Funding Breakthrough Research. An Update and Research Agenda

Chiara Franzoni and Reinhilde Veugelers

September 2025

Chiara Franzoni
School of Management
Polytechnic University of Milan
Piazza Leonardo da Vinci, 32
Milan, ITALY 20123
chiara.franzoni@polimi.it

Reinhilde Veugelers
Department of Management, Strategy & Innovation, KU Leuven
Naamsestraat 69, 3000 Leuven, BELGIUM
Bruegel
Peterson Institute for International Economics
reinhilde.veugelers@kuleuven.be

1. Introduction

Scientific progress and technological change are central to solving human problems and fostering economic growth and social well-being. Yet, the emergence of breakthroughs has failed to keep pace with the large increase of scientific productivity and total investments over time (Bloom et al., 2020; Park et al., 2023). Breakthrough research is often pioneering, relying on unconventional, innovative approaches that take a long and bumpy road and require patience and a tolerance for failure. Because the probability of success is typically small but the potential payoff is very large, breakthrough research is often described as high-risk, high-gain research. Concern exists that funders of research are insufficiently supportive of the pioneering science needed to achieve breakthroughs (Alberts et al., 2014; Petsko, 2012).

In Franzoni, Veugelers and Stephan (2022) we highlighted the challenges faced by scientists who did pioneering-research related to mRNA-based drugs in getting support for research. We used that exceptional moment for science to reflect on whether the government funding system is sufficiently supportive of research needed for key breakthroughs, and whether the system of funding encourages sufficient risk-taking to induce scientists to explore transformative research paths. We suggested interventions to avoid such bias.

Since then, the landscape has evolved considerably in terms of new empirical indicators and analyses of conservatism in funding, as well as new policy experimentation. This updated chapter revisits the core arguments, integrates recent evidence, highlighting the newest areas of experimentations, and reflects on emerging trends, most notably a noticeable tightening of public science funding in the U.S., growing emphasis on directing research towards national priorities, and tangible impact, but also effort to reduce the use of bibliometrics. These shifts further underscore the tension between breakthrough science and the conservative structures that shape funding decisions. We conclude by drawing a research agenda for scholars and policy makers moving forward.

2. Breakthrough research: features, measures and implications

Pathbreaking research that can lead to breakthroughs is characterized by unconventional and often untested approaches that carry a high probability of failure but a high potential value, if successful. Such approaches may adopt novel ideas, deviate from prior trajectories, recombine distant bodies of knowledge, sometimes, though not always, across the boundaries of disciplinary domains. When successful, they lead to the emergence of new concepts and constructs that are embraced by others, sometimes disrupting prior knowledge and other times branching out in new areas.

Scholars seeking to study breakthrough research empirically have developed measures that focus on several of the elements and dimensions just recalled, including risk, adoption or introduction of new ideas, knowledge recombination, pivoting, disruptiveness and interdisciplinarity. All indicators capture a partial, yet complementary perspective of breakthrough research.

Measuring risk-taking in research is particularly challenging. Risk is a broad and ill-defined concept that takes different meanings in different fields (Althaus, 2005; Aven, 2011; Franzoni & Stephan, 2023; Hansson, 2018). In science, *risk* has a *speculative* meaning and refers to the uncertainty of research outcomes. Unlike other domains of speculative risk, such as finance, outcomes tend to be positive, ranging from groundbreaking discoveries to no progress, rather than direct damage and risk/rewards are not necessarily correlated (Franzoni & Stephan, 2023). In practice, however, measuring *risky research* empirically remains an open challenge.

Azoulay and Greenblatt (2025) propose, among other indicators, to measure "extreme tail outcomes", defined as the distance between the highest and lowest percentile in citations of the realized outcomes. Other measures of breakthrough research look at the degree to which research deviates from the past and/or look at the building blocks upon which the research is based. Foster and colleagues (2015) working on chemical research, distinguish three types of papers. Research that makes a *jump* explores previously unexplored chemical relationships -jumping beyond current knowledge-. Such research arguably is more likely to fail but, if the research succeeds, is more likely to make a breakthrough. Research that explores relationships between previously studied entities is subdivided into research that tests a *new* relationship, not published before, or research that *repeats* an analysis of a previously studied relationship. Their findings suggest that taking the risk associated with *jump* and *new* research makes it more likely to achieve high impact, but also to failure, consistent with risk-taking. Interestingly. the additional rewards associated with *jump* papers are, relatively small and may not compensate sufficiently for the possibility of failing, suggesting higher expected returns of a safer research path.

Wang et al. (2017) take a similar approach: They measure whether a published paper makes first time ever combinations of scientific knowledge, but they use combinations of referenced journals, accounting for the difficulty of making such combinations. Almost all new combinations cross subject categories. They show that novel papers have higher mean and variance in citation performance and also a higher probability of becoming highly cited, or have no/low citations. Wang et al. (2017) also find strong evidence that novel research takes more time to become top-cited and that it is published in lower-tiers journals, as measured by the Journal Impact Factor. These findings suggest that bibliometric indicators based on citation counts and journal Impact Factor with a short citation window, may be biased against risky, novel research. They also show that citations to novel papers are more likely to come from a broader set of disciplines and from approaches that are more distant from their "home" field, suggestive that novel research has a tendency to be best appreciated and spark applications across disciplinary boundaries.

Uzzi, Mukherjee, Stringer, and Jones (2013) focus on atypical combinations of prior knowledge as measured by the proximity between pairs of cited journals. Operationally, they calculate the relative commonness of journal pairs and identify *atypicality* as the lowest 10^{th} percentile commonness score and the median commonness score as an indication of *conventionality*. They find that papers with both atypicality and conventionality are more likely to become top cited.

Funk & Owen-Smith (2017) as well as other authors (Park et al., 2023; Wu et al., 2019) focus on patterns of citations over time and introduce a metric that characterizes if a paper is disruptive or consolidating. A focal paper is assumed to be more disruptive if the papers that cite the focal paper do not cite the focal paper's predecessors, i.e. the papers cited by the focal paper. A paper is considered consolidating if it is mostly cited together with its predecessors. Disruption thus involves novelty that has a disruptive impact, although only a particular subset of novel research qualifies as disruptive research (Park et al., 2023).

The measures just reviewed are based on patterns of citations. Yet, this approach has important limitations. Citations capture prior art but not the scientific content and contribution of the paper itself. Other approaches have focused on new scientific ideas in the content of a paper. Azoulay, Graff-Zivin and Manso (2011) and Packalen and Bhattacharya (2020) look at the research ideas contained in biomedical papers, as represented by the Medical Subject Headings (MeSH), a collection of keywords curated by the National Library of Medicine. They identify the age at which the ideas were first coined, as a proxy for novelty. Unusual combinations of MeSH descriptors can also be used to characterize recombination of ideas (Azoulay et al., 2011). Boudreau, Guinan, Lakhani, and Riedl (2016) take a similar approach to characterize research proposals, rather than papers. To do so, they extrapolate MeSH terms from proposal texts and identify the fraction of new MeSH pairs combinations as those which have never appeared in the previous literature in PubMed.

Following the seminal contribution of Kuhn (1962), who argued that new scientific ideas can only be identified through shifts in language, new text mining techniques have been used to identify novelty ideas and concepts expressed in scientific texts. A substantial body of research, scattered across various disciplines from physics to sociology, is exploring high-dimensional data from NLP for detecting novel scientific ideas as well as trace the diffusion and impact of these ideas over time (Kelly et al., 2021; Shi & Evans, 2023; Shibayama et al., 2021; Yin et al., 2023). More recently, Arts, Melluso and Veugelers (2025) use natural language processing (NLP) techniques to identify the origin and impact of new scientific ideas in the population of scientific papers. They use the titles and abstracts of all scientific publications covered in the latest version of Microsoft Academic Graph (MAG). To detect new scientific ideas and measure a paper's novelty at publication, they identify words, noun phrases, and novel combinations of words or noun phrases that appear for the first time. Alternatively, they measure a paper's novelty based on the similarity of its entire text to all prior papers, using SPECTER, a pre-trained document-level embedding of scientific papers (Cohan et al., 2020). To measure the impact or influence of new scientific ideas, they count the number of subsequent papers reusing new words, noun phrases, or combinations of either words or noun phrases. To validate their methods, they analyze Nobel Prize-winning papers, which likely pioneered impactful new ideas, and literature review papers, which typically consolidate existing knowledge. The results illustrate the improvement of text-based metrics over traditional citation-based measures to predict Nobel Prize winning papers. They also show that papers introducing new ideas, particularly those with greater follow-on reuse, attract more citations and that novel papers have more intellectual neighbors published after them, indicating they are ahead of their intellectual peers.[1]

3. A review of empirical evidence on conservatism in funding

New empirical studies on conservatism in funding that use new methods and indicators for measuring risky research have multiplied in recent years. Commentators on

[1] They provide open access to all data and code. Data: https://zenodo.org/records/13869486. Code: https://github.com/nicolamelluso/science-novelty. Accessed September 18, 2025.

science policy have long lamented that science funders are too conservative and risk averse and skimp on supporting breakthrough research (e.g., (Laudel, 2017; Mazzucato, 2015; Viner et al., 2004). The growing body of empirical evidence, reviewed in this section, finds indications confirming conservatism in science in funding.

Azoulay et al. (2011) found that HHMI investigators use more novel keywords and produce more hits and more flops, compared with the NIH investigators. It is not clear, as they are quick to point out, whether the results depend on the criteria for selection or other factors, such as the longer duration of grants and the practice of HHMI to not require preliminary results or to expect early research results. Wagner and Alexander (2013) evaluate the SGER NSF program designed to support high risk, high reward research that ran from 1990 to 2006. Funding decisions were made entirely by program officers with no external review. The authors find that program officers routinely used but a small percent of available funds. The authors interpret the findings as suggesting that either officers were averse to funding risky research, despite the number of funded proposals that had transformative results, or that risk taking was not rewarded within NSF. In an experiment conducted at the Harvard Medical School, Boudreau, Guinan, Lakhani, and Riedl (2016) found that more novel research proposals, as measured by the percent of keyword-pairs that did not previously exist in the published scientific literature, receive more negative evaluations during peer-review than do less novel ones. The result is driven by proposals with particularly high levels of novelty (Boudreau et al., 2016). Packalen and Bhattacharya (2020), using the age of MeSH terms embodied in a paper as a proxy for novelty, find that NIH was more likely to fund ideas of intermediate maturity (7 to 10 years old) and were less likely to fund consolidated ideas, but also the ideas that were more recently born. Moreover, the NIH propensity to fund projects that build on the most recent ideas had declined over the last several decades. Lanoë (2019) studied the French ANR funding programs directed towards new areas and found that although individuals with a history of novel research are more likely to apply, they are not more likely to get funding. They use pairwise combinations of author keywords to identify novel publications Azoulay and Greenblatt (2025) study the renewals of R01, the most common type of NIH grant. They find that grants whose first-stage outcomes reflected greater risk-taking were significantly less likely to be renewed, based on four indicators. The magnitude of the effect - between 4% and 9.5% lower renewal rate- appears large enough to suggest that it may not entirely depend on lack of trying to renew.

Veugelers, Stephan and Wang (2025) examine whether the ERC, the most important funding agency of the European Commission, set up in 2007 with the explicit aim to fund “high gain/high risk” research, is biased against novelty. They find a significantly negative selection effect for applicants with a track record of producing highly novel publications prior to application. The bias against novelty holds also for top-cited researchers. Moreover, the negative selection against novelty is larger and more significant for early career applicants than for advanced career applicants, suggesting that panel members are less willing to tolerate highly novel approaches for early career applicants than for established researchers. This holds only for early career applicants in non-top host environments. Early career applicants operating in top host environments escape this bias against novelty suggesting that panel members are more likely to tolerate novel approaches from juniors working in “trusted” environments. In a difference-in-differences analysis they find no significant treatment effects for advanced career recipients, only positive treatment effects for early career grantees. This positive treatment effect is however in part due to unsuccessful applicants cutting back on highly novel research. The authors interpret as a strategy on the part of the failed junior applicants to prepare for resubmission to a research funding system that they perceive to be biased against novel research. Using the novelty measured developed by Wang et al. (2017), Ayoubi, Pezzoni, and Visentin (2021) investigated the Swiss National Science Foundation’s SINERGIA program that promotes interdisciplinary, collaborative and breakthrough research and found that scientists inclined towards novel research are more likely to apply but less likely to be funded.

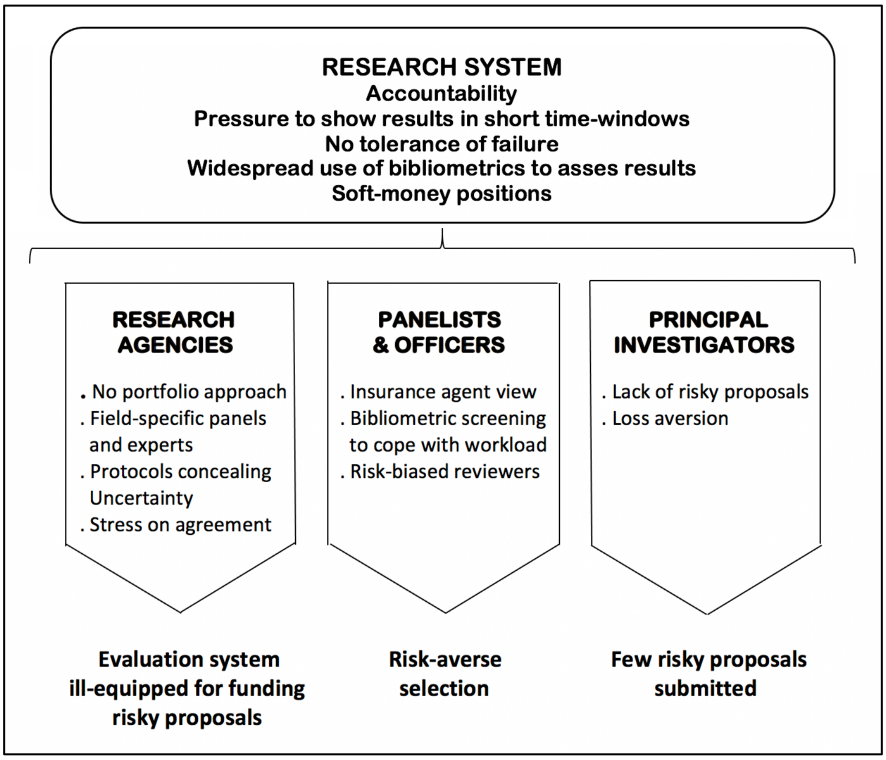
Feliciani, Christensen and Franzoni (2025) estimate the novelty of over 40 thousand proposals submitted to 273 calls of the Novo Nordisk Foundation over a ten-year period. They extrapolate four metrics among those described in the previous section from the text of both accepted and rejected proposals. They focus on conservatism in panel decisions, controlling for the scores with which proposals enter the discussion. Their estimates show a small but persistent penalty for novel proposals in panel choices for regular calls -on the order of 1-2%- regardless of the metric used. By contrast, those calls designed to support high-risk high-reward and interdisciplinary research to not exhibit such penalty, suggesting that novel proposals are more likely to be disadvantaged when competing directly against more conventional research.

The evidence accumulated over time, and discussed above is now substantial and consistently indicates that research projects most likely to generate breakthroughs are disadvantaged in the competition for funding. This appears to be particularly the case when bold proposals compete directly against more conventional ones, but it is observed even in the case of the ERC, whose programs should be designed to support especially high-risk high-reward science. In the next sections we turn to the question of why funding agencies eschew supporting risky research and thereby miss opportunities to fund breakthroughs. Is this a deliberate or unconscious choice, embedded in their modes of operation? And what can be done to encourage greater risk-taking among funders?

4. Why Is Science Conservative?

Given the complexity of the research system and the many actors involved, multiple factors are likely to interact and play a role. In this section we provide an overview of potential explanations, following the structure adopted in Franzoni et al. (2022) and summarized in Figure 1. While these factors remain in many cases possible rather than proved explanations, we also summarize recent progress made in understanding some of them.

Figure 1. Incentives and opportunities regarding risky research: A summary.



Source: Franzoni et al. (2022)

We start with hypotheses that arise from outside the funding agencies and relate to the broader research system. We then discuss how these translate into a set of incentives and opportunities that could induce principle investigators to refrain from writing and submitting grant proposals, panelists and officers to select more conventional research and funding agencies to set-up programs and evaluation procedures that disadvantage pathbreaking proposals.

4.1 Research system

In the last decades the research system has arguably become less supportive of breakthrough research and less tolerant to failure. First, there has been a generalized push for more accountability when using public funds. Many national governments have set up regular evaluations of universities that put increasing pressure on publicly funded science institutions to show results, especially those aligned with political cycles, favoring shorter and targeted results. Shorter windows for results bias against untargeted research programs in general; witness the heated discussions on the share of overall public R&D budgets for bottom-up untargeted research programs like ERC, NSF and NIH, compared to more targeted, top-down agendas and applied research programs. But even within untargeted research programs, the pressure to show results quickly may discourage publicly funded agencies from funding more breakthrough research. A major factor is impatience: Breakthroughs take a long time to materialize (Wang et al., 2017). Funding agencies may feel that they cannot afford to wait long to show impact and opt instead to fund safer projects that give tangible results in the near term, even if these are less likely to be breakthroughs.

Second, pressure for accountability is particularly impactful when it is combined with the widespread use of bibliometric indicators. Commonly-used metrics, like the journal Impact Factor or short windows for calculating citations to measure research “quality” can discourage risky research, as these measures appear to be biased against novel research (Stephan et al., 2017).

Third, tenure positions, which in academia mark the passage from a period of short-term performance scrutiny to a period of job security, have become less abundant over time, are reached at a later age, particularly in the biomedical fields, and are increasingly subjected to mandatory periodic re-examination that can lead to the dismissal of underperforming professors (Clement, 2022). Empirical research has shown that, after tenure, scholars engage more in projects that depart from pre-existing research trajectories and produce more novel works, likely because tenure shields researchers from the worst consequences of a project failure (Franzoni & Rossi-Lamastra, 2017; Tripodi et al., 2025). Consequently, policies that limit, delay or constraint tenure can be effective in boosting short-term productivity but also discourage engagement in bold scientific projects. A similar problem concerns the large use in the US of "soft money" positions, i.e. in positions where salary is funded from grants that the researcher is responsible for obtaining, making the university free to cut its losses and hire another individual into the position when the grant expires or is not renewed (Stephan, 2007). To the extent that they require short-term results, untenured positions, tenure re-exams and soft money positions discourage breakthrough research on the part of the faculty, particularly in periods, like the one unfolding, of large budget cuts.

4.2 Principal investigators

A system of incentives based on short-time outcomes and reliant on risk-biased metrics inevitably may constrain the supply of pathbreaking research proposals from Principal Investigators. Consistent with this, Veugelers, Stephan and Wang (2025) find that non-funded junior ERC applicants who fail in the second stage have significantly lower likelihood of producing novel papers after being rejected, compared to the successful ones, suggesting that rejected applicants learn that novel research is not rewarded. Faced with the pressures to (re-) apply for funding, they adjust their research portfolio away from risky research, something which the successful applicants are "freed" from doing.

The findings from Foster and colleagues (2015) reported supra, also suggest that returns may be higher for following a safer research path. Stephan (2019) has called this the *"Quad effect"*, referring to the fact that competitive female figure skaters attempt fewer quadruple jumps, arguably because the incremental score they can earn for completing a quad, compared to successfully completing a triple jump, is insufficient to compensate for the risk of failing to complete the quad jump.

The preferences of scientists for the level of risk involved in the projects they wish to pursue may not only reflect biases against risk in the reward structure of science, as discussed supra, but also loss aversion on the part of scientists, in line with the general human tendency to over-estimate the magnitude of perspective losses and under-estimate the magnitude of perspective gains (Kahneman & Tversky, 1979; Tversky & Kahneman, 1991).

Short supply of innovative approaches may also reflect the limited diversity of the scientific workforce, which—despite progress over time—remains predominantly composed of white men in most of the STEMs, especially in senior and tenured positions (Ceci & Williams, 2011; Hoppe et al., 2019). Hofstra and colleagues (2020), analyzing more than thirty years of U.S. doctoral dissertations, found that research conducted by scholars from minority groups—defined as underrepresented genders and ethnicities within their discipline—tended to be more innovative (i.e., more recombinatorial) than that of majority groups. Yet, these novel approaches were taken-up at lower rates by other scholars, compared to those from majority groups and led to less successful careers. Thus, limited diversity in the research community may be listed as one additional factor hindering the supply of breakthrough research.

4.3 Research agencies

The design of evaluation and selection procedures in research agencies is a critical step, with direct repercussions on what is funded and indirect repercussions on what is seen as fundable and therefore worth pursuing. Several elements of the selection process can in turn favor or disfavor breakthrough research.

To begin with, regular and risky research are normally competing in the same programs, with the results that risky research is disadvantaged due to low probability of success, early-stage ideas, lack of preliminary data and penalty for interdisciplinarity. Only a minority of research agencies operate special programs targeting high-risk, high-gain proposals, where applicants and evaluators are explicitly instructed to prioritize bold projects. Feliciani, Christensen and Franzoni (2025) find that programs explicitly dedicated to high-risk high-gain research, where they exist, do not exhibit the conservatism penalty observed in general funding schemes. Yet, such initiatives remain rare and account for only a small fraction of total research budgets. For example, the NIH has created four programs with these aims, designed for rapid approval and without requirements for preliminary data, though one of them has recently been put on hold. Even so, the programs are modest in scope: representing only about 0.15% of the NIH extramural budget, with success rates of roughly 5% (Stephan & Franzoni, 2023).

Research agencies also exert direct control on the procedures and practices used to evaluate proposals. These involve decisions such as the disciplinary composition of panels, the protocols defining evaluation criteria, the choice and number of evaluators, and the methods used to aggregate their opinions. While these may appear to be neutral procedural technicalities, they can have subtle unintended repercussions on the evaluation of breakthrough research and its chances to get funded.

Review panels are often designed by funding agencies to be discipline-based. This, for example, is generally the case at NSF, and ERC. The latter for instance operates with 25-panels which are mostly discipline-focused. As noted supra, papers of high novelty are often interdisciplinary. Wang et al. (2017) find that novel work that is highly cited is more likely to garner citations from outside, not from within its own field, suggesting that the research is appreciated more by others than by colleagues. Monodisciplinary panels may thus more likely be biased against risks associated with novel interdisciplinary research. Consistent with this view, Banal-Estanol and colleagues (2019) show that proposals from interdisciplinary teams submitted to the UK's Engineering and Physical Sciences Research Council—where evaluation also relies on disciplinary panels—were less likely to be funded, although when funded, they tended to be more successful.

Evaluation protocols, intended as the form used to elicit experts' opinions by a formal set of criteria and scores may are generally unsuitable to capture the essence of breakthrough research. Suppose, for example, a reviewer believes that a project could deliver a breakthrough but with a low chance of success. If the protocol asks to assess the project impact giving only a single score on a scale -say from 1 to 6-, the reviewer will likely give a low or average value. Even though the reviewer has recognized the possibility of an exceptional outcome in the tail of the distribution, the protocol would not allow the possibility of express this insight. Whereas in peer review of 'standard' science, the provision of a single point-estimate may provide a necessary time-saving compromise, in evaluations of risky research, the outcomes of interest can be expected to be in the tails, and a single-point estimate may have little meaning (Franzoni & Stephan, 2023).

Evaluations of proposals are done by collecting several expert opinions, which are then that *aggregated*. In practice, however, aggregation mechanisms may not be neutral with respect to breakthrough research. Proposals involving radically novel approaches are generally more controversial, they are not grounded on preliminary findings and allow ample room to challenge their validity. They are consequently more likely to spark disagreement, given the larger uncertainty involved. Major funding agencies, like NIH and ERC use *consensus meetings* to aggregate reviewers' opinions and resolve disagreement through panel discussion. In such meetings, the need to take collegial decisions may favor proposals that are uncontroversial but also less innovative, because these are more palatable to the majority of panelists and harder to criticize (Lamont, 2009). The work by Felician et al. (2025) cited supra supports this view. Collegial decisions are also vulnerable to groupthink that can bias outcomes towards mainstream especially if endorsed by reputable peers (Lamont, 2009). Franzoni and colleagues (2025) find that prestigious scholars exert a strong influence on funding decisions but are not more accurate. This is consistent with the findings from Della Vigna and Pope (2018) that academics wrongly overestimate the accuracy of prominent scholars. Consequently, consensus meetings may lead people to herd away from the truth, following influential mainstream opinions.

Panels are generally organized to review proposals on a "one by one" basis, expressing the merit of each proposal separately, without considering an overall desirable level of funding for bolder research.[2] This "one by one" approach leaves no room for hedging risk, which would be possible by taking a portfolio approach. To the extent that they are risk averse, the "one by one" approach only aggravates the risk-taking problem.

4.4 Panelists and research officers

The workload of agencies and evaluators is typically substantial. For example, the NIH evaluates approximately 80,000 applications annually, engaging over 2,000 reviewers per years and has more than 150 standing Study Sections (Franzoni et al., 2022). The ERC averages 15 members on each of its 25 separate panels. The average panel member for the Starting Grants looks at more than 100 proposals per call; Given the heavy workload, it is not surprising that reviewers and panel members may seek ways to rapidly screen proposals, especially on a first pass. One of the easiest ways to do so is to focus on the publishing record of the scientist proposing the research, by examining readily available bibliometric indicators on platforms such as Google Scholar and Scopus. The use of bibliometrics may affect the panel's decision when it comes to supporting risky research. The work by Wang, Veugelers and Stephan (2017) suggests that novel research is systematically less likely to be published in journals with high Impact Factor and takes longer to be a top hit than does non-novel research. This bibliometric penalty can lead panels to select against individuals with a

[2] At some agencies, such as NSF, program officers have some leeway in making decisions, but this is not common.

history of novel (risky) research, especially when the applicant is young and has a short history of citations. More generally, a focus on bibliometrics shifts the basis of decisions away from the substance of the proposal to an easily accessible metric.

Peer review evaluations may also disadvantage breakthrough research in more subtle ways. When panelists fear that modest results could undermine future support for a program, they may focus excessively on "what can go wrong" and favor proposals that appear feasible or present strong "preliminary findings", offering reassurance that the research will not come up "empty-handed." In effect, this creates an implicit requirement that research be de-risked before it is funded. Consistent with this view, empirical analyses that disentangle reviewers' scores across sub-criteria suggest that reviewers place a heavy emphasis on feasibility (Lane et al., 2022). Specifically, sub-criteria assessing the strength of a proposal's approach and methodology are weighted roughly twice as heavily as those assessing its potential impact on science and society (Franzoni et al., forthcoming). Because breakthrough research is typically early-stage, but high-potential —weaker in methodological rigor but stronger in potential impact—this weighting can systematically disadvantage novel and exploratory proposals, consistent with findings from Boudreau and colleagues (2016).

5. Reducing Conservatism: Emerging Approaches

The discussion in the prior section has outlined a number of possible root causes that may contribute to hinder breakthrough research. This section follows up with a set of possible remedies and related evidence regarding their efficacy, when available.

Failure-tolerant research system

Breakthrough research requires environments that tolerate failure. One way to foster tolerance to failure is by reducing the pressure for showing short-term outcomes and instead emphasizing progress toward broader long-term goals. An interesting benchmark in this respect is provided by entrepreneurial communities, where failure is not stigmatized, and narratives highlighting lessons learnt from failure are actively encouraged. PhD programs and initiatives for early-career scholars could play an important role in cultivating a similar culture (Pramanik, 2024).

Journal policies that permit the publication of registered reports, including those that do not produce the expected results, are another step in this direction. In recent years, such policies have been promoted by several communities, including the Center for Open Science[3] and over 300 journals have adhered to the policy, including the flagship journal *Nature*.

Deemphasize bibliometrics

Bibliometric indicators, as we discussed, are particularly problematic for novel and pathbreaking research. Although still widely used, there is increasing awareness of their limitations. Universities and funding agencies can reinforce this trend, by embracing the principles of responsible research assessments, as articulated in the DORA Declaration[4] and Leiden Manifesto.[5] One example is the curricula template introduced in 2022 by the Swiss National Science Foundation, which requires "short narratives in combination with a limited number of research results, rather than an extensive publication lists".[6] Although we do not yet have evidence of their effectiveness, there is considerable hope that such practices should discourage bibliometric screening.

Dedicated funding programs for breakthrough risk research

Programs dedicated to breakthrough research can be an effective way to bolster more risk-taking (Feliciani et al., 2025). But even when these programs are operative, there is a need to carefully select, instruct and direct panelists to ensure they factually prioritize such research, differently than the usual excellence selection. The experience of the ERC suggests that this cannot be taken for granted (Veugelers et al., 2025). Moreover, it is possible that these programs are particularly effective if they are part of a well-balanced portfolio that has separate funding lines for both regular and risky research, so that investigators can chose the most suitable funding line and different types of proposals compete in their own category. When funding is consolidated in a common program, like at the ERC, the competition of high-quality but not so innovative projects may simply be too strong (Veugelers et al., 2025).

[3] https://www.cos.io/initiatives/registered-reports. Accessed September 17, 2025.
[4] https://sfdora.org/. Accessed September 17, 2025.
[5] https://www.leidenmanifesto.org/. Accessed September 17, 2025.
[6] https://www.snf.ch/en/wBR6E3emu8PP1ZSY/news/a-new-cv. Accessed September 17, 2025.

Allow for disagreement

We stressed that breakthrough research may lead to more polarized views in evaluations. This warns against aggregation methods, such as consensus meetings, that emphasize collegiality and calls for alternative aggregation mechanisms. Considering this limitation, several funding agencies are currently experimenting mechanisms that relax the requirement of consensus in various ways (Feliciani et al., 2025). One of these is the use of "golden tickets" that allow each panelist to promote one application per call, even without the consensus of the other panelists. Recent experimentation from the Volkswagen Foundation seem to indicate that this mechanism may be useful in promoting novel research (Haroff et al., 2025). Another approach, implemented by the Swiss National Science Foundation, the Novo Nordisk Foundation, and others is to use partial lotteries, where the panel can use a random draw to select among applications that are in a tie. Although the impact of this policy is still unknown, simulations of its impact suggest that partial lotteries may have a limited impact on novel research and could work best only when they are used in targeted programs designed to prioritize breakthrough research (Feliciani et al., 2025).

Portfolio approach and staging

The 'one by one' approach typically used in panels works against selecting risky proposals. At a minimum, panels need to think about diversifying the risk profile of the proposals they are funding. Requiring that all successful proposals have convincing preliminary findings would imply no portfolio approach. More generally, a portfolio approach to address risk aversion could require panels to put in different baskets highly risky and moderately risky proposals and provide a way to choose proposals from each. In practice such a portfolio approach could be quite challenging to implement for research projects. A portfolio approach poses issues of fairness, as some proposals may have to be eliminated to balance or de-risk the portfolio, even if they have merit. Moreover, portfolio theory requires that the research paths be sufficiently uncorrelated and correlation between research paths, in and of itself, can be hard to determine, particularly for scientific and basic research covering vastly different goals across different fields and with different research approaches. Such conditions may be more suited for mission-oriented research, where funding in parallel of competing strategies is used to overcome major technological bottlenecks or achieve breakthrough milestones. This strategy has long been adopted by DARPA. In recent years similar approaches in mission oriented programs have emerged in other countries, including the ARIA[7] in the UK and the SPRIND[8] in Germany.

Similar considerations apply for the approach to fund in stages, common at DARPA and in the Venture Capital industry, where increasingly larger amounts of funding are allocated, depending on whether interim milestones are being met. Can such a staging approach also be used by science funding agencies, allowing them to take more risk? Interim evaluation is especially useful for projects that can start small and make quick tests (Ewens et al., 2018; Vilkkumaa et al., 2015). These conditions are more the exception than the norm in the natural sciences, where the share of research that requires expensive equipment and substantial effort is large (Stephan, 2012). This is probably one reason why we have not seen much progress in this direction.

6. Conclusions: What have we learnt and what do we need to learn

Society needs scientific breakthroughs to tackle the many challenges it faces. Because many of the paths to such breakthroughs require embracing unconventional approaches with high chances of failure, its science system, and particularly its public science funding system, needs to ensure that risk taking is encouraged or, at a minimum, that the system is not biased against risky research. The previous sections have made clear that we cannot take for granted that our science system will deliver this and have outlined possible ways forward, some of which are currently under testing.

As discussed supra, the promotion of risk needs to be addressed within the entire science system: It cannot be solved by an individual program or funding agency. It requires a holistic perspective on the science enterprise, activating not only funders and their reviewers, but also universities and research centers, journal editors and their reviewers and, last but not least, researchers themselves. Nevertheless, within the overall science system, science funding agencies can play a pivotal role, given the relevance of funding in science. Science funding agencies should be encouraged to pave the way for promoting risk taking in scientific research, if we don't want to miss breakthroughs.

---

[7] https://www.aria.org.uk/. Accessed September 17, 2025.
[8] https://www.sprind.org/en. Accessed September 17, 2025.

[Add here projections about what the future will bring: cuts in funding, less tenured positions, more directedness. TBD at workshop]

We conclude by sketching a research agenda for advancing our understanding of risk-taking, novelty, and conservatism in science funding. Future academic research on the topic should exploit new and complementary metrics to capture research trajectories for breakthroughs more accurately. Empirical studies are needed to assess potential biases of funding programs against breakthrough research. Experimentation with new designs should be more common and new approaches (such as narrative CVs, lotteries, long-term investigator grants) warrant systematic evaluation to determine their effects on risk-taking behavior and research outcomes. An open question remains whether, and how -mission-oriented research programs, often inspired by and mirrored on the DARPA model for bold funding—translate to basic science funding more broadly.

7. References

Alberts, B., Kirschner, M. W., Tilghman, S., & Varmus, H. (2014). Rescuing US biomedical research from its systemic flaws. *Proceedings of the National Academy of Sciences*, *111*(16), 5773–5777. https://doi.org/10.1073/PNAS.1404402111

Althaus, C. E. (2005). A disciplinary perspective on the epistemological status of risk. *Risk Analysis*, *25*(3), 567–588.

Arts, S., Melluso, N., & Veugelers, R. (2025). Beyond Citations: Measuring Novel Scientific Ideas and their Impact in Publication Text. *Review of Economics and Statistics*, 1–33. https://doi.org/10.1162/REST_A_01561/127741/BEYOND-CITATIONS-MEASURING-NOVEL-SCIENTIFIC-IDEAS

Aven, T. (2011). On Some Recent Definitions and Analysis Frameworks for Risk, Vulnerability, and Resilience. *Risk Analysis*, *31*(4), 515–522. https://doi.org/10.1111/j.1539-6924.2010.01528.x

Ayoubi, C., Pezzoni, M., & Visentin, F. (2021). Does It Pay to Do Novel Science? The Selectivity Patterns in Science Funding. *Science and Public Policy*, *48*(5), 635–648. https://doi.org/10.1093/SCIPOL/SCAB031

Azoulay, P., Graff Zivin, J., & Manso, G. (2011). Incentives and Creativity: Evidence from the Howard Hughes Medical Investigator Program. *The Rand Journal of Economics*, *42*(3), 527–554.

Azoulay, P., & Greenblatt, W. (2025). Does Peer Review Penalize Scientific Risk Taking? Evidence from NIH Grant Renewals. In *NBER Working Paper* (33495). https://doi.org/10.2139/ssrn.5140742

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

Bloom, N., Jones, C. I., van Reenen, J., & Webb, M. (2020). Are Ideas Getting Harder to Find? *American Economic Review*, *110*(4), 1104–1144. https://doi.org/10.1257/AER.20180338

Boudreau, K. J., Guinan, E. C., Lakhani, K. R., & Riedl, C. (2016). Looking across and looking beyond the knowledge frontier: Intellectual distance, novelty, and resource allocation in science. *Management Science*, *62*(10), 2765–2783. https://doi.org/10.1287/mnsc.2015.2285

Ceci, S., & Williams, W. (2011). Understanding current causes of women's underrepresentation in science. *Proceedings of the National Academy of Sciences*, *108*(8), 3157–3162.

Clement, T. P. (2022). How can we reform the STEM tenure system for the 21st Century? *Proceedings of the National Academy of Sciences*, *119*(33), e2207098119. https://doi.org/10.1073/PNAS.2207098119

Cohan, A., Feldman, S., Beltagy, I., Downey, D., & Weld, D. S. (2020). SPECTER: Document-level Representation Learning using Citation-informed Transformers. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 2270–2282. https://doi.org/10.18653/V1/2020.ACL-MAIN.207

DellaVigna, S., & Pope, D. (2018). Predicting Experimental Results: Who Knows What? *Journal of Political Economy*, *126*(6), 2410–2456.

Ewens, M., Nanda, R., & Rhodes-Kropf, M. (2018). Cost of experimentation and the evolution of venture capital. *Journal of Financial Economics*, *128*(3), 422–442. https://doi.org/10.1016/j.jfineco.2018.03.001

Feliciani, T., Nørding Christensen, R., & Franzoni, C. (2025). Addressing conservatism in research funding with lotteries, golden tickets, and novelty-dedicated panels. In *2025*.

Foster, J. G., Rzhetsky, A., & Evans, J. A. (2015). Tradition and Innovation in Scientists' Research Strategies. *American Sociological Review*, *80*(5), 875–908. https://doi.org/10.1177/0003122415601618

Franzoni, C., Guerini, M., Nørding-Christensen, R., & Stanaj, A. (2025, June 18). Expert Predictions and Errors in Research Funding Decisions | Academy of Management Proceedings. *Proceedings of the Academy of Management Annual Meeting*. https://journals.aom.org/doi/10.5465/AMPROC.2025.11186abstract

Franzoni, C., & Rossi-Lamastra, C. (2017). Academic tenure, risk-taking and the diversification of scientific research. *Industry and Innovation*, *24*(7), 691–712. https://doi.org/10.1080/13662716.2016.1264067

Franzoni, C., & Stephan, P. E. (2023). Uncertainty and Risk-Taking in Science: Meaning, Measurement and Management. *Research Policy*, *52*(3), 104706. https://doi.org/10.2139/ssrn.3804560

Franzoni, C., Stephan, P., & Veugelers, R. (2022). Funding Risky Research. *Entrepreneurship and Innovation Policy and the Economy*, *1*, 103–133.

Funk, R. J., & Owen-Smith, J. (2017). A Dynamic Network Measure of Technological Change. *Management Science*, *63*(3), 791–817. https://doi.org/10.1287/MNSC.2015.2366

Hansson, S. O. (2018). Risk. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2018). Stanford University. https://plato.stanford.edu/archives/fall2018/entries/risk/

Hofstra, B., Kulkarni, V. V., Galvez, S. M. N., He, B., Jurafsky, D., & McFarland, D. A. (2020). The Diversity–Innovation Paradox in Science. *Proceedings of the National Academy of Sciences*, *117*(17), 9284–9291. https://doi.org/10.1073/PNAS.1915378117

Hoppe, T. A., Litovitz, A., Willis, K. A., Meseroll, R. A., Perkins, M. J., Hutchins, B. I., Davis, A. F., Lauer, M. S., Valantine, H. A., Anderson, J. M., & Santangelo, G. M. (2019). Topic choice contributes to the lower rate of NIH awards to African-American / black scientists. *Science Advances*, *5*(10), 1–13.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263–292.

Kelly, B., Papanikolaou, D., Seru, A., & Taddy, M. (2021). Measuring Technological Innovation over the Long Run. *American Economic Review: Insights*, *3*(3), 303–320. https://doi.org/10.1257/AERI.20190499

Kuhn, T. S. (1962). *The structure of Scientific Revolutions* (C. U. Press, Ed.; IV edition).

Lamont, M. (2009). *How professors think. Inside the Curious World of Academic Judgement*. Harvard University Press.

Lane, J. N., Szajnfarber, Z., Crusan, J., Menietti, M., & Lakhani, K. R. (2022). *Are Experts Blinded by Feasibility? Experimental Evidence from a NASA Robotics Challenge* (Working Paper 22-071).

Lanoë, M. (2019). *The evaluation of competitive research funding: an application to French programs* [PhD Thesis]. L'Universite de Bordeaux.

Laudel, G. (2017). How do National Career Systems Promote or Hinder the Emergence of New Research Lines? *Minerva*, *55*(3), 341–369.

Mazzucato, M. (2015). *The entrepreneurial state. Debunking public vs. private sector myths*. Anthem Press.

Packalen, M., & Bhattacharya, J. (2020). NIH funding and the pursuit of edge science. *PNAS*, *117*(22), 12011–12016.

Park, M., Leahey, E., & Funk, R. J. (2023). Papers and patents are becoming less disruptive over time. *Nature*, *613*(7942), 138–144. https://doi.org/10.1038/s41586-022-05543-x

Petsko, G. A. (2012). Goodbye, Columbus. *Genome Biology*, *13*(5), 1–2. https://doi.org/10.1186/GB-2012-13-5-155/METRICS

Shi, F., & Evans, J. (2023). Surprising combinations of research contents and contexts are related to impact and emerge with scientific outsiders from distant disciplines. *Nature Communications*, *14*(1), 1–13. https://doi.org/10.1038/S41467-023-36741-4;TECHMETA

Shibayama, S., Yin, D., & Matsumoto, K. (2021). Measuring novelty in science with word embedding. *PLoS ONE*, *16*(7 July), 1–16. https://doi.org/10.1371/journal.pone.0254034

Stephan, P. (2007). Early Careers for Biomedical Scientists: Doubling (and Troubling) Outcomes presentation Harvard University. *Presentation Harvard University*. http://users.nber.org/~sewp/Early Careers for Biomedical Scientists.pdf

Stephan, P. (2012). *How Economics Shapes Science*. Harvard University Press.

Stephan, P. (2019). *Practices and Attitudes Regarding Risky Research*. https://www.metascience2019.org/presentations/paula-stephan/

Stephan, P., & Franzoni, C. (2023). *Encouraging high-risk high-reward research at NIH*. https://www.brookings.edu/collection/building-a-better-nih/

Stephan, P., Veugelers, R., & Wang, J. (2017). Blinkered by bibliometrics. *Nature*, *544*(7651), 411–412.

Tripodi, G., Zheng, X., Qian, Y., Murray, D., Jones, B. F., Ni, C., & Wang, D. (2025). Tenure and research trajectories. *Proceedings of the National Academy of Sciences*, *122*(30), e2500322122. https://doi.org/10.1073/PNAS.2500322122

Tversky, A., & Kahneman, D. (1991). Loss Aversion in Riskless Choice: A Reference-Dependent Model. *Quarterly Journal of Economics*, *106*(4), 1039–1061.

Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical Combinations and Scientific Impact. *Science*, *342*(6157), 468–472. https://doi.org/10.1126/science.1240474

Veugelers, R., Wang, J., & Stephan, P. (2025). Do funding agencies select and enable novel research: evidence from ERC. *Economics of Innovation and New Technology*. https://doi.org/10.1080/10438599.2025.2486344

Vilkkumaa, E., Salo, A., Liesiö, J., & Siddiqui, A. (2015). Fostering breakthrough technologies - How do optimal funding decisions depend on evaluation accuracy? *Technological Forecasting and Social Change*, *96*, 173–190. https://doi.org/10.1016/j.techfore.2015.03.001

Viner, N., Powell, P., & Green, R. (2004). Institutionalized biases in the award of research grants: A preliminary analysis revisiting the principle of accumulative advantage. *Research Policy*, *33*(3), 443–454. https://doi.org/10.1016/j.respol.2003.09.005

Wagner, C., & Alexander, J. M. (2013). Evaluating transformative research programmes: A case study of the NSF Small Grants for Exploratory Research programme. *Research Evaluation*, *22*(3), 187–197.

Wang, J., Veugelers, R., & Stephan, P. (2017). Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Research Policy*, *46*(8), 1416–1436. https://doi.org/10.1016/j.respol.2017.06.006

Wu, L., Wang, D., & Evans, J. A. (2019). Large teams develop and small teams disrupt science and technology. *Nature*, *566*(7744), 378–382. https://doi.org/10.1038/s41586-019-0941-9

Yin, D., Wu, Z., Yokota, K., Matsumoto, K., & Shibayama, S. (2023). Identify novel elements of knowledge with word embedding. *PLOS ONE*, *18*(6), e0284567. https://doi.org/10.1371/JOURNAL.PONE.0284567