong enough to carry weight on its own. The penalty applies to women and men, to physicians and PhDs, to members of the National Academies and the rest, to specialists and generalists, and to investigators with and without other funding. Its size differs in one consistent way: the penalty is steeper for new investigators, the group for which reviewers have the least to go on.

Table I1: The Penalty for Risk Taking and the Traits of the Investigator

	(1) Extreme tail outcomes	(2) Disruption	(3) Pivoting	(4) Standing out
Panel A. Gender				
Risk measure	-0.0143**	-0.0128**	-0.1088**	-0.1833**
	(0.0010)	(0.0008)	(0.0126)	(0.0200)
Risk measure × woman	0.0004	-0.0002	-0.0043	0.0279
	(0.0018)	(0.0015)	(0.0248)	(0.0367)
Share of cycles with the trait	0.202	0.207	0.207	0.204
Nb. of grant cycles	87,718	99,504	101,435	83,052
Adjusted R^2	0.1650	0.1717	0.1679	0.1667
Panel B. Degree				
Risk measure	-0.0142**	-0.0127**	-0.1182**	-0.1892**
	(0.0010)	(0.0008)	(0.0124)	(0.0195)
Risk measure × MD	-0.0001	-0.0011	0.0460†	0.0667
	(0.0018)	(0.0017)	(0.0270)	(0.0420)
Share of cycles with the trait	0.184	0.183	0.185	0.186
Nb. of grant cycles	87,718	99,504	101,435	83,052
Adjusted R^2	0.1650	0.1717	0.1679	0.1667
Panel C. Career stage				
New investigator	0.0084	0.0056	0.0579**	-0.0300
	(0.0098)	(0.0110)	(0.0195)	(0.0272)
Risk measure	-0.0129**	-0.0120**	-0.0854**	-0.1810**
	(0.0010)	(0.0008)	(0.0131)	(0.0209)
Risk measure × new investigator	-0.0056**	-0.0032*	-0.0938**	0.0115
	(0.0017)	(0.0014)	(0.0239)	(0.0331)
Share of cycles with the trait	0.239	0.251	0.253	0.258
Nb. of grant cycles	87,718	99,504	101,435	83,052
Adjusted R^2	0.1651	0.1717	0.1680	0.1667
Panel D. Status				
Elite investigator	-0.0298†	0.0022	0.0353	-0.0154
	(0.0175)	(0.0250)	(0.0464)	(0.0848)
Risk measure	-0.0146**	-0.0129**	-0.1090**	-0.1779**
	(0.0009)	(0.0008)	(0.0116)	(0.0184)
Risk measure × elite investigator	0.0097*	0.0022	-0.0282	0.0242
	(0.0039)	(0.0039)	(0.0649)	(0.1022)
Share of cycles with the trait	0.033	0.032	0.031	0.030
Nb. of grant cycles	87,718	99,504	101,435	83,052
Adjusted R^2	0.1651	0.1717	0.1679	0.1667
Panel E. Breadth of earlier research				
Broad earlier research	-0.0380**	-0.0399**	-0.0196	-0.0377
	(0.0079)	(0.0106)	(0.0191)	(0.0310)
Risk measure	-0.0165**	-0.0135**	-0.0875**	-0.1589**
	(0.0015)	(0.0013)	(0.0191)	(0.0298)
Risk measure × broad earlier research	0.0041*	0.0029†	-0.0040	0.0084
	(0.0018)	(0.0016)	(0.0261)	(0.0386)
Share of cycles with the trait	0.497	0.497	0.497	0.493
Nb. of grant cycles	54,850	61,716	63,440	51,276
Adjusted R^2	0.1614	0.1669	0.1625	0.1620
Panel F. Other funding				
Another R01 when the cycle starts	-0.0020	-0.0033	-0.0268†	-0.0128
	(0.0068)	(0.0087)	(0.0162)	(0.0257)
Risk measure	-0.0152**	-0.0133**	-0.1271**	-0.1850**
	(0.0011)	(0.0009)	(0.0139)	(0.0211)
Risk measure × another R01	0.0024	0.0013	0.0454*	0.0217
	(0.0014)	(0.0013)	(0.0216)	(0.0318)
Share of cycles with the trait	0.415	0.407	0.405	0.398
Nb. of grant cycles	87,718	99,504	101,435	83,052
Adjusted R^2	0.1651	0.1717	0.1679	0.1667

Notes: Estimates stem from a linear probability model in which the unit of analysis is the grant cycle and the dependent variable is an indicator for whether the grant cycle was renewed, with the controls of Table 2. Each column of each panel is a separate regression with one measure as a continuous regressor (extreme tail outcomes and disruption in units of 10 percentile points, pivoting and standing out as shares of MeSH descriptor pairs), an indicator for the trait named in the panel, and their interaction, on the cycles of Table 2 for which the trait is defined. Every trait is fixed when the cycle starts. Panel A: women. Panel B: an MD alone, against PhDs, MD-PhDs, and other degrees. The fixed effects of Table 2 include gender and degree, which absorb these two traits' own coefficients. Panel C: new investigators, whose cycle starts in their first year of an R01-equivalent award (R01, R37, R29, R23, R56, R35, RF1, RL1, DP1, DP2, DP5, U01). Panel D: elite investigators, members of the National Academy of Sciences or of the National Academy of Medicine (formerly the Institute of Medicine), or investigators of the Howard Hughes Medical Institute, in the year before the cycle starts. Panel E: breadth of earlier research, the mean MeSH dissimilarity between pairs of the investigator's first- or last-author publications of the five years before the cycle, above the median of the cycle's study section; defined for cycles with at least five such publications, and the regressions include indicators for their number. Panel F: another R01, R37, R29, or R23 grant held in the cycle's first year.

Standard errors in parentheses, clustered by investigator and by grant. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Appendix J: Risk Taking and Interdisciplinary Teams

Funding agencies encourage interdisciplinary research, yet their review panels appear to treat it less favorably. Among some 18,000 proposals to the Australian Research Council, Bromham et al. (2016) find that those spanning a broader range of fields were less likely to be funded. Banal-Estañol et al. (2019) find that teams whose members differ in knowledge and skills, education, or scientific ability were less likely to obtain funding from the United Kingdom's Engineering and Physical Sciences Research Council, although such teams were more likely to succeed. A skeptic could suspect that our measures of risk taking, pivoting above all, partly record the interdisciplinarity of the work, so that the penalty we document would be this penalty under another name.[1]

Interdisciplinarity can enter our data in three ways. A grant's work can move into fields that are new to the investigator, which our pivoting measure records and Appendix D dissects. The investigator can be a generalist whose earlier publications range across fields; Appendix I shows that such investigators are not penalized more than specialists for taking risks. And the team can bring together scientists from different fields, which is what interdisciplinary research most often means and what this appendix examines. The collaborators we study are those of a single principal investigator: a cycle whose starting award names several principal investigators does not enter our sample (Appendix A). We measure how far the research of the investigator's collaborators lies from her own, and we ask two questions. Is a grant cycle conducted with intellectually distant collaborators less likely to be renewed? And does the distance of the team change the penalty on risk taking?

Measuring the distance of the team. A collaborator is a coauthor on the publications that acknowledge the grant during the cycle who, during the same years, last-authored an original research article with at least one other author and without the focal investigator. Such a collaborator runs a research program of her own, which we can compare with the focal investigator's. We identify coauthors with the Author-ity disambiguation of *PubMed* authors (Torvik and Smalheiser 2009) and last authors from *PubMed*'s author lists. Students, postdoctoral fellows, and technicians, who publish mostly with the investigator, rarely qualify. Across all cycles, the publications that acknowledge a grant list 16.1 coauthors on average, of whom 6.0 qualify; among the cycles studied below, 22.8 and 8.8.

For each collaborator, we compare two bodies of work published during the cycle's years: the collaborator's original research articles in English, and the investigator's. On both sides we set aside any article that acknowledges the grant and any article the two wrote together; the investigator's articles that acknowledge her other grants are kept. The comparison is thus between two research programs, each observed outside the grant. Comparing collaborators with the grant's own publications would have brought every collaborator closer the more the grant published, mixing the composition of the team with the productivity of the grant. We represent each body of work by how often each MeSH descriptor appears in its articles, giving more weight to descriptors that are rare in *PubMed* (the TF-IDF weighting common in text analysis). The distance between the investigator and a collaborator is one minus the cosine similarity of the two representations, and the distance of the team is its average over the cycle's collaborators. It runs from 0 to 1.

The measure requires that the investigator have at least three articles with MeSH descriptors outside the grant during the cycle, since one or two articles give too thin a picture of her work, and every collaborator then looks distant. Above that floor, more articles still bring collaborators closer (the correlation between the distance of the team and the number of the investigator's articles is -0.29), so every regression includes indicators for that number. The measure is also taken during the cycle, not at its start. Whom an investigator works with during a cycle may depend on how the work is going, so the distance of the team describes the cycle, as our four measures do, rather than a trait of the investigator fixed in advance, as in Appendix I.

The measure is defined for 72,284 cycles, 70 percent of the cycles in our sample: 60 percent of those evaluated before 1996, and 76 percent of those evaluated after 2008. Most of the cycles without it have no qualifying collaborator. Among the cycles with the measure, the focal investigator published a median of 10 articles outside the grant during the cycle, 23 percent of which acknowledge another of her grants. Few of these articles involve her collaborators: in the median cycle, a collaborator coauthored, on average, 4 percent of them. Those shared articles are left out of both bodies of work when the focal investigator and that collaborator are compared. The

[1] We thank the Associate Editor and an anonymous referee for raising the question of interdisciplinary research.

distance of the team averages 0.81, with a standard deviation of 0.11. Its correlations with our four measures are small: 0.03 with extreme tail outcomes, 0.05 with disruption, 0.12 with standing out, and 0.28 with pivoting. A cycle that moves into a new field brings in collaborators from that field, whose work lies far from the investigator's other work, which accounts for the correlation with pivoting. The correlation with the breadth of the investigator's earlier research (Appendix I) is 0.11, so the distance of the team measures something other than the investigator's breadth. Figure J1 shows the distribution of the measure for renewed and non-renewed cycles. The distribution for renewed cycles lies slightly to the left, and a larger share of non-renewed cycles falls among the most distant teams (means of 0.802 and 0.814).

The distance of the team and renewal. Table J1 relates renewal to each of our measures and to an indicator for a distant team, with the controls of Table 2 and the indicators described above. We call a team distant when its distance lies above the median of the cycles reviewed by the same study section. Fields index their articles differently, and a single cutoff would make whole fields distant; Appendix I applies the same rule to breadth. Holding the risk measure fixed, distant teams are renewed 2.2 to 2.4 percentage points less often than close ones, 5 to 6 percent of the mean rate (the second column of each measure).

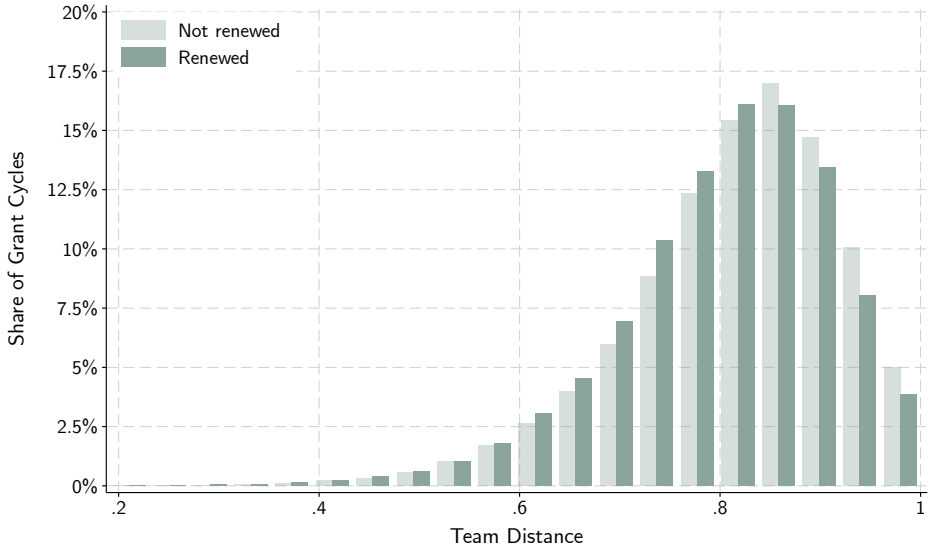
Figure J2 shows where along the distribution the association lies. It follows the design of Figure 4: each panel comes from one regression with indicators for the ventiles of one risk measure and for the ventiles of the distance of the team, and shows both sets of coefficients. In all four panels, cycles conducted with the closest collaborators do better: relative to the middle of the distribution, the probability of renewal is higher by 8.5 to 11.6 percent of the mean rate in the two closest ventiles, and by 5 to 8 percent up to the 25^{th} percentile. From the 25^{th} to the 90^{th} percentile, the distance makes little difference. Between the 90^{th} and the 95^{th} percentiles, renewal is lower by 4 to 7 percent of the mean rate, and the most distant ventile cannot be told apart from the middle. The gap in Table J1 thus comes mostly from an advantage of close teams rather than from a penalty on distant ones.

The four measures, holding the team fixed. Holding the distance of the team fixed leaves the association of each measure with renewal nearly unchanged. From the first column of each measure in Table J1 to the second, the coefficient falls by 2 percent for extreme tail outcomes (from -0.0136 to -0.0133), 3 percent for disruption (-0.0126 to -0.0122), 7 percent for standing out (-0.191 to -0.176), and 13 percent for pivoting (-0.103 to -0.089), the measure most correlated with the distance of the team. All four remain significant at the 1 percent level. In Figure J2, the gradient of each measure, with the distance of the team held fixed, is close to its gradient in Figure 4, where novelty is held fixed. In the riskiest ventile of cycles, the probability of renewal is lower by 9.3, 9.3, 12.6, and 9.4 percent of the mean rate for the four measures, against 8.2, 10.2, 14.9, and 9.3 percent in Figure 4.

The penalty for close and distant teams. In the third column of each measure in Table J1, we ask whether the penalty on risk taking depends on the distance of the team. It adds the interaction of the measure with the indicator for a distant team, the measure centered at its mean, so that the indicator's coefficient remains the gap between distant and close teams at the average level of risk taking; it is the same as in the second column. The penalty on risk taking is present for close and distant teams alike. For extreme tail outcomes, it is the same for both (an interaction of 0.0018, standard error 0.0016); for disruption, it is slightly smaller when the team is distant (0.0026, standard error 0.0016, significant at the 10 percent level only). For pivoting, it is steeper when the team is distant: -0.111, against -0.069 for close teams, an interaction of -0.042 (standard error 0.025) that is significant at the 10 percent level only. For standing out, the point estimate goes in the same direction and is imprecise (-0.050, standard error 0.038).

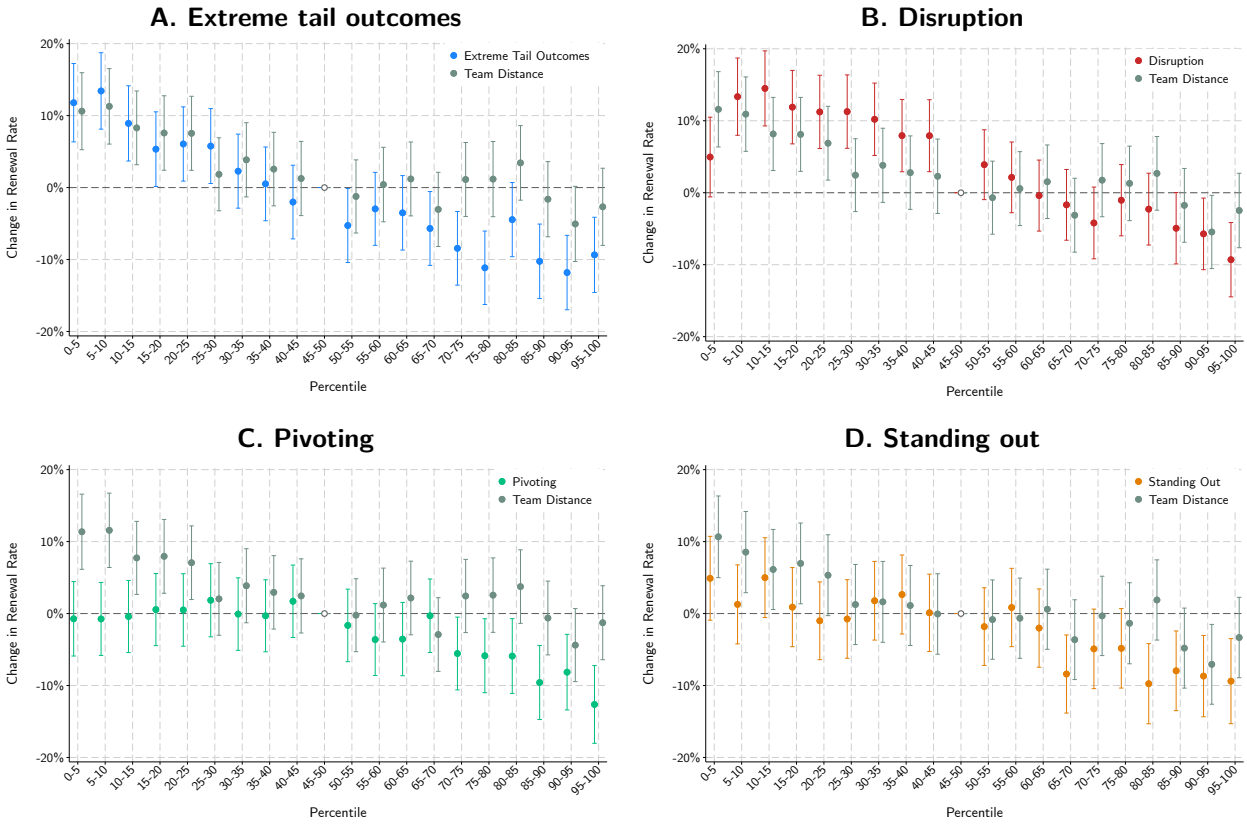
Interpretation. Interdisciplinary research enters our analysis in three ways. Moving into fields that are new to the investigator is one of the forms of risk taking we measure, and it is penalized, the more so the more distant the new fields are from her earlier work (Appendix D). Investigators whose earlier work ranges widely are not penalized more for taking risks (Appendix I). The composition of the team, studied here, has a modest association of its own with renewal: cycles conducted with intellectually distant collaborators are renewed somewhat less often, consistent with Bromham et al. (2016) and Banal-Estañol et al. (2019), although in our data the gap comes mostly from the success of close teams. That association leaves the penalty on risk taking intact: each of our four measures keeps its association with renewal when the distance of the team is held fixed, and the penalty applies whether the team is close or distant. The penalty on risk taking is not a penalty on interdisciplinary teams.

Figure J1: The Distance of the Team, by Renewal Status



Notes: The cycles for which the distance of the team is defined, by renewal status, in 20 bins of equal width from the minimum to the maximum of the measure. Each bar gives the share of the cycles of its group (renewed or not renewed) that fall in the bin. The distance of the team is the average, over the investigator's collaborators during the cycle, of one minus the cosine similarity between the MeSH descriptors of the collaborator's articles and those of the investigator's, both published during the cycle's years, excluding any article that acknowledges the grant and any article the two wrote together, with TF-IDF weights. Means for renewed and non-renewed cycles: 0.802 and 0.814.

Figure J2: Risk Taking, the Distance of the Team, and Grant Renewal



Notes: The plots present coefficient estimates, with 95 percent confidence intervals, from a linear probability model in which the unit of analysis is the grant cycle and the dependent variable is an indicator for whether the grant cycle was renewed, with the controls of Table 2 and indicators for the number of the investigator's articles with MeSH descriptors outside the grant during the cycle (3 to 10 one by one, 11-15, 16-20, 21-30, 31-50, more than 50). Each regression includes indicators for 19 of the 20 ventiles of the risk measure and 19 of the 20 ventiles of the distance of the team, all computed within evaluation year, the 45^{th} to 50^{th} percentile omitted (hollow marker); each panel shows both sets of coefficients from the same regression, the risk measure's in its color and the distance of the team's in teal. The design is that of Figure 4, with the distance of the team in the place of novelty. Coefficients and confidence intervals are divided by the mean renewal rate of the estimation sample. Standard errors are clustered by investigator and by grant.

Table J1: The Penalty for Risk Taking and the Distance of the Team

	Extreme tail outcomes			Disruption			Pivoting			Standing out		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Risk measure	-0.0136**	-0.0133**	-0.0142**	-0.0126**	-0.0122**	-0.0135**	-0.1026**	-0.0892**	-0.0690**	-0.1906**	-0.1764**	-0.1514**
	(0.0010)	(0.0010)	(0.0013)	(0.0009)	(0.0009)	(0.0012)	(0.0145)	(0.0146)	(0.0192)	(0.0234)	(0.0235)	(0.0305)
	[-0.0309]	[-0.0302]	[-0.0322]	[-0.0283]	[-0.0274]	[-0.0304]	[-0.0142]	[-0.0123]	[-0.0095]	[-0.0193]	[-0.0179]	[-0.0153]
Distant team		-0.0240**	-0.0239**		-0.0240**	-0.0238**		-0.0216**	-0.0215**		-0.0235**	-0.0235**
		(0.0040)	(0.0040)		(0.0038)	(0.0038)		(0.0038)	(0.0038)		(0.0042)	(0.0042)
Risk measure × distant team			0.0018			0.0026†			-0.0421†			-0.0505
			(0.0016)			(0.0016)			(0.0253)			(0.0382)
			[0.0041]			[0.0058]			[-0.0058]			[-0.0051]
Mean of dependent variable	0.443	0.443	0.443	0.426	0.426	0.426	0.424	0.424	0.424	0.429	0.429	0.429
Share of cycles with a distant team	0.490	0.490	0.490	0.496	0.496	0.496	0.497	0.497	0.497	0.498	0.498	0.498
Nb. of investigators	27,264	27,264	27,264	29,448	29,448	29,448	29,868	29,868	29,868	26,419	26,419	26,419
Nb. of grant cycles	64,208	64,208	64,208	69,986	69,986	69,986	71,184	71,184	71,184	57,431	57,431	57,431
Adjusted R^2	0.1662	0.1667	0.1667	0.1738	0.1743	0.1743	0.1704	0.1708	0.1709	0.1691	0.1695	0.1696

Notes: Estimates stem from a linear probability model in which the unit of analysis is the grant cycle and the dependent variable is an indicator for whether the grant cycle was renewed, with the controls of Table 2 and indicators for the number of the investigator's articles with MeSH descriptors outside the grant during the cycle (3 to 10 one by one, 11-15, 16-20, 21-30, 31-50, more than 50). Each column is a separate regression; the three columns of a measure use the same cycles, those of Table 2 for which the distance of the team and the measure are defined. The first column of each measure includes the measure as a continuous variable (extreme tail outcomes and disruption in units of 10 percentile points, pivoting and standing out as shares of MeSH descriptor pairs). The second adds an indicator for a distant team. The third adds the indicator's interaction with the measure, with the measure centered at its mean on these cycles, so that the coefficient on the indicator is the gap in renewal between distant and close teams at the measure's mean; the coefficient on the measure is its association with renewal for close teams. A distant team is one whose distance (Figure J1) lies above the median of the cycles reviewed by the same study section, or above the overall median for study sections with fewer than 20 such cycles. Under the measure and the interaction, in brackets, the coefficient times the standard deviation of the measure on these cycles.

Standard errors in parentheses, clustered by investigator and by grant. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Appendix K: Risk Taking and the Structure of NIH

Our estimates pool the renewal decisions of NIH's institutes over four decades. NIH is not a single funder, however, and the money its institutes could spend varied a great deal over the period. A skeptic could object that the penalty we document belongs to some institutes rather than to NIH peer review as a whole, or to the lean years that followed the doubling of NIH's budget, when a renewal became much harder to obtain.[1] This appendix describes how NIH reviews and funds a competing renewal, then examines the penalty by institute, over time, and against the scarcity of money at the institute that decided the renewal.

Institutes and study sections. NIH comprises 27 institutes and centers, 24 of which receive their own appropriations from Congress and award grants (Sekar and Angert 2026). Most institutes are organized around a disease or an organ system, such as the National Cancer Institute or the National Heart, Lung, and Blood Institute; some around a stage of life, such as the National Institute on Aging; and the National Institute of General Medical Sciences supports research tied to no particular disease. The institutes fund research, but for the most part they do not evaluate it. Most applications for an R01 grant are reviewed by study sections, standing committees of scientists convened by NIH's Center for Scientific Review and organized around fields of science rather than diseases. NIH staff assign each application to a study section according to its science, and to the institute whose mission it fits. An applicant may request a study section but does not choose it.

A field of science serves several diseases, so the applications reviewed by one study section are funded by several institutes. Of the 439 study sections that reviewed the grants we study, 87 percent had the renewals they reviewed decided by two institutes or more. The pattern holds within a year: 91 percent of the cycles were reviewed by a study section whose renewals that year went to at least two institutes, and the median institute drew on 15 study sections in a given year. Study sections change slowly. The Center for Scientific Review redrew their boundaries between 2001 and 2005, and investigators tend to stay with the study section that knows their work: in a typical five-year period, close to nine NIH grantees in ten are reviewed by a single study section (Azoulay et al. 2019b). Our regressions include a fixed effect for each combination of institute and study section, and our measure of standing out compares a cycle's publications with the funded applications of its institute and study section in the same year.

How a renewal is reviewed and funded. An application to renew a grant competes for funds as a new one does. It reports the progress made during the cycle that is ending and proposes the aims of the next. Since October 1997, reviewers have judged applications on five criteria: significance, approach, innovation, the investigator, and the environment (National Institutes of Health 1997). The members of the study section score the applications they discuss, the weakest applications go unscored, and the scores are converted into percentiles among the applications the study section reviewed. Reviewers often disagree about the same application. In an exercise that replicated NIH's review, 43 reviewers evaluating a common set of 25 applications showed no agreement on their quality, in their scores or in their written critiques (Pier et al. 2018); in an experiment at the National Science Foundation, whether a proposal was funded depended in a large proportion of cases on which reviewers happened to be selected for it (Cole et al. 1981). The funding decision is made separately, by the institute to which the application is assigned. Each institute funds applications in order of their percentile until its budget, net of its commitments to grants already under way, is spent. The last percentile funded is the payline, and the institute's staff and advisory council may fund some applications out of order (Azoulay et al. 2019b). How scarce money is for a renewal therefore depends on the appropriation that Congress votes for the funding institute that year, net of the commitments the institute has already made to grants under way. Two renewals that a study section judged alike in the same year can meet different fates, depending on the institute that funds them; Table K1 uses this variation.

What changed between 1980 and 2022. The conditions under which these decisions were made changed considerably over the period of our sample. NIH's budget doubled in nominal terms between 1998 and 2003, from $13.6 billion to $27.1 billion. It then stayed nearly flat for a decade, so that by 2013 its purchasing power was 22 percent below its 2003 level (Johnson 2013). The Recovery Act of 2009 added $10.4 billion, to be spent

[1] We are grateful to an anonymous referee for asking whether the penalty varies with how competitive funding is.

by September 2010 (National Institutes of Health 2009a). The renewal rate of the grants we study did not rise during the doubling, but it fell after it. Among all the cycles of the grants we study, it was between 40 and 52 percent in every year from 1982 to 2003, between 35 and 42 percent from 2004 to 2008, and between 23 and 28 percent from 2009 to 2019.

Peer review changed as well. For applications reviewed from 2009 onward, NIH replaced its scale of 1.0 to 5.0 with a scale of 1 to 9, and asked reviewers to score each criterion as well as the application as a whole (National Institutes of Health 2008a). An application submitted from 2009 onward that was not funded could be resubmitted only once (National Institutes of Health 2008b), and from 2010 the research plan of an R01 application, the renewal's progress report included, was limited to twelve pages (National Institutes of Health 2009b). These changes coincide with one another and with the tightening of funds, and our design cannot separate them. We show in Figure K2 how the penalty evolved; we do not attribute that evolution to any one of these changes.

The penalty by institute. We estimate the penalty separately for each of the eleven largest institutes, in one regression per measure with the controls of Table 2 (Figure K1). All 44 coefficients are negative. They are statistically significant at the 5 percent level in 10 of the 11 institutes for extreme tail outcomes, in all 11 for disruption, in 7 for pivoting, and in 9 for standing out. The size of the penalty differs across institutes. For pivoting, it ranges from -0.03 at the National Eye Institute to -0.19 at the National Institute of General Medical Sciences; for standing out, from -0.07 at the National Institute of Allergy and Infectious Diseases to -0.50 at the National Institute on Aging.

The penalty over time. We estimate the penalty separately for each evaluation year from 1982 to 2019, in one regression per measure with the controls of Table 2 held in common, as we do for institutes (Figure K2). The penalty is present throughout the period. The coefficient is negative in all 38 years for extreme tail outcomes and disruption, in 36 for pivoting, and in 34 for standing out, and no year has a positive coefficient that is statistically significant. Each year rests on 700 to 4,000 cycles, so the yearly estimates are noisy, especially for pivoting and standing out: their 95 percent intervals exclude zero in 15 years for pivoting and 19 for standing out, against 34 years for extreme tail outcomes and 32 for disruption.

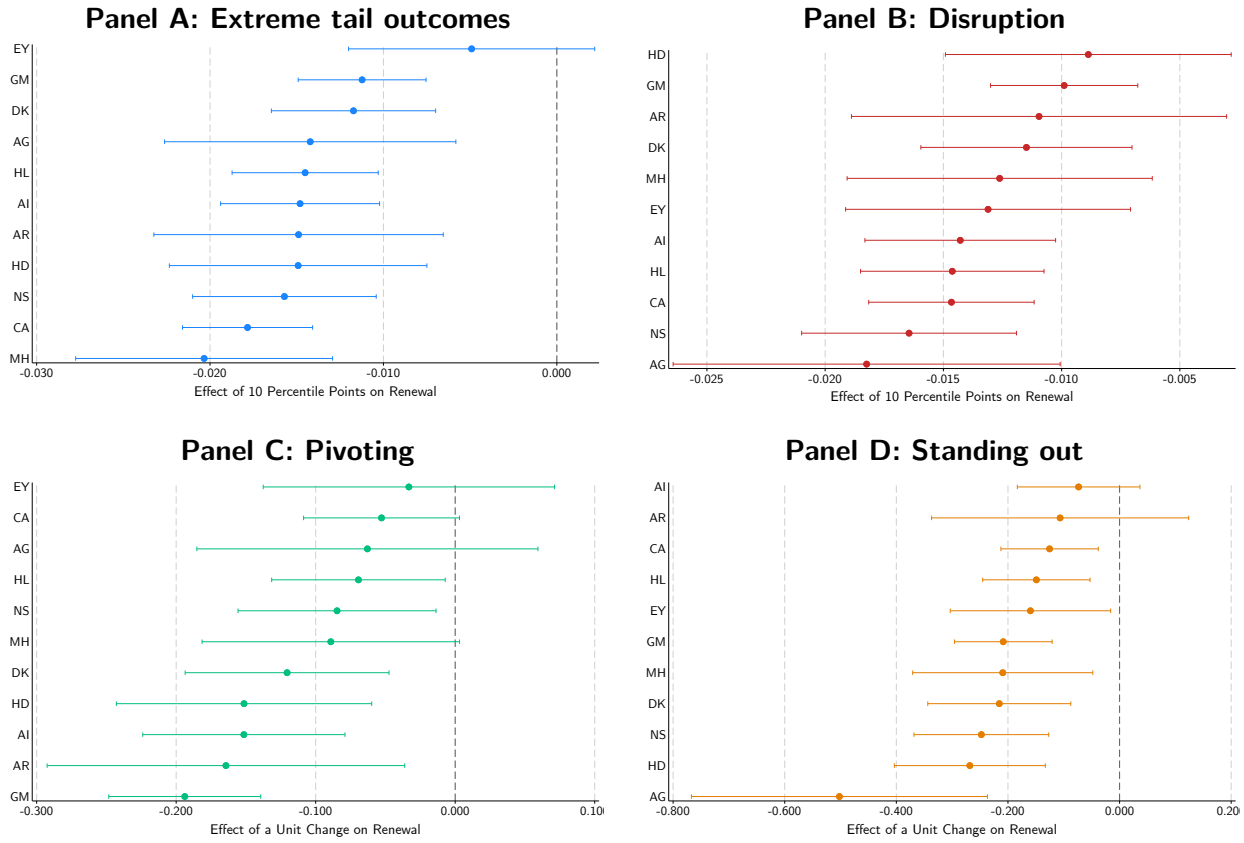
The size of the penalty changes over time, and differently for each measure. Pooling the years into five-year blocks in the same regression makes the changes easier to see. The penalty on extreme tail outcomes is largest in the early 1980s (-0.023 per 10 percentile points in 1981-85) and lies between -0.010 and -0.016 from 2001 onward. The penalty on disruption lies between -0.009 and -0.014 until 2010, then reaches -0.017 in 2011-15 and -0.024 in 2016-22. The penalty on pivoting is larger after 2001 (-0.13 to -0.16) than before (-0.06 to -0.10), but each block is estimated with a standard error of about 0.03, and the difference is within the noise. Standing out shows no trend.

The scarcity of money. We ask whether the penalty is larger where money is scarcer (Table K1). We measure scarcity by the renewal rate among the other cycles of the same institute and evaluation year: a higher rate means that money was more plentiful. For each measure, the first column repeats the regression of Table 2 with the measure entered as a continuous variable, and the second adds the renewal rate and its interaction with the measure. As everywhere in the paper, the regressions include fixed effects for each evaluation year and for each combination of institute and study section.

Where money is more plentiful, renewal is more likely: one standard deviation of the renewal rate, about 13 percentage points, is associated with a probability of renewal higher by 3.4 to 3.7 points. The penalty itself barely changes when the rate is added, and it varies little with the rate. The interaction is positive for extreme tail outcomes and disruption, which means a smaller penalty where money is plentiful, and negative for pivoting; it is statistically significant at the 5 percent level for disruption only, at the 10 percent level for pivoting, and not significant for extreme tail outcomes and standing out. One standard deviation of the rate changes the effect of a one-standard-deviation increase in the measure by 0.25 percentage points of renewal for extreme tail outcomes, 0.75 for disruption, 0.28 for pivoting, and 0.15 for standing out, against effects of 3.2, 3.1, 1.5, and 1.9 points. We therefore do not claim that peer review penalizes risk taking more when money is tight: the claim would hold, weakly, for some of our measures only.

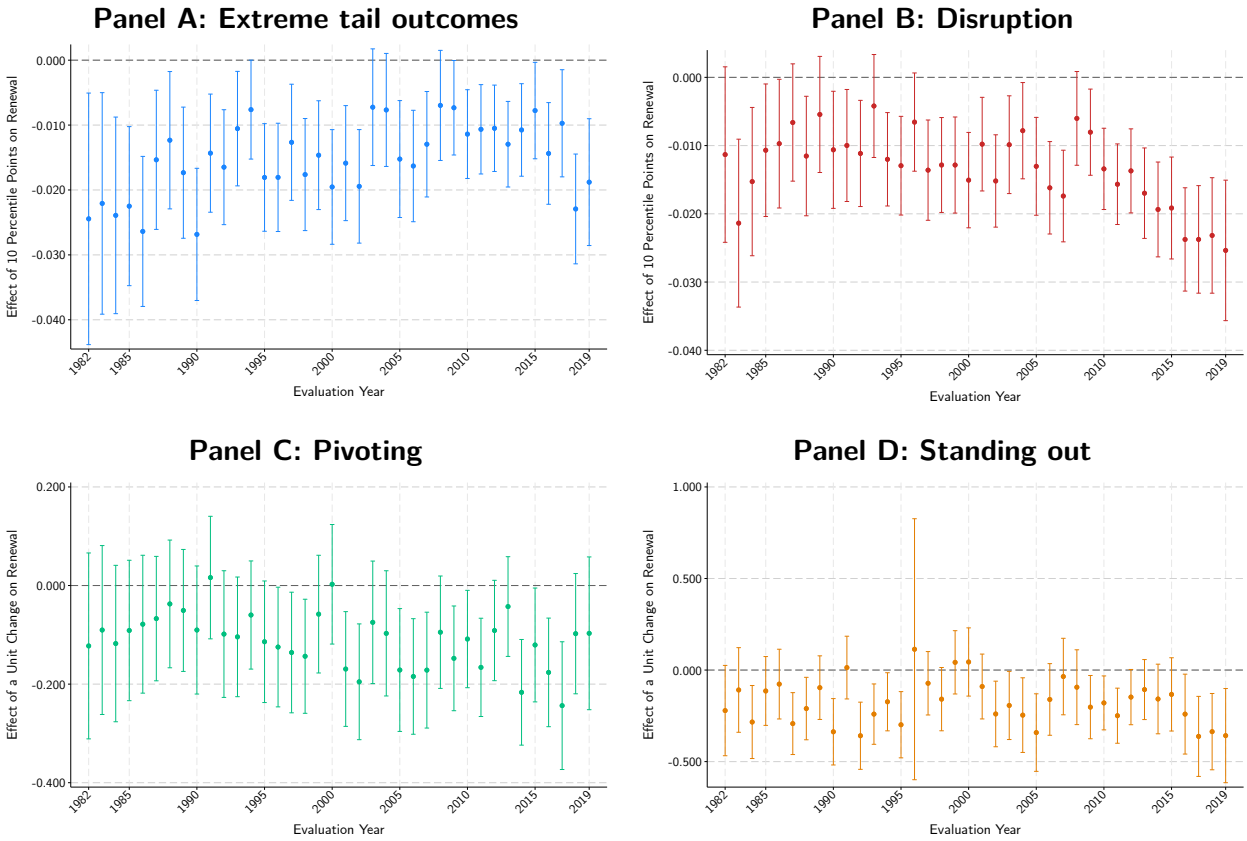
Interpretation. The penalty on risk taking is a feature of NIH peer review across its institutes and across four decades, under budgets that doubled and then lost a fifth of their purchasing power, and under two scales of scoring. Its size varies across institutes and over time, sometimes considerably: the penalty on disruption has roughly doubled since 2010. But the variation differs from one measure to the next, and it does not follow the scarcity of money in a consistent direction. Whether the penalty has grown over time depends on the measure as well: it has for disruption, and not for extreme tail outcomes or standing out, while the estimates for pivoting are too imprecise to settle the question.

Figure K1: The Penalty for Risk Taking by NIH Institute



Notes: Coefficients, with 95% confidence intervals, of each measure interacted with an indicator for each of the eleven largest NIH institutes (GM: general medical sciences; CA: cancer; HL: heart, lung and blood; AI: allergy and infectious diseases; NS: neurological disorders and stroke; DK: diabetes, digestive and kidney diseases; HD: child health; EY: eye; MH: mental health; AG: aging; AR: arthritis and musculoskeletal and skin diseases), in one linear probability model of renewal per measure, with the controls of Table 2 and no common slope; cycles of other institutes are excluded. Extreme tail outcomes and disruption in units of 10 percentile points, pivoting and standing out as shares of MeSH descriptor pairs. Institutes are sorted by their coefficient. Standard errors are clustered by investigator and by grant.

Figure K2: The Penalty for Risk Taking over Time



Notes: Coefficients, with 95% confidence intervals, of each measure interacted with an indicator for each evaluation year, in one linear probability model of renewal per measure, with the controls of Table 2 and no common slope. Evaluation years 1980 and 1981 are pooled with 1982, as in the fixed effects, and 2020 to 2022 (164 cycles) with 2019. Extreme tail outcomes and disruption in units of 10 percentile points, pivoting and standing out as shares of MeSH descriptor pairs. In 1996, standing out is defined for only 280 cycles, against 930 to 3,128 in other years: the abstracts of most applications funded in fiscal year 1996 are missing from NIH's RePORTER files, so few comparison pools reach the 20 tagged applications the measure requires. Standard errors are clustered by investigator and by grant.

Table K1: The Penalty and the Scarcity of Money at the Institute

	Extreme tail outcomes		Disruption		Pivoting		Standing out	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Risk measure	-0.0142**	-0.0142**	-0.0129**	-0.0133**	-0.1106**	-0.1080**	-0.1751**	-0.1759**
	(0.0009)	(0.0009)	(0.0008)	(0.0008)	(0.0115)	(0.0114)	(0.0182)	(0.0181)
	[-0.0322]	[-0.0322]	[-0.0304]	[-0.0314]	[-0.0157]	[-0.0153]	[-0.0187]	[-0.0188]
Renewal rate of the institute-year		0.0368**		0.0341**		0.0346**		0.0342**
		(0.0050)		(0.0046)		(0.0045)		(0.0050)
Risk measure × renewal rate of the institute-year		0.0011		0.0032**		-0.0195†		0.0138
		(0.0007)		(0.0006)		(0.0103)		(0.0168)
		[0.0025]		[0.0075]		[-0.0028]		[0.0015]
Mean of dependent variable	0.446	0.446	0.424	0.424	0.422	0.422	0.427	0.427
SD of the institute-year renewal rate	0.130	0.130	0.130	0.130	0.129	0.129	0.129	0.129
Nb. of investigators	34,659	34,659	38,634	38,634	39,184	39,184	35,278	35,278
Nb. of grant cycles	87,365	87,365	99,089	99,089	101,019	101,019	82,785	82,785
Adjusted R^2	0.1651	0.1656	0.1716	0.1723	0.1679	0.1684	0.1667	0.1671

Notes: Estimates stem from a linear probability model in which the unit of analysis is the grant cycle and the dependent variable is an indicator for whether the grant cycle was renewed, with the controls of Table 2. Each column is a separate regression; both columns of a measure use the same cycles, those whose institute and evaluation year count at least 10 cycles. The first column of each measure includes the measure as a continuous variable (extreme tail outcomes and disruption in units of 10 percentile points, pivoting and standing out as shares of MeSH descriptor pairs). The second adds the renewal rate among the other cycles of the same institute and evaluation year, standardized to mean 0 and standard deviation 1 on these cycles (its standard deviation before standardizing is reported), and its interaction with the measure. In the interaction, the measure is centered at its mean on these cycles, so that the coefficient on the renewal rate is its association with renewal for a cycle of average risk. A higher rate means less competitive funding, so a positive interaction means a smaller penalty when money is plentiful. Under the measure and the interaction, in brackets, the coefficient times the standard deviation of the measure on these cycles: the effect of a one-standard-deviation increase in the measure, and the change in that effect for each standard deviation of the renewal rate.

Standard errors in parentheses, clustered by investigator and by grant. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Appendix L: Measuring the Downstream Outcomes

In Table 3, cycles are compared on six outcomes recorded in the five years after the cycle's last payment; descriptive statistics for these outcomes are presented in Table 1e. Each outcome rests on a source built for purposes other than ours: NIH's records of the publications that acknowledge its grants and of the patents that its grantees report; a database of the citations from patents to scientific articles; lists of the patents on drugs and medical devices; estimates of what patents are worth to the firms that own them; and a model that predicts which articles patents will cite. We describe each outcome, how it is matched to a grant cycle, and the limitations to bear in mind when interpreting the results of Table 3.

The publications that follow the cycle. Five of the six outcomes are built from the same publications: those that acknowledge the grant, list the investigator among their authors, and appeared in the five years after the year of the cycle's last payment. The acknowledgments come from NIH RePORTER and the authorship from Author-ity, as for the measures of risk taking (Section 3 and Appendix A); only the window differs. The measures of risk taking for a cycle are built from its own publications, from its first to its last funded year, and its outcomes from those of the five years that follow, so that no publication enters both a measure of risk taking and an outcome for the same cycle. A publication of those five years counts whether it appeared during a later cycle of the grant or after the grant ended. For a renewed cycle, the five years therefore include work funded by the renewal, whose publications also enter the measures of risk taking for the next cycle; for a terminated cycle, they follow the end of the grant's funding. The cycles of Table 1e are followed by 561,989 distinct publications. Since the five years after one cycle of a grant can overlap those after the next, they form 667,266 pairs of a cycle and a publication. Nearly every renewed cycle is followed by at least one such publication (98.4 percent), and most terminated cycles are as well (87.8 percent).

In Table L1, the outcomes of these five years are set beside the same outcomes computed on the publications of the cycle itself and on the patents filed during it. Terminated cycles are followed by fewer publications than they produced (4.2 against 5.9 on average), and renewed cycles by more (10.1 against 8.0). For terminated cycles, the decline is steeper for the three outcomes built from citing patents than for publications and their citations. Taken together, the publications that follow a terminated cycle receive 5.57 citations from patents on average, against 13.59 for the publications of the cycle itself, while those that follow a renewed cycle receive about as many as the cycle's own (24.05 against 25.67). The comparisons of Table 3 are therefore made among renewed cycles or among terminated ones, never across the two.

Article citations. Article citations are the sum, over these publications, of the citations each receives from other *PubMed* publications from the year it appeared through the fifth year after, as recorded in OpenAlex, an open index of scholarly publications and of the citations among them. The citation data end in 2025, so the citations of a publication of 2021 are counted through the fourth year after it appeared, and those of a publication of 2025 in that year only. Such publications are few: 2.1 percent of the pairs of a renewed cycle and a publication, and 3.6 percent of those of a terminated cycle, involve a publication that appeared after 2020.

Patent citations. Patent citations are the citations that these publications receive from US patents, which we take from Reliance on Science (Marx and Fuegi 2020, 2022), in its release of July 2026: the citations from granted US utility patents, issued through 2025, to articles indexed in *PubMed*. We count every citation in the release, however long after the publication the citing patent was granted. The database links patents to the scientific articles they cite, both on the front page of the patent, where the prior art is listed, and in the body of its text, from which Marx and Fuegi (2022) extract references with a combination of rules and machine learning. The two kinds of citation serve different purposes. Citations on the front page "serve a *legal* purpose by delimiting the validity of the patent" and are reviewed by patent attorneys, while those in the text "tend to be supplied by the inventors themselves" (Marx and Fuegi 2022, p. 370). We use both kinds.

Each link carries a score, from 1 to 10, of the confidence that the reference was matched to the right article, and we keep the links with a score of 6 or more. For the citations in the text, Marx and Fuegi (2022, Table 1) estimate that 99.5 percent of the links at or above that score are correct, against 99.4 percent at or above the score of 4 that they recommend, and that the higher threshold finds about 5 percent fewer of the true citations.

The release contains no link with a score below 4, and on our publications the links scored 4 or 5, which we drop, are 2.1 percent of all links; 88 percent have the highest score. Of the links we keep, 53 percent appear only on the front page, 28 percent only in the text, and 19 percent in both. A cycle's count is the number of pairs of a publication and a citing patent, so that a patent citing two of the publications that follow the cycle counts twice. On the cycles of Table 1e, these publications are cited by 259,912 distinct patents.

A citation tells us that the inventors, their attorneys, or the examiner knew of the publication, not that the invention drew on it, and front-page citations, chosen for their legal purpose, are weaker evidence of such use than citations in the text (Marx and Fuegi 2022, p. 370).

Patents cite publications long after they appear. For the publications of 1995 to 2000 that follow our cycles, the median citing patent was granted 16 years after the publication and filed 12 years after it; only 5.8 percent of their citations come from patents granted within five years, and 22 percent from patents granted within ten. Every cycle's count is therefore incomplete, and more so the later the cycle: the publications that follow a cycle evaluated in 2020 appeared from 2021 to 2025, and the release ends with the patents granted in 2025. The regressions of Table 3 include evaluation-year effects, so a shortfall that scales down the counts of all the cycles evaluated in the same year by the same proportion leaves their coefficients unchanged. A difference in the timing of citations between cycles above and below the median of a measure of risk taking would not be absorbed.

FDA patent citations. FDA patent citations are those, among these citations, that come from patents on drugs and medical devices. No public source lists all such patents, so we assembled a list from five sources. We call it the Russet Book: the FDA lists the patents on small-molecule drugs in its Orange Book and some of those on biologics in its Purple Book, and russet is the color that painters obtain by mixing orange and purple.

The first two sources are two versions of the Orange Book. Since the Hatch-Waxman Act of 1984, a firm that markets a brand-name drug approved under a New Drug Application must list with the FDA every patent that claims the drug's active ingredient, its formulation, or a method of using it for an approved indication; patents on manufacturing processes cannot be listed (Durvasula et al. 2023, p. 4). The listing matters because a generic manufacturer that seeks to enter before the patents expire must challenge each listed patent. Each edition of the Orange Book lists the patents in force at the time, so patents that have expired or been withdrawn disappear from later editions. Durvasula et al. (2023) digitized every edition from 1985 to 2016, recovering 5,511 patents, and the FDA's current data files, of March 2024, list 5,545. Together, the two contain 9,694 distinct patents. The third source, a list of the patents on 2,832 New Drug Applications compiled by Bhaven Sampat, adds 418.

Biologics, such as therapeutic proteins, are approved under a different application, the Biologics License Application, and are therefore largely absent from the Orange Book (Durvasula et al. 2023, p. 3). Biologics account for about 2 percent of US prescriptions but close to half of net spending on prescription drugs (Frakes and Wasserman 2026, p. 1315), and their share of the new molecular entities approved by the FDA rose from 6.3 percent in 1985-1990 to 26 percent in 2015-2020 (Durvasula et al. 2023, p. 3, fn. 14). Their patents are also much harder to identify. No law requires the maker of a biologic to list all of its patents. The Purple Book has published patent lists for biologics only since 2021, and only those that the maker discloses in an exchange with a biosimilar applicant that is itself elective and limited to the patents that this applicant's product might infringe; as a result, it contains fewer than 4 percent of the patents associated with approved biologics (Frakes and Wasserman 2026, pp. 1338-1339). A biologic is associated with about 23 patents on average, against fewer than 4 for a small-molecule drug, and a quarter of them cover its manufacture (Frakes and Wasserman 2025, pp. 5 and 16). Frakes and Wasserman (2026, pp. 1342 and 1349) built a list of these patents for the 515 therapeutic biologics approved by January 2025, reading and classifying each candidate patent, 11,595 pairs of a drug and a patent in all. They plan to make it public through the NBER, and they shared with us the patents it contains, without the drug that each covers or its type: 4,821 issued patents and 1,162 applications still pending. The fourth source is the issued patents; a pending application cannot be matched to Reliance on Science, which covers granted patents only. Of these 4,821 patents, 86 are in the Orange Book and 17 more in Sampat's list, so the fourth source adds 4,718. It also includes the patents on a biologic's manufacture, which the Orange Book excludes, and the list we received does not identify them.

The fifth source is the list of patent term extensions. Under 35 U.S.C. 156, the owner of an approved product may extend the term of one of its patents to recover part of the time lost to clinical trials and regulatory review (Durvasula et al. 2023, p. 5). The extension is available for human drugs, biologics among them, and for medical

devices, but only for the first approval of an active ingredient (US Patent and Trademark Office 2026a, Poon and You 2022), and nearly all the extended patents that are not in the Orange Book cover biologics or devices (Durvasula et al. 2023, p. 5, fn. 29). The USPTO's list of the patents whose term it had extended by September 2026 (US Patent and Trademark Office 2026a) holds 1,140 patents, of which 438 are in neither the Orange Book nor Sampat's list. Frakes and Wasserman list 145 of these 438, among them 94 of the 99 that another list identifies as patents on biologics: a list, assembled by Bhaven Sampat, of the new drugs and new biologics that the FDA approved from 1985 to 2023 (US Food and Drug Administration 2026), each matched to its extended patent if it has one. The term extensions add 293 patents to the other four sources.

The Russet Book thus holds 15,123 patents, of which 210, reissued or design patents, cannot be matched to the utility patents of Reliance on Science. On the cycles of Table 1e, the publications of the five years after the cycle receive 35,608 citations from these patents: 45.5 percent from the Orange Book, 3.5 percent from the patents that only Sampat's list contains, 50.5 percent from the patents on biologics that only Frakes and Wasserman list, and 0.5 percent from the term extensions alone. Patents on biologics give half of these citations, and all of them for 2,259 of the 4,602 cycles whose publications receive at least one (49.1 percent). The list remains incomplete for devices, of which it contains only the patents whose term was extended, one per product, so the outcome undercounts the citations from patents on devices.

KPSS patent value. KPSS patent value is the value of the patents that cite these publications, from the estimates of Kogan et al. (2017). They infer the value of a patent to the firm that owns it from the firm's stock return, net of the market's, over the three days that begin on the day the patent is granted. Since the stock price changes over those days for many reasons besides the patent, they extract the part of the return attributable to the patent under assumptions about the distributions of the patent's value and of other news, and divide the value among the patents that the firm receives on the same day (Kogan et al. 2017, pp. 673-678). We use their estimates for the patents granted through 2023, deflated by the consumer price index of the month in which each patent was granted and expressed in millions of 2023 dollars (at the price level of December 2023). A cycle's value is the sum over the distinct patents that cite the publications of the five years after the cycle, so that a patent citing several of them counts once.

The measure has four limitations. First, it exists only for patents owned by firms whose stock is publicly traded. Patents held by universities, hospitals, and privately held firms, start-ups among them, count as zero. Of the 259,912 patents that cite the publications following our cycles, 21.6 percent have a value (23.7 percent of those granted by 2023), and the share has fallen, from 30.9 percent of those granted from 1995 to 2009 to 19.3 percent of those granted from 2020 to 2023. Only 34.1 percent of renewed cycles and 12.8 percent of terminated ones are followed by a publication that a patent with a value cites, so the median cycle has a value of zero (Table 1e). Second, the estimate measures the private value of the patent to its owner, as investors assessed it when the patent was granted, not its value to patients or to the other firms and scientists who build on it; Kogan et al. (2017, p. 666) note that this private value need not coincide with the scientific value of the patent, which forward citations typically measure. Third, a citing patent's value is that of the whole invention, to which the cited publication may have contributed little, and a patent that cites the publications of several cycles counts in full for each of them. The sum is therefore the private value of the patented technology that cites the work, not an estimate of the value that the work created. Fourth, the estimate is noisy: Kogan et al. (2017, pp. 677-678, fn. 9) caution that their estimates contain measurement error, which aggregating over patents only partly removes.

Commercial potential. Commercial potential requires no patent to have cited the publications. Masclans et al. (2025) train a language model on the abstracts of scientific articles to predict whether an article will be cited by a patent that its owner later keeps in force by paying the first maintenance fee (a "renewed" patent, in their terms); since an owner pays only for a patent worth keeping, maintained patents have some commercial value. They estimate a separate model for each year, on articles published several years earlier and the citations they received, so that an article's own citations never enter its score. We use the scores that the authors computed for *PubMed* publications from 1990 onward, and a cycle's commercial potential is the mean score of the publications of the five years after it that have a score. Of the pairs of a cycle and a publication, 7.4 percent involve a publication that appeared before 1990, which has no score, and 94.2 percent of the others have one; the measure is missing for the cycles with no scored publication in the five years after.

The scores predict what kind of work a publication reports, not what became of it: a high score means that the abstract resembles those of articles that maintained patents have cited. That is a limitation, but also an advantage over the counts of citations: the score requires no citation, so it is not truncated for the latest cycles. The scores also rank publications rather than estimate probabilities. In the version that Masclans et al. (2025) describe, each year's model is trained on equal numbers of articles cited and not cited by maintained patents, whereas maintained patents cite only about one article in twenty in their data. A mean score of 0.56, as in Table 1e, therefore does not mean that 56 percent of the publications will be cited by such a patent; only the ranking is informative. Since each year has its own model, the ranking holds among the publications of the same year, not across years. The evaluation-year effects of Table 3 absorb differences in the average level of the scores across years, although not differences in their spread. Finally, Masclans et al. (2025) note that their measure reflects findings with relatively near-term commercial use, not the long-run potential of early-stage research, which makes it an incomplete measure of the value of much of the science that NIH funds.

Direct grant patents. Direct grant patents are not built from publications. The Bayh-Dole Act of 1980 requires the universities, nonprofit organizations, and small businesses that receive federal research funds to disclose to the funding agency each invention made with its support, and to state in any patent that issues on it that the invention was made with government support and that the government has certain rights in it (35 U.S.C. 202(c)). NIH's grantees report their inventions and patents through iEdison, an interagency reporting system, and NIH RePORTER lists, for each issued patent reported to NIH, the projects reported as having supported it; a project is a grant, identified by its activity code, institute, and serial number. We use the file of these links that NIH ExPORTER, RePORTER's service for bulk downloads, posts (National Institutes of Health 2026a), in its version of 11 October 2026: 93,357 records linking 50,398 patents to 32,376 projects, of which 33,765 patents are linked to R01-equivalent grants. A link names the grant but no year of it: patents "cannot be identified with any particular year of the project or fiscal year of funding" (National Institutes of Health 2026d). We therefore date each patent by the year its application was filed, taken from PatentsView, the USPTO's public database of patent records (US Patent and Trademark Office 2026b), and count, for each cycle, the patents linked to its grant whose applications were filed in the five years after the cycle's last payment. Such a patent may draw on the work of the cycle, of an earlier one, or, for a renewed cycle, of the renewal. The filing date is missing for 1,196 of the 44,607 links between a patent and an R01-equivalent grant (2.7 percent), and these links are dropped. Nine in ten of the linked patents were filed in 1995 or later, and their median filing year is 2009. Among the cycles of Table 1e, 10.8 percent of renewed cycles and 6.0 percent of terminated ones are followed by at least one such patent, but only 3.0 percent of the cycles evaluated in the 1980s, against 8 to 10 percent of those evaluated in each later decade; the evaluation-year effects of Table 3 absorb this difference across years.

These links are only as complete as grantees' compliance with the law. Rai and Sampat (2012) find that many academic biomedical patents arising from federally funded research omit the statement of government support that the law requires, and that even among the patents that grantees reported to NIH, fewer than 90 percent included it as late as 2007. NIH itself warns that the patent information in RePORTER is incomplete, since "not all recipients of NIH funding are compliant with the iEdison reporting requirements, particularly after their NIH support has ended" (National Institutes of Health 2026d). This last point bears on our comparisons. The grant of a terminated cycle has lost its NIH support, so its patents may be reported less completely than those of a renewed grant, which is one more reason why, in Table 3, terminated cycles are compared only with one another. Within that comparison, incomplete reporting matters only insofar as it differs between cycles above and below the median of a measure of risk taking. RePORTER also lists only issued patents, so patents filed in the last few years, many of them not yet issued, are missing for the latest cycles; the evaluation-year effects absorb this shortfall insofar as it scales down the counts of all the cycles of a year by the same proportion.

Interpretation. Five of the outcomes are built from the publications that follow the cycle, and the sixth from the patents that grantees report to NIH. Each has gaps. Counts of citations from patents are incomplete, the more so the later the cycle, because patents cite publications many years after they appear. Patent values exist only for the patents of publicly traded firms, so most cycles have a value of zero. The list of drug and device patents covers devices only in part. Grantees also report their patents incompletely, particularly once their NIH support has ended, as NIH itself warns. The regressions of Table 3 include evaluation-year effects, which absorb what such gaps do to all the cycles of a year alike.

Table L1: Downstream Outcomes during the Grant Cycle and in the Five Years after Its Last Payment

	Renewed	Terminated
Cycles	42,127	57,597
A. During the cycle		
Publications	8.0	5.9
Article Citations	239.9	158.2
Direct Grant Patents	0.16	0.10
Patent Citations	25.67	13.59
FDA Patent Citations	0.67	0.37
KPSS Patent Value	414.6	204.6
Commercial Potential	0.547	0.551
Cycles with a scored publication	34,758	50,404
B. The five years after the cycle's last payment		
Publications	10.1	4.2
Article Citations	280.4	93.5
Direct Grant Patents	0.22	0.10
Patent Citations	24.05	5.57
FDA Patent Citations	0.63	0.16
KPSS Patent Value	359.8	77.8
Commercial Potential	0.562	0.554
Cycles with a scored publication	39,083	46,062

Notes: Means over the cycles of Table 1e: cycles evaluated by 2020 with at least one publication acknowledging the grant and a known number within the grant, renewed or terminated (not renewed, and the grant not revived later). Every outcome but direct grant patents is built from the publications that acknowledge the grant and list the investigator among their authors. Panel A: the publications that appeared from the cycle's first to its last funded year, those from which the measures of risk taking are built, and the patents linked to the grant whose applications were filed in those years. Panel B: the publications of the five years after the year of the cycle's last payment and the patents filed in those years, the outcomes of Table 3, whose means are those of Table 1e. For renewed cycles, these five years include work funded by the renewal. Article citations: five-year citations from *PubMed* publications. Direct grant patents: the patents that NIH RePORTER links to the grant. Patent citations: citations from US patents (Reliance on Science, match confidence 6 or more, front-page and in-text). FDA patent citations: those from drug- and device-related patents. KPSS patent value: the value of the distinct citing patents, from Kogan et al. (2017), in millions of 2023 dollars. Commercial potential: the mean score of the publications, from Masclans et al. (2025), over the publications with a score (from 1990); the line below gives the number of cycles with at least one.

Appendix References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J. Molecular Biol.* 215(3):403–410.

Arts S, Fleming L (2018) Paradise of novelty—or loss of human capital? Exploring new fields and inventive output. *Organ. Sci.* 29(6):1074–1092.

Arts S, Melluso N, Veugelers R (2025) Beyond citations: Measuring novel scientific ideas and their impact in publication text. *Rev. Econom. Statist.*, ePub ahead of print January 28, https://doi.org/10.1162/rest_a_01561.

Azoulay P, Fons-Rosen C, Graff Zivin JS (2019a) Does science advance one funeral at a time? *Amer. Econom. Rev.* 109(8):2889–2920.

Azoulay P, Graff Zivin JS, Li D, Sampat BN (2019b) Public R&D investments and private-sector patenting: Evidence from NIH funding rules. *Rev. Econom. Stud.* 86(1):117–152.

Banal-Estañol A, Macho-Stadler I, Pérez-Castrillo D (2019) Evaluation in research funding agencies: Are structurally diverse teams biased against? *Res. Policy* 48(7):1823–1840.

Berlin I (1953) *The Hedgehog and the Fox: An Essay on Tolstoy's View of History* (Weidenfeld & Nicolson, London).

Bornmann L, Devarakonda S, Tekles A, Chacko G (2020) Are disruption index indicators convergently valid? The comparison of several indicator variants with assessments by peers. *Quant. Sci. Stud.* 1(3):1242–1259.

Boudreau KJ, Guinan EC, Lakhani KR, Riedl C (2016) Looking across and looking beyond the knowledge frontier: Intellectual distance, novelty, and resource allocation in science. *Management Sci.* 62(10):2765–2783.

Bromham L, Dinnage R, Hua X (2016) Interdisciplinary research has consistently lower funding success. *Nature* 534(7609):684–687.

Bu Y, Waltman L, Huang Y (2021) A multidimensional framework for characterizing the citation impact of scientific publications. *Quant. Sci. Stud.* 2(1):155–183.

Cohan A, Feldman S, Beltagy I, Downey D, Weld DS (2020) SPECTER: Document-level representation learning using citation-informed transformers. *Proc. 58th Annual Meeting Assoc. Comput. Linguistics* (Association for Computational Linguistics, Stroudsburg, PA), 2270–2282.

Cole S, Cole JR, Simon GA (1981) Chance and consensus in peer review. *Science* 214(4523):881–886.

Durvasula M, Hemphill CS, Ouellette LL, Sampat B, Williams HL (2023) The NBER Orange Book dataset: A user's guide. *Res. Policy* 52(7):104791.

Frakes MD, Wasserman MF (2025) Strategic patenting: Evidence from the biopharmaceutical industry. NBER Working Paper 34024, National Bureau of Economic Research, Cambridge, MA.

Frakes MD, Wasserman MF (2026) Unveiling the patent landscape of biologic drugs. *Northwestern Univ. Law Rev.* 120(5):1315–1374.

Funk RJ, Owen-Smith J (2017) A dynamic network measure of technological change. *Management Sci.* 63(3):791–817.

Hill R, Yin Y, Stein C, Wang X, Wang D, Jones BF (2025) The pivot penalty in research. *Nature* 642(8069):999–1006.

Johnson JA (2013) Brief history of NIH funding: Fact sheet. CRS Report R43341, Congressional Research Service, Washington, DC.

Knecht L, Auld VA, McGhee MA (2001) Changes in the treatment of chemical data in MEDLINE citations. *NLM Tech. Bull.* 323:e7.

Kogan L, Papanikolaou D, Seru A, Stoffman N (2017) Technological innovation, resource allocation, and growth. *Quart. J. Econom.* 132(2):665–712.

Le D, Mork JG (2023) Check tags MeSH terms indexing research project. Report, Applied Clinical Informatics Branch, Lister Hill National Center for Biomedical Communications, National Library of Medicine, Bethesda, MD, https://lhncbc.nlm.nih.gov/LHC-publications/PDF/CheckTagsMeSHTermsResearchTechnicalReport.pdf.

Lerchenmueller MJ, Sorenson O (2016) Author disambiguation in PubMed: Evidence on the precision and recall of Author-ity among NIH-funded scientists. *PLoS One* 11(7):e0158731.

Lin D (1998) An information-theoretic definition of similarity. *Proc. 15th Internat. Conf. Machine Learn.* (Morgan Kaufmann, San Francisco), 296–304.

Lin J, Wilbur WJ (2007) PubMed related articles: A probabilistic topic-based model for content similarity. *BMC Bioinformatics* 8:423.

Lin Y, Li L, Wu L (2026) The disruption index measures displacement between a paper and its most cited reference. *Quant. Sci. Stud.* 7:229–239.

Manso G (2011) Motivating innovation. *J. Finance* 66(5):1823–1860.

Marx M, Fuegi A (2020) Reliance on science: Worldwide front-page patent citations to scientific articles. *Strategic Management J.* 41(9):1572–1594.

Marx M, Fuegi A (2022) Reliance on science by inventors: Hybrid extraction of in-text patent-to-article citations. *J. Econom. Management Strategy* 31(2):369–392.

Masclans R, Hasan S, Cohen WM (2025) Measuring the commercial potential of science. *Strategic Management J.* 46(9):2199–2236.

Mork J, Aronson A, Demner-Fushman D (2017) 12 years on—Is the NLM medical text indexer still useful and relevant? *J. Biomedical Semantics* 8:8.

National Institutes of Health (1997) Review criteria for and rating of unsolicited research grant and other applications. *NIH Guide for Grants and Contracts* 26(22), Notice NOT-97-010 (June 27), https://grants.nih.gov/grants/guide/notice-files/not97-010.html.

National Institutes of Health (2008a) Enhancing peer review: The NIH announces new scoring procedures for evaluation of research applications received for potential FY2010 funding. *NIH Guide for Grants and Contracts*, Notice NOT-OD-09-024 (December 2), https://grants.nih.gov/grants/guide/notice-files/NOT-OD-09-024.html.

National Institutes of Health (2008b) New NIH policy on resubmission (amended) applications. *NIH Guide for Grants and Contracts*, Notice NOT-OD-09-003 (October 8), https://grants.nih.gov/grants/guide/notice-files/NOT-OD-09-003.html.

National Institutes of Health (2009a) NIH's role in the American Recovery and Reinvestment Act (ARRA). From the Director (February 25), archived copy, https://webarchive.library.unt.edu/eot2008/20090505234535/http://www.nih.gov/about/director/02252009statement_arra.htm.

National Institutes of Health (2009b) Restructured application forms and instructions for submissions for FY2011 funding. *NIH Guide for Grants and Contracts*, Notice NOT-OD-09-149 (September 16), https://grants.nih.gov/grants/guide/notice-files/NOT-OD-09-149.html.

National Institutes of Health (2026a) ExPORTER: RePORTER patents information with project references for all fiscal years. File of October 11, https://reporter.nih.gov/exporter/patents.

National Institutes of Health (2026b) NIH Grants Policy Statement. Revised March 2026, https://grants.nih.gov/grants/policy/nihgps/HTML5/section_8/8.1.2_prior_approval_requirements.htm.

National Institutes of Health (2026c) NIH research project grants and R01-equivalent grants: Competing applications, awards, success rates, and total funding by type and submission number, fiscal year 2015–2025. NIH RePORT, Table 210 (data updated February 18), https://report.nih.gov/reportweb/web/displayreport?rId=571.

National Institutes of Health (2026d) RePORT frequently asked questions. Accessed October 16, https://report.nih.gov/faqs.

Oster E (2019) Unobservable selection and coefficient stability: Theory and evidence. *J. Bus. Econom. Statist.* 37(2):187–204.

Park M, Leahey E, Funk RJ (2023) Papers and patents are becoming less disruptive over time. *Nature* 613(7942):138–144.

Petersen AM, Arroyave FJ, Pammolli F (2025) The disruption index suffers from citation inflation: Re-analysis of temporal CD trend and relationship with team size reveal discrepancies. *J. Informetrics* 19(1):101605.

Pier EL, Brauer M, Filut A, Kaatz A, Raclaw J, Nathan MJ, Ford CE, Carnes M (2018) Low agreement among reviewers evaluating the same NIH grant applications. *Proc. Natl. Acad. Sci. USA* 115(12):2952–2957.

Poon MM, You Y (2022) Patent term extension for biologics. MoFo Life Sciences, Morrison Foerster, February 15, https://lifesciences.mofo.com/topics/patent-term-extension-for-biologics.

Rai AK, Sampat BN (2012) Accountability in patenting of federally funded research. *Nature Biotech.* 30(10):953–956.

Resnik P (1995) Using information content to evaluate semantic similarity in a taxonomy. *Proc. 14th Internat. Joint Conf. Artificial Intelligence*, volume 1 (Morgan Kaufmann, San Mateo, CA), 448–453.

Sekar K, Angert J (2026) National Institutes of Health (NIH) funding: FY1996–FY2026. CRS Report R43341, Congressional Research Service, Washington, DC, https://www.congress.gov/crs-product/R43341.

Smith BN, Singh M, Torvik VI (2013) A search engine approach to estimating temporal changes in gender orientation of first names. *Proc. 13th ACM/IEEE-CS Joint Conf. Digital Libraries* (Association for Computing Machinery, New York), 199–208.

Torvik VI, Smalheiser NR (2009) Author name disambiguation in MEDLINE. *ACM Trans. Knowledge Discovery Data* 3(3):11.

US Food and Drug Administration (2026) Compilation of CDER new molecular entity (NME) drug and new biologic approvals. Edition covering 1985 to 2023, used here; the edition current as of May 12, 2026 covers 1985 to 2025, https://www.fda.gov/drugs/drug-approvals-and-databases/compilation-cder-new-molecular-entity-nme-drug-and-new-biologic-approvals.

US Patent and Trademark Office (2026a) Patent terms extended under 35 USC 156. List of certificates issued through September 8, downloaded October 9, https://www.uspto.gov/patents/laws/patent-term-extension/patent-terms-extended-under-35-usc-156.

US Patent and Trademark Office (2026b) PatentsView: Data download tables. Table g_application, downloaded February 9, https://patentsview.org/download/data-download-tables.

Willis J (1999) Year-end processing. *NLM Tech. Bull.* 311:e2.

Wu L, Wang D, Evans JA (2019) Large teams develop and small teams disrupt science and technology. *Nature* 566(7744):378–382.

Wu Q, Yan Z (2019) Solo citations, duet citations, and prelude citations: New measures of the disruption of academic papers. Preprint, submitted May 9, https://doi.org/10.48550/arXiv.1905.03461.