Working Paper 24-064

Greenlighting Innovative Projects: How Evaluation Format Shapes the Perceived Feasibility of Novel Ideas

Jacqueline N. Lane
Tianxi Cai
Michael Menietti
Griffin Weber
Eva C. Guinan

Harvard Business School

Greenlighting Innovative Projects: How Evaluation Format Shapes the Perceived Feasibility of Novel Ideas

Jacqueline N. Lane*
Harvard Business School

Tianxi Cai
Harvard TH Chan School of Public Health
Harvard Medical School

Michael Menietti
Harvard Business School
Digital Data and Design Institute at Harvard

Griffin Weber
Beth Israel Deaconess Medical Center
Harvard Medical School

Eva C. Guinan
Dana-Farber Cancer Institute
Harvard Medical School
Digital Data and Design Institute at Harvard

Working Paper 24-064

*Corresponding author.

Funding for this research was provided in part by Harvard Business School.

Greenlighting Innovative Projects:

How Evaluation Format Shapes the Perceived Feasibility of Novel Ideas

Jacqueline N. Lane*
Harvard Business School &
Digital Data and Design Institute at Harvard
jnlane@hbs.edu

Tianxi Cai
Harvard T.H. Chan School of Public Health & Harvard Medical School
tcai@hsph.harvard.edu

Michael Menietti
Harvard Business School &
Digital Data and Design Institute at Harvard
mmenietti@hbs.edu

Griffin Weber
Beth Israel Deaconess Medical Center
& Harvard Medical School
griffin_weber@hms.harvard.edu

Eva C. Guinan
Dana-Farber Cancer Institute, Harvard Medical School
& Digital Data Design Institute at Harvard
eva_guinan@dfci.harvardedu

*Corresponding author

Abstract

Evaluation of novel projects is essential for scientific and technological advancement. However, evaluator bias toward a project's potential can obscure its limitations. This study investigates evaluation formats by contrasting combined assessments of novelty and feasibility with feasibility-only evaluations. In collaboration with a leading research university, we executed a field experiment during a grant-funding opportunity where evaluators were exogenously assigned either a "combined-standard" evaluation or a "feasibility-only" evaluation. Results indicate that, relative to feasibility-only evaluators, combined-standard evaluators assign higher feasibility scores to novel projects, possibly overlooking limitations that could hinder a project's implementation. In contrast, a feasibility-only approach might improve early detection of reasons novel projects could fail, suggesting that single-criterion evaluations of project feasibility could be more effective in recognizing potential pitfalls in innovative projects.

Keywords: scientific and technological innovation, project evaluation, feasibility, novelty, field experiment

INTRODUCTION

Evaluating innovative projects is fundamental to organizational performance, technology adoption decisions, allocation of resources for research and development, responses to competition, and firm growth (Abernathy & Clark, 1985; Boudreau et al., 2016; Criscuolo et al., 2017; Lane et al., 2022; Tan, 2022; Tripsas & Gavetti, 2000). Essential aspects of evaluation processes often include the assessment of the novelty of new projects and ideas to examine the degree to which they are original and depart from existing knowledge, technologies, products, and services (Berg, 2016; Boudreau et al., 2016; Criscuolo et al., 2017; Falchetti et al., 2022; Mount et al., 2021). At the heart of innovation across science and technology is the creation of something new (Schumpeter, 1942). Novel products and innovations have the potential to open new markets and drive long-run economic growth (Hasan & Tucci, 2010; Mowery & Rosenberg, 1999; Schumpeter, 1942), and novel research can lead to scientific breakthroughs and open new paradigms and fields (Kuhn, 1977; Stephan, 2012; Uzzi et al., 2013). Similarly, technological discontinuities are more likely to occur when firms pursue radical, as opposed to incremental, innovation that builds on existing capabilities and technologies (Tripsas & Gavetti, 2000; Tushman & Anderson, 1986). However, their potential for significant reward notwithstanding, novel projects are also associated with a higher risk of failure (Baucells & Heukamp, 2012; Fleming, 2001; Leiponen & Helfat, 2010). Hence, firms may favor incremental innovations that build on existing knowledge and capabilities, as they are typically associated with a lower risk of failure compared to the uncertainties of novel activities (Benner & Tushman, 2003; Gupta et al., 2006; Katila & Ahuja, 2002).

Building on the notion that innovative initiatives often yield greater impact but fail more often, this paper investigates the effectiveness of different evaluation formats in identifying potential feasibility issues of early-stage novel projects. Here, we define novelty as the extent to which a project is original and departs from the existing knowledge frontier (Boudreau et al., 2016; Kuhn, 1977; Weitzman, 1998), and feasibility as the likelihood that a project can be successfully implemented (Franzoni & Stephan, 2023; Lane et al., 2022). A high-quality project balances both aspects (Amabile, 1982; Berg, 2016; Diehl & Stroebe, 1987; Rindova & Petkova, 2007). We propose that ex ante feasibility-only reviews that focus on assessments of a project's feasibility may be useful in identifying a project's critical limitations before resources are allocated. Detecting these challenges early on can increase the likelihood of successfully developing novel projects into completed products or innovations.

The need to thoroughly assess the feasibility of novel early-stage projects is particularly important due to their uncertainty (Fleming, 2001; Hoetker, 2005; Leiponen & Helfat, 2010) and, in most cases, the staged progression of development of scientific and technological projects (Azoulay & Li, 2020; Guler, 2007; Kerr et al., 2014). For instance, recent evidence indicates that many of the factors leading to late-stage failures in the pharmaceutical industry could have been detected before the research began, making these failures potentially *preventable* (Briel et al., 2016; Fogel, 2018). Despite uncertainty in the innovation process and the potential for a project's feasibility to change over time (Dougherty, 1992), early identification of the critical issues hindering a project's implementation can enable corrections or timely termination prior to investment (Åstebro & Elhedhli, 2006; Khanna et al., 2016).

Most organizations conduct evaluations of innovative proposals before resource allocation decisions have been decided (Bian et al., 2021; Boudreau et al., 2016; Criscuolo et al., 2017; Knudsen & Levinthal, 2007). Yet assessments of novel projects as high-risk or less feasible, however, are often based on observed outcomes (Fleming, 2001; Scherer & Harhoff, 2000; Wang et al., 2017). Despite recent studies examining how evaluation formats shape experts' assessments of novel projects (Bian et al., 2021; Criscuolo et al., 2021; Lane et al., 2022), little research has explored how novelty and feasibility are perceived in their evaluations. This leaves an incomplete understanding of whether evaluative processes are effective in exposing the potential limitations or flaws that may hinder the likelihood of developing and carrying out the project. The standard evaluation formats used to assess the novelty and feasibility of early-stage projects can lead to correlated assessments (Cooper, 1981; Nisbett & Wilson, 1977; Sahoo et al., 2012; Thorndike, 1920) due to such factors as prioritization of certain criteria over others (Falchetti et al., 2022; Mount et al., 2021) and the use of heuristics (Åstebro & Elhedhli, 2006; Haas et al., 2015; Kahneman et al., 1982). As the primary objective of innovation activities within firms is to identify novel projects (Piezunka & Dahlander, 2015; Rindova & Petkova, 2007), it is possible that project novelty could shape evaluators' assessments of feasibility.

Evaluators may perceive an *inverse* relation between novelty and feasibility if novel projects are viewed as riskier (Ferguson & Carnabuci, 2017; Howell & Higgins, 1990; Leiponen & Helfat, 2010; Rosner, 1968) or if the proposed work is less technically familiar or appears in a new context (Resch, Ernst, & Garrow, 2000). Conversely, they may see a *positive* correlation if novelty triggers optimism that leads to underestimating risks (Amore et al., 2021; Wells et al., 2010). These perceptions could lead decision-makers to greenlight novel projects without a complete understanding of their feasibility.

As standard evaluation formats combine assessments of novelty and feasibility (Amabile et al., 1996; Boudreau et al., 2016; Diehl & Stroebe, 1987), this means there is typically no baseline for comparing how evaluations might differ if feasibility were considered separately, posing a challenge to investigating these issues. Given limited attention and time, evaluators may use mental shortcuts to prioritize certain criteria over others, such as a project's novelty over its feasibility.

To better dissect these issues, we partnered with a U.S. medical school's translational science program to design and execute a field experiment that involved modifying its pilot grant-funding process. We chose this setting for its focus on knowledge creation and scientific peer review of novelty and feasibility in early-stage projects. The experiment design established a control group of expert evaluators exposed to a standard form that involved assessing both novelty and feasibility, termed "combined-standard," and a treatment group exposed to a form focused exclusively on feasibility, termed "feasibility-only."[1] Evaluators in the treatment condition were instructed to consider only a project's ease of implementation given proposed resources. All evaluators used identical feasibility criteria, first judging a project's overall feasibility, and then assessing how they factored various sub-criteria, such as proposed resources, technical feasibility, and ethical or safety concerns into their score. We mobilized 97 evaluators from seven institutions to assess 47 projects, which yielded 641 proposal-evaluation pairs and $500,000 in funding.

Our experimental results revealed that feasibility-only evaluators assign feasibility scores lower by 6–10 percentage points (pp) than those of the combined-standard evaluators. The former also considered 25% more feasibility sub-criteria, evidencing a heightened focus of attention on evaluating the kinds of issues that limit a project's feasibility. Furthermore, we examined the effect of proposal novelty on these relationships. We find that combined-standard

[1] Whereas evaluators also judged the proposed impact of proposals, we focus on the relationships between novelty and feasibility given the primacy of novelty in evaluations of early-stage projects and positive association between novelty with impact documented in the literature (Azoulay, Fuchs, et al., 2019; Girotra et al., 2010; Katila & Ahuja, 2002).

evaluators perceive novelty and feasibility as *positively* related: their feasibility scores ranged from 9–13% higher for more novel projects compared to those of feasibility-only evaluators. Notably, this discrepancy was most significant among moderately novel proposals that balance conventional knowledge with some atypicality. Lastly, we investigated in post hoc analyses the variability in feasibility scores between feasibility-only and combined-standard evaluators. Our results show feasibility scores exhibit greater dispersion in feasibility-only compared to combined-standard evaluations. Further analysis suggests that this higher variability in scores is driven by projects at the right tail of the novelty distribution. These results are consistent with the notion that single-criterion evaluations of feasibility are more likely to reveal that novel projects may yield multiple possible paths to execution, which makes their ex-ante feasibility more uncertain (Garud & Rappa, 1994).

Our research suggests that novel early-stage projects may benefit from greater consideration of feasibility-only evaluations that potentially foster more in-depth, critical assessments of a project's potential for successful implementation and might prove beneficial in identifying highly novel projects for which many possible paths to viability may exist.

EVALUATIONS OF SCIENTIFIC PROJECTS AND JUDGMENTS OF NOVELTY AND FEASIBILITY

Schumpeter's theory of creative destruction highlighted the integral role of scientists and inventors in driving economic change (Schumpeter, 1942). Academic science has emerged as a crucial catalyst for economic growth, with discoveries from basic research fueling productivity, innovation, and the creation of new industries (Bikard & Marx, 2020; Fleming & Sorenson, 2004; Rosenberg & Nelson, 1994; Sauermann & Stephan, 2013). Scientific progress results not from a series of isolated experiments; rather, it follows an accretive pattern within a prevailing paradigm (Kuhn, 1977). The aim of new research is to extend the current state of knowledge beyond the existing scientific frontier, often balancing originality and tradition (Boudreau et al., 2016; Kuhn, 2003). Because successful knowledge creation in academic science is often a lengthy and uncertain process (Franzoni & Stephan, 2023; Lane et al., 2021; Scherer & Harhoff, 2000), it is crucial to identify any critical limitations associated with novel projects before carrying out, completing, and disseminating the work. Examples of critical obstacles that could hinder a project's feasibility might include limited resources, overambitious scope, unrealistic timelines, or challenging regulatory constraints (Guler, 2007; Sleesman et al., 2018).

The task of evaluating novel projects encompasses consideration of problem formulation as well as idea generation and implementation of the innovation process (Baer, 2012; Baer et al., 2013) and thus constitutes a crucial juncture in deciding which ideas to pursue and which to reject or defer (Azoulay & Li, 2020; Criscuolo et al., 2017; Lane et al., 2022). Accordingly, the *quality* of a new idea depends on both its novelty and its feasibility (Amabile, 1983; Boudreau et al., 2016; Diehl & Stroebe, 1987).

Standard evaluation formats in scientific peer review typically ask evaluators to distribute attention across multiple criteria, which often include novelty and impact as well as feasibility (Azoulay & Li, 2020; Lee, 2015). In single-criterion evaluation, in contrast, evaluators focus exclusively on evaluating a single aspect of the proposed work, such as feasibility. Feasibility assessments are used in a wide range of industries, such as real estate, manufacturing, energy, and healthcare, in which proposed projects may require significant investments of time, money, and resources (Justis & Kreigsmann, 1979).

In the remainder of this section, we examine how the format of evaluation criteria, whether combined-standard (with feasibility being one among several criteria) or feasibility-only (with feasibility being the sole focus), shapes evaluators' perceptions of feasibility.

The Role of Evaluation Format in Evaluations of Project Feasibility

We theorize that feasibility-only evaluators will produce lower feasibility scores than combined-standard evaluators. This concept can be best understood in light of Kahneman's (1973) theory of attentional strategies, which distinguishes between divided and focused attention. In the context of evaluation, combined-standard evaluators employ a divided attention strategy that involves either processing criteria sequentially or alternating between them (Kahneman, 1973; Ocasio, 2011). This approach allows for broader evaluation but may sacrifice depth, potentially overlooking details that require extra cognitive effort to recognize, understand, and interpret (Piezunka & Dahlander, 2015; Rhee & Leonardi, 2018; Sullivan, 2010). Single-criterion assessments, in contrast, adopt a focused attention strategy concentrated on one project attribute, thus avoiding expending perception and effort on any number of other attributes (Kahneman, 1973; Ocasio, 1997; Ocasio et al., 2020).

Due to the need to assess multiple criteria, such as novelty and feasibility, combined-standard evaluators assessing project quality may need to resort to heuristics or shortcuts to direct their attention to specific cues, while ignoring other, less salient information, to facilitate timely decision-making (Gigerenzer & Gaissmaier, 2011; Maitland & Sammartino, 2015). One heuristic that combined-standard evaluators may deploy during multi-criteria evaluations is to prioritize assessments of novelty over feasibility (Cooper, 1981; Nisbett & Wilson, 1977). Novelty, frequently characterized by its unexpectedness and statistical rarity (Ernst et al., 2020; Fantz, 1964; Johnston et al., 1990), tends to overshadow more familiar information in the evaluation process (Berlyne, 1954; Fiske, 1980; Haas et al., 2015; Taylor & Fiske, 1978). Novelty is often assessed based on abstract criteria that can vary according to personal interpretation, rather than relying on objective or quantifiable factors (Falchetti et al., 2022). This subjectivity may make evaluating novelty less cognitively demanding compared to assessing a project's feasibility, which typically requires more concrete, and quantifiable analysis (Gigerenzer & Gaissmaier, 2011).

Feasibility is a measurable feature of a project that can be decomposed into sub-criteria (Falchetti et al., 2022; Franzoni & Stephan, 2023), such as whether the proposed work can be accomplished within the given timeframe and the practicality of planned experiments and suitability of proposed methods (Luukkonen, 2012). Although evaluators may have differing perspectives on the feasibility of these sub-criteria, the mere fact that they exist and can be assessed makes the process of evaluating feasibility more demanding (Gigerenzer & Gaissmaier, 2011). Evaluators may utilize their first impressions of a project's novelty to inform their feasibility assessments (Cooper, 1981; Nisbett & Wilson, 1977). This practice may, however, lead them to conflate assessments of distinct criteria, potentially limiting the range of unique information considered in assessing a project's quality (Cooper, 1981; Sahoo et al., 2012).

Feasibility-only evaluations enable evaluators to attend exclusively to that aspect of a proposed project. When time and attention are devoted to feasibility rather than divided across multiple criteria, evaluators are more likely to be able to carefully scrutinize project details (Haas et al., 2015; Kahneman, 1973), taking into consideration a greater number of feasibility sub-criteria relative to their combined-standard counterparts. On one hand, greater emphasis on project feasibility may lead to more thorough assessments of feasibility sub-criteria and could potentially lead to higher feasibility scores if the projects are indeed more feasible. On the other hand, feasibility-only evaluators might scrutinize a project's feasibility aspects more critically, potentially leading to lower feasibility scores.

We propose three reasons why evaluators, in the context of academic science, might expend more effort assessing a project's potential limitations over its merits. The first relates to Robert Merton's argument that essential for the scientific method to function effectively is the norm of "organized skepticism," which emphasizes that new claims must be thoroughly tested and validated before acceptance (Merton, 1973). Indeed, evaluators notice more limitations when given more time to reassess a project's strengths and weaknesses (Lane et al., 2022).

The second reason is that evaluating prospective projects and completed work involves different challenges (Gross & Bergstrom, 2021). Whereas the emphasis in evaluating completed research is on the validity of findings and accuracy of interpretation, evaluators of prospective projects scrutinize proposed methods, resources, and achievability (Azoulay, Fuchs et al., 2019) and offer suggestions for enhancing the quality of the science (Franzoni & Stephan, 2023; Gallo et al., 2016). This difference in focus, together with the uncertainties inherent in proposed projects, suggests that feasibility-related issues may be more likely to be detected by evaluators employing a single-criterion evaluation format.

The third reason is that, from a statistical perspective, a heightened focus on the evaluation of feasibility may translate to closer scrutiny of sub-criteria that reveal limitations in a proposed project (Dahan & Mendelson, 2001). In other words, an evaluator specifically charged to weed out projects based solely on feasibility (Page, 2019) might be more likely to identify issues in the proposed work. Collectively, these reasons suggest the following hypothesis.

Hypothesis 1 (H1). *Feasibility-only evaluations will produce lower feasibility scores compared to combined-standard evaluations.*

The magnitude of the difference in feasibility scores suggested by the foregoing hypothesis likely reflects a project's degree of novelty. We next theorize how feasibility scores between combined-standard and feasibility-only evaluations vary based on project novelty.

Relationships between Evaluation Format, Project Novelty, and Feasibility

We begin by examining the perspective that combined-standard evaluators may view project novelty and feasibility as being inversely related. After that, we explore the alternative theory, suggesting that these two factors might, in fact, be positively related. In theorizing the two competing perspectives, we propose that, on average, combined-standard evaluators may perceive novelty and feasibility as positively related and hypothesize how this perception may affect the differences in scoring between combined-standard evaluators and feasibility-only evaluators.

Combined-standard evaluators' view of novelty and feasibility as inversely related implies that they perceive novel projects as less feasible. Novel breakthroughs in science and technology are often linked to higher risk-taking compared to more incremental innovations (Azoulay et al., 2011; Carson et al., 2022; Fleming, 2001; Krieger et al., 2022; Kuhn, 1977). However, their higher failure rates and longer times to yield results (Criscuolo et al., 2014) tend to be offset by the potential for considerable rewards (Manso, 2011; Park & Tzabbar, 2016). The tension between pursuing a safer, proven approach or a novel, riskier one (Bourdieu, 1975; March, 1991; Whitley, 2000) may result in evaluators perceiving novel projects as less feasible.

Novelty is often a recombinant process that can be assessed based on the degree that existing knowledge is recombined in unprecedented or original ways (Boudreau et al., 2016; Criscuolo et al., 2017; Uzzi et al., 2013; Weitzman, 1998). The recombination of existing ideas has given rise to many innovative discoveries (Henderson & Clark, 1990; Nelson & Winter, 1982; Weitzman, 1998) ranging from CRISPR-Cas9 gene editing to immunotherapy cancer treatments, but the frequency of failure tends to be higher for novel than for more conventional recombinations (Fleming, 2001; Scherer & Harhoff, 2000).

If combined-standard evaluators expect experimentation with new combinations to increase the variability of outcomes that determine failure or breakthrough (Ferguson & Carnabuci, 2017; Fleming, 2001; Rosenberg, 1998), then they may view incremental projects that build on existing knowledge to be more feasible than highly novel projects (Carson et al., 2022; Fleming, 2001; Tan, 2022). Incremental work typically incurs a lesser burden of proof than novel work that deploys a rare strategy (Foster et al., 2015), introduces an unconventional technique or line of inquiry (Muller, 1980), or makes unorthodox claims (Chai, 2017; Luukkonen, 2012), thereby generating a heightened sense of uncertainty in their feasibility (Fox & Tversky, 1995; Mueller et al., 2012).

Although these arguments suggest that combined-standard evaluators may employ a mental shortcut that posits novelty and feasibility as inversely related, we suggest that there are compelling arguments why combined-standard evaluators may perceive project novelty and feasibility as positively related, resulting in higher ratings for more novel projects.

As the primary function of academic science is the creation of new knowledge (Boudreau et al., 2016; Merton, 1973), we might expect combined-standard evaluators to demonstrate enthusiasm for novel projects over incremental ones. More specifically, evaluators for whom novelty fosters optimistic judgments of a project's potential might be prompted to overestimate favorable future outcomes and underestimate potential setbacks (Amore et al., 2021; Bracha & Brown, 2012; Kahneman & Lovallo, 1993; Weinstein, 1980). Such optimism can lead evaluators to misconceive the steps needed to achieve anticipated successful outcomes (Lant, 1992), resulting in a positive perception of novel projects independent of the likelihood of successful implementation (Amore et al., 2021). Novelty can spark a sense of potential and lower the threshold for what evaluators consider high-quality projects, namely, those that strike a balance between novelty and feasibility (Amabile et al., 1996; Diehl & Stroebe, 1987).

Moreover, novelty triggers both cognitive and affective responses (Falchetti et al., 2022; Rindova & Petkova, 2007) and attracts more attention than familiar information (Berlyne, 1954; Taylor & Fiske, 1978). These emotional responses, which are often independent of and precede cognitive reactions, can guide subsequent judgment (Zajonc, 1980), leading evaluators to prioritize the envisioned desired outcomes over potential execution issues (Falchetti et al., 2022). The promise of higher rewards might make the risks more acceptable, thereby increasing perceived feasibility. Indeed, prior work indicates that combined-standard evaluators view novelty and feasibility as positively related (Amabile et al., 1996; Girotra et al., 2010).

The inclination of combined-standard evaluators to downplay factors that might diminish a novel project's feasibility may be particularly critical for early-stage scientific projects compared to completed products, for which the research has already been implemented (Gross & Bergstrom, 2021). The feasibility of early-stage scientific projects depends on myriad factors including specialized equipment, worker knowledge sets, institutional support, relational contracting, the regulatory environment, and the quality of the research design needed to carry out the proposed work (Catalini et al., 2020; Dahlander & McFarland, 2013). Unlike evaluations of novelty, which highlight potential opportunities (Falchetti et al., 2022), assessments of feasibility aim to narrow alternatives based on available resources, techniques, and occurrences of natural phenomena (Franzoni & Stephan, 2023). These factors are more nuanced (Boudreau et al., 2016), and some feasibility sub-criteria may be unknowable at the time of the evaluation

(Maitland & Sammartino, 2015). As project feasibility issues only become more certain during the implementation stage (Baer, 2012), combined-standard evaluators might overlook potential obstacles to execution, believing that they can be resolved over time rather than raising them as concerns in the evaluation. This approach allows them to focus on the appeal of novel projects and their potential for valuable knowledge creation, rather than foreseeing the risks associated with their positive outcomes.

Taken together, these arguments suggest that the differences in feasibility scores between feasibility-only and combined-standard evaluators are likely to be larger for more novel projects and thus lead us to suggest the following hypothesis.

Hypothesis 2 (H2). *If feasibility-only evaluations are lower, on average, and combined-standard evaluators perceive project novelty and feasibility to be positively related, the magnitude of the difference in feasibility scores between the two groups will be larger for more novel projects.*

RESEARCH DESIGN

We describe in this section key aspects of our research design including setting, recruitment of evaluators, and treatment conditions.

Research Setting

We modified in collaboration with the translational science program at a large U.S. medical school the evaluation process for a translational research pilot grant competition. Translation is the "process of turning observations in the laboratory, clinic, and community into interventions that improve the health of individuals and the public—from diagnostics and therapeutics to medical procedures and behavioral changes."[2] We chose this setting for three reasons, (i) the

[2] https://ncats.nih.gov/translation.

medical school's desire to enhance its evaluation process, (ii) the significance of feasibility evaluations in translational medicine, and (iii) the vital role of evaluation in grant allocations for academic research.

The grant competition, focused on the "Five Senses and Perception," was open to any member with a university appointment pursuing an investigation related to human sensory systems in health or disease including, but not limited to, sight, sound, smell, taste, touch, and internal sensory systems. The competition received, between October 19, 2021, and November 10, 2021, 47 complete proposals covering a variety of topics from using biomarkers to diagnose dizziness to employing MRI to detect hidden hearing loss, and it awarded $50,000 to each of the top ten proposals for a total of $500,000 in one-year funding.

Evaluator Recruitment and Selection

To aid in evaluator recruitment, at the end of the competition the administrative team categorized the submitted proposals into six topic areas—sight; hearing; pain; smell/taste; touch; and proprioception/somatosensation—and, using a university-wide database and Medical Subject Headings (MeSH) terms, identified relevant researchers for each topic area. A subset of the top 100 most relevant evaluators within each topic area was randomly selected by the grant administrators and invited to participate. Ninety-seven accepted, resulting in 641 evaluator-proposal pairs across the six topics. Evaluators were randomly assigned to either the feasibility-only treatment (52 evaluators) or combined-standard control (45 evaluators) review condition. To allow multiple observations of evaluator behavior, each evaluated four to eight proposals within their area. On average, evaluators reviewed 6.47 proposals (s.d. = 1.93, min = 1, max = 9), and each proposal received a mean of 13.64 reviews (s.d. = 4.47, min = 7, max = 20).

Evaluator Instruction and Treatment

Evaluations were conducted online and triple-blinded: applicants were blinded to evaluators' identities, and evaluators to applicants' and each other's identities. Evaluators in the control condition conducted a standard, multi-criteria review assessing impact, novelty, and feasibility. The treatment condition, which evaluated only feasibility, was given the following prompt to signal the non-standard review process:

> [The grant organization] is interested in learning more about the scientific review process and is trying an alternative approach to reviewing proposals. For each of the proposals you have been assigned, we would like you to provide a focused review of their feasibility by answering the following questions. Please provide your review of the feasibility component in isolation, ignoring any considerations of the proposal's impact and innovation components.

The content and presentation of the feasibility criteria were identical in both the control and treatment conditions. Evaluators were first asked to assign a proposal an overall feasibility score (1= Highly unlikely or not feasible, 4 = Fully feasible) using the following definition of feasibility.

> Is the project feasible? Feasible is defined here as addressing the proposed specific aims within the year of support, by following the suggested research plan and using the proposed resources ($50K, direct costs, for 12 months).

As prior research connects overconfidence, risk-taking behavior, and willingness to undertake novel projects (Galasso & Simcoe, 2011; Hirshleifer et al., 2012), evaluators rated their confidence in their feasibility assessments (1 = Not confident at all, 4 = Completely confident).

Additionally, we examined the extent of consideration of different feasibility sub-criteria in evaluators' overall scores. Evaluators' feasibility assessments included in their scoring (1 = Not at all, 4 = To a great extent) the importance of *Experimental design*, *Proposed analysis plan*, *Proposed resources (includes access to samples/patients)*, *Alternative plan (unforeseen obstacles)*, *Timeframe (12 months)*, *Scope ($50K)*, and *Regulatory requirements*. These criteria, informed by NIH guidelines,[3] were accompanied by open-ended comments documenting evaluators' concerns about a project's feasibility.

Evaluators in the feasibility-only treatment cohort conducted a mean of 6.38 reviews (s.d. = 2.03, min = 1, max = 9) per proposal. Each proposal received a mean of 7.19 (s.d. = 2.62, min = 3, max = 12) feasibility-only treatment reviews per proposal, for a total of 338 evaluator-proposal pairs. Evaluators in the combined-standard control cohort conducted a mean of 6.59 reviews (s.d. = 1.82, min = 2, max = 8). Each proposal received a mean of 6.45 (s.d. = 2.22, min = 3, max = 10) combined-standard control reviews per proposal, for a total of 303 evaluator-proposal pairs.

Main Variables

Dependent variables. The main dependent variable, *Feasibility score*, is the overall feasibility score evaluators assigned to a proposal. Feasibility was assessed on a four-point scale (1 = Highly unlikely or not feasible, 2 = Possibly feasible, 3 = Likely feasible, and 4 = Fully feasible).

The extent to which the evaluators' overall feasibility scores considered different sub-criteria of feasibility was assessed using the variable, *Nb. of feasibility sub-criteria reviewed*, which counts the number of feasibility sub-criteria an evaluator considered *"To a great extent"* in terms of their contribution to a proposal's overall feasibility score.

Treatment variable. Our main independent variable, *Feasibility-only treatment*, is a dummy variable that corresponds to whether an evaluator was exogenously assigned to the feasibility-only treatment (equal to 1) or the combined-standard control condition (equal to 0).

Independent variables. Our main approach to measuring novelty is using the Medical Subject Headings (MeSH) term lexicon. Provided by the U.S. National Library of Medicine, MeSH

[3] See https://www.niaid.nih.gov/research/review-criteria.

terms represent unique biomedical concepts (e.g., humans, mice, diabetes mellitus, insulin) and are employed to index PubMed articles in the life sciences. These terms are systematically generated, resulting in less bias compared to author-selected keywords (Azoulay, Fons-Rosen, et al., 2019). We extracted MeSH terms from the abstracts using the MeSH on Demand analyzer,[4] which identifies MeSH terms in texts of up to 10,000 characters. Each abstract yielded an average of 12.90 MeSH terms (s.d. = 3.87, min = 6, max = 24). We identified proposals that were either *New-to-field* or *New-to-subfield* (Azoulay, Fons-Rosen, et al., 2019) by creating all possible pairs of MeSH terms for each proposal and determining how often the pairs had appeared in PubMed literature over the past decade. For *New-to-field* novelty, we calculated the percentage of MeSH term pairs not previously seen in the literature. For *New-to-subfield* novelty, we limited pairs to those that contained at least one term from the same subfield. Both measures were categorized into low, moderate, or high levels of novelty in three equal-sized intervals.

Other variables and controls. Our analysis strategy relied on the random assignment of evaluators to feasibility-only or combined-standard evaluation conditions and multiple observations per proposal and evaluator. That said, we also consider various evaluator-proposal and evaluator attributes. For each evaluator-proposal pair we considered two types of expertise: self-reported *Topic area expertise* (1 = Not at all familiar, 4 = Extremely familiar),[5] and *Intellectual similarity*, calculated by matching the MeSH terms in a proposal to the MeSH terms in evaluators' first- and last-authored publications. Each proposal had a mean of 12.79 (s.d. = 3.87, min = 6, max = 24) MeSH terms. We used Scopus to identify for each evaluator all first- and last-authored publications as of December 2021. In biomedicine, the first author is typically

[4] https://www.nlm.nih.gov/oet/ed/mesh/meshondemand.html.

[5] We did not collect evaluators' self-reported expertise in the topic area of pain (N = 40 evaluator-proposal pairs), instead imputing self-reported expertise in pain as the mean self-reported expertise in the other five topic areas (mean = 3.17).

the researcher who makes the most significant contribution to the research, such as acquiring or analyzing results or writing the manuscript. The last author tends to be the lead principal investigator (PI) who supervises and finances the work and is considered the main person responsible for the project (Pain et al., 2021). Each author had a mean of 74.02 (s.d. = 105.46) first- and last-authored publications, which corresponds to a mean of 231.39 (s.d. = 223.61) MeSH terms. We measured the *Intellectual similarity* between each evaluator-proposal pair by computing the cosine similarity, expressed as a percentage, between each proposal and reviewer's MeSH terms.

Additionally, we accounted for time-invariant unobserved evaluator and proposal characteristics. We controlled for status using *Early career* (= 1 if assistant professor, instructor, or other) and evaluators' confidence ratings in their feasibility assessment as well as various proposal characteristics, including counts of *Figures*, *Tables*, *References*, and *Abstract* word count. In some models, we substituted dummy variables for specific proposal characteristics.

Estimation Approach

We performed ordinary least squares (OLS) regressions to estimate the relationships between evaluators' *Feasibility scores* and the *Feasibility-only treatment* (H1) as well as the effect of *Proposal Novelty* (H2) on the estimated relationships. We tested H2 by adding the interaction term between *Feasibility-Only Treatment* x *Proposal Novelty* to the OLS models. We performed robustness checks with ordinal logit models, in which we model the dependent variable as a categorical variable with four ordered possible values. For regressions examining *Nb. of feasibility sub-criteria reviewed*, we used Poisson regression models to account for the discrete, non-negative integer values of the outcome variable.

RESULTS OF QUANTITATIVE ANALYSES

Table 1 presents the covariate balance checks for the 641 evaluator-proposal pairs exogenously exposed to either the feasibility-only treatment (N = 338) or combined-standard control (N = 303) evaluation. The covariates are not associated with the treatment vs. control assignments.

[Table 1 about here]

To test H1, which theorized lower scores from the feasibility-only than the combined-standard evaluations, we used OLS models to estimate the relationship between the feasibility scores and feasibility-only treatment (Table 2, Models 1–5), and Poisson models to estimate the number of feasibility sub-criteria reviewed and the feasibility-only treatment (Table 2, Models 6–10). Models 1 and 6 begin with the most basic comparison between the feasibility scores and feasibility-only treatment. Models 2 and 7 add evaluator and evaluator-proposal level attributes, and Models 3 and 8 add proposal characteristics. Models 4 and 9 add proposal dummies, and Models 5 and 10 include interaction terms between the feasibility-only treatment and evaluator and evaluator-proposal attributes.

Looking first at the relationships between feasibility scores and the feasibility-only treatment in Table 2, Model 1, we observe that the feasibility scores assigned by the feasibility-only treatment evaluators were 0.184 points lower, on average, than those assigned by the combined-standard control evaluators, corresponding to a 5.9 percent decrease in feasibility scores (Model 1: -0.184, $p = 0.004$). We observe that in Model 2 the coefficient for the *Feasibility-only treatment* remains stable and meaningful (Model 2: -0.199, $p < 0.001$). The coefficients for *Feasibility-only treatment* also remain consistent in Models 3 and 4 (Model 3: -0.201, $p < 0.001$; Model 4: -0.210, $p = 0.001$), which aligns with the experimental design that randomized evaluators into the combined-standard versus feasibility-only conditions. Model 5 indicates that the *Feasibility-only treatment* effect is stable across observed evaluator and evaluator-proposal attributes. Taken together, the estimated relationships in Table 2 associate the feasibility-only treatment condition with lower evaluator feasibility scores. Supplementary analyses confirm the reported results to be robust to ordinal logit models.

Turning next to the relationships between *Nb. of feasibility sub-criteria reviewed* and the feasibility-only treatment,[6] in Table 2, Model 6, we observe, on average, 26.1% more feasibility sub-criteria being reviewed in the feasibility-only treatment than in the combined-standard control condition (Model 6: 0.232, $p = 0.001$). These relationships remain consistent in Model 7, which adds the evaluator and evaluator-proposal covariates (Model 7: 0.210, $p = 0.001$), as well as in Models 8 and 9, which incorporate proposal characteristics (Model 8: 0.212, $p = 0.001$) and proposal dummies (Model 9: 0.219, $p < 0.001$). In Model 10, we observe the effect of the feasibility-only treatment on the count of feasibility sub-criteria to be strengthened among intellectually closer (Model 10: 0.0403, $p = 0.023$) and early-career-stage (Model 10: 0.254, $p = 0.067$) evaluators. Taken together, the estimated relationships in Table 2 indicate that the feasibility-only treatment caused evaluators to consider roughly 0.2 more feasibility sub-criteria in their assessments of proposals' overall feasibility scores.[7]

Overall, the estimated relationships in Table 2 suggest that the feasibility-only treatment is associated with both lower feasibility scores and greater consideration of additional feasibility sub-criteria. Thus, we find support for Hypothesis H1.

[Table 2 about here]

---

[6] We observed that the feasibility-only evaluators assigned greater weight to the consideration of the feasibility sub-criteria of experimental design, analysis plan, and timeframe of the proposed work, relative to combined-standard evaluators.

[7] Further investigating the use of novelty- and feasibility-related synonyms in the evaluators' qualitative comments documenting their concerns with a project's feasibility (N = 338 comments), we found mention of novelty-related words to be meaningfully more likely in the feasibility comments of combined-standard than feasibility-only evaluators (combined-standard, 0.111; N = 99; feasibility-only, 0.022; N = 186; z = 8.68; $p = 0.003$), but there was no difference in the respective evaluator's use of feasibility-related words (combined-standard, 0.212; N = 99; feasibility-only, 0.274; z = -1.01; $p = 0.315$).

To test H2, which theorized that combined-standard evaluators would perceive novelty and feasibility as positively related, in Table 3, we present the OLS models that correspond to proposals' *New-to-field* (Table 3, Models 1–5) and *New-to-subfield* (Table 3, Models 6–10) novelty. We begin by adding the proposal novelty score to Model 1 and Model 6, and then add, in Model 2 and Model 7, respectively, the interaction term between *Feasibility-only treatment* x *New-to-field novelty* and that between *Feasibility-only treatment* x *New-to-subfield novelty*. Models 3–5 and Models 8–10 add evaluator and evaluator-proposal covariates, proposal characteristics, and proposal dummies.

We observe in Model 1 that the coefficient for *Feasibility-only treatment* is negative and noteworthy after adding *New-to-field novelty* (Model 1: -0.188, $p = 0.003$). In Model 2, compared to the baseline category of *Feasibility-only treatment* x *Low new-to-field novelty*, we observe the coefficient for *Feasibility-only treatment* x *Moderate new-to-field novelty* to be negative and somewhat meaningful (Model 2: -0.264, $p = 0.085$), and the coefficient for *Feasibility-only treatment* x *High new-to-field novelty* to be directionally negative (Model 2: -0.119, $p = 0.445$). The estimated relationships remain robust in Model 3 (Moderate: -0.293, $p = 0.028$; High: -0.068, $p = 0.623$), and Model 4 (moderate: -0.293, $p = 0.029$; high = -0.067, $p = 0.630$), which add evaluator and evaluator-proposal covariates and proposal characteristics, respectively, as well as in our most stringent specification, in Model 5, which adds proposal dummies to allow for *within*-proposal comparisons (Model 5: -0.281, $p = 0.029$) and is directionally negative between *Feasibility-only treatment* x *High new-to-field novelty* (Model 5: -0.112, $p = 0.391$).

Figure 1A shows the margins plot with 95% CIs between *Feasibility score*, *Feasibility-only treatment*, and *New-to-field novelty* from Table 3, Model 4. We observe the feasibility-only treatment effect to be driven by proposals with a moderate level of new-to-field novelty, the average feasibility score in the control condition being 3.115 [2.992, 3.239] compared to 2.736 in the feasibility-only condition [2.608, 2.566], a meaningful difference of 0.379 points (which is a decrease of 12.2%).

Turning next to *New-to-subfield-novelty*, in Model 6, the coefficient for the *Feasibility-only treatment* is negative and meaningful (Model 6: -0.187, p = 0.003) and there is no meaningful relationship between *New-to-subfield novelty* and the feasibility scores. In Model 7, which adds the interaction term between *New-to-subfield novelty* and the *Feasibility-only treatment* condition, we observe a negative and meaningful relationship between *Feasibility-only treatment* x *Moderate new-to-subfield novelty* (Model 7: -0.440, p = 0.003) as well as a directionally negative relationship between *Feasibility-only treatment* x *High new-to-subfield novelty* (Model 7: -0.107, p = 0.488). The coefficients for the interaction terms suggest that the effect on feasibility scores of the *Feasibility-only treatment* condition is driven by proposals that are moderately novel relative to their subfield. The coefficient for the interaction term is roughly 1.7 times larger than the coefficient for the interaction term with a moderate degree of *New-to-field novelty* in Model 2. The reported relationships for the interaction term between *Feasibility-only treatment* x *Moderate new-to-subfield novelty* remain meaningful in Models 8–10, which add evaluator (Model 8: -0.328, p = 0.014) and evaluator-proposal covariates (Model 9: -0.332, p = 0.013) as well as proposal characteristics (Model 10: -0.369, p = 0.004).

Figure 1B presents the margins plot with 95% CIs between the *Feasibility score*, *Feasibility-only treatment*, and *New-to-field novelty* from Table 3, Model 9. Figure 1B shows the mean feasibility score for moderately novel proposals in their subfields to be 3.148 [3.000, 3.296] in the combined-standard control and 2.712 [2.569, 2.864] in the feasibility-only treatment condition, a noteworthy difference of 0.436 points (which is a decrease of 13.9%).

We perform several robustness checks in supplementary analyses. Our results in Table 3 are consistent with ordinal logit models and our findings are stable using four- and five-interval specifications of *New-to-field* and *New-to-subfield novelty*.

Taken together, the patterns reported in Table 3 indicate that for more innovative proposals, particularly those with moderate levels of novelty in their field or subfield, we observed lower feasibility scores in the feasibility-only treatment than in the combined-standard control condition. Hence, we find partial support for H2.

[Table 3 and Figure 1 about here]

IMPLICATIONS OF FEASIBILITY-ONLY EVALUATIONS FOR EVALUATION OUTCOMES

We next examined the impact of the feasibility-only treatment on the variation of feasibility scores and proposal rank orders. These analyses considered the effect of the feasibility-only treatment across all proposals and according to each proposal's level of *New-to-field novelty*.

Variation in Feasibility Scores by Experimental Condition

For each of the 47 proposals, we computed the standard deviations of feasibility scores by experimental condition, as illustrated in Figure 2A. Overall, the standard deviation of feasibility scores is seen to be marginally lower in combined-standard (mean = 0.643) than in feasibility-only (mean = 0.740) evaluations (paired t-test = -1.799, p = 0.078; N = 47 proposals). We then examined how the standard deviation in feasibility scores differs for low (Fig. 2B), moderate (Fig. 2C), and high (Fig. 2D) levels of new-to-field novelty. In Figures 2B and 2C, no difference is observed in the standard deviations of proposals for which new-to-field novelty is low (mean control = 0.664; mean treatment = 0.738; paired *t*-test = -0.758, *p* = 0.460; N = 16 proposals) or moderate (mean control = 0.657; mean treatment = 0.684; paired *t*-test = -0.282, *p* = 0.782; N = 16 proposals). The differences between the combined-standard and feasibility-only treatment conditions observed in Figure 2D are driven by proposals with high new-to-field novelty (mean control = 0.609; mean treatment = 0.795; paired *t*-test = -2.070, *p* = 0.056; N = 15 proposals).

[Figure 2 about here]

The observation that highly novel proposals elicit a wider range of feasibility scores under the feasibility-only treatment condition aligns with evidence that suggests that evaluators often express divergent views about the feasibility of novel projects that deviate from the current knowledge frontier (Boudreau et al., 2016; Garud & Rappa, 1994; Muller, 1980). Varying opinions may offer more potential paths (Baldwin & Clark, 2000; Leiponen & Helfat, 2010; Nelson, 1961), but their heightened uncertainty can make it challenging to predict the success or failure of any particular path (Garud & Rappa, 1994; Tushman & Anderson, 1986). The increased variability in feasibility scores among feasibility-only evaluators relative to those in the control condition, who also assessed proposal novelty, could explain why the feasibility scores between the two groups exhibited the greatest divergence for moderately novel proposals.

Rank Ordering of Proposals by Experimental Condition

To assess whether feasibility-only and combined-standard evaluations yielded different rankings based on proposals' merit scores, we compared the rank-order correlation between the actual and imputed merit scores. The grant administrators used the merit scores to rank the proposals and allocate funding decisions. The actual merit scores are derived from the sum of the combined-standard evaluators' mean novelty, impact, and feasibility scores for each proposal (i.e., the merit score is a sum of averages based on each criterion). The imputed merit scores substitute the mean feasibility scores of the combined-standard evaluators with those of the feasibility-only evaluators, keeping all other factors constant.

We subsequently rank-ordered the proposals from 1 (best) to 47 (worst) based on their actual and imputed merit scores. Figure 3 displays the scatter plot depicting the ranking of proposals based on the imputed (y-axis) and actual (x-axis) merit scores. The overall high correlation between the rankings derived from the imputed and actual merit scores ($\rho = 0.833$, $p < 0.001$) observed in Figure 3A is not surprising considering the high degree of correlation between the novelty and impact scores in the combined-standard control condition ($\rho = 0.794$).

Figures 3B–3D present the Spearman rank-order correlations organized by each proposal's *New-to-field novelty* scores. The correlation between the actual and imputed ranks remains high among less novel proposals, at $\rho = 0.967$ ($p < 0.001$), and drops to $\rho = 0.576$ ($p = 0.025$) for moderately novel, and jumps back up to $\rho = 0.817$ ($p < 0.001$) for highly novel proposals. The Spearman rank-order correlations indicate that the divergence between the combined-standard and feasibility-only scores meaningfully affects the ranking of the proposals.[8]

[Figure 3 about here]

DISCUSSION AND CONCLUSION

Evaluations of novel early-stage ideas are pivotal in assessing the feasibility of alternative projects given current resources, skills, and technology. Standard evaluation formats that amalgamate various criteria, such as a project's novelty and feasibility, can potentially introduce biases due to cognitive limitations related to expertise, time, and attention (Boudreau et al., 2016;

[8] Three, or 30%, of the top ten proposals eligible for funding based on their actual merit scores would have dropped out of eligibility for funding had the proposals been ranked instead by their imputed merit scores.

Criscuolo et al., 2017; Li et al., 2017; Stephan et al., 2017). Given that decision-makers often leverage evaluator judgments to allocate resources, such biases can exert a significant impact on the selection process for innovative projects. This paper explores specifically the degree to which standard evaluations that combine assessments of a project's novelty and feasibility may disregard obstacles that could potentially impede project implementation.

We investigated this research question by designing an intervention that exogenously varied the format of the evaluation to accommodate a comparison of feasibility assessments between combined-standard evaluators, who assess both project novelty and feasibility, and feasibility-only evaluators, whose focus is exclusively on assessing project feasibility. We chose the context of academic science, given that peer review lies at the heart of scientific discovery and advancement of knowledge (Lamont, 2009; Stephan, 2012), but analogous evaluations are used more broadly to steer firm strategy and the direction of innovation (Bian et al., 2021; Criscuolo et al., 2021).

We report a number of noteworthy patterns. First, we find that evaluations focused solely on feasibility tend to yield lower feasibility scores compared to multi-criteria, standard evaluation formats. Second, our findings indicate that this discrepancy is more pronounced for moderately novel projects that balance conventional knowledge with atypical combinations (Boudreau et al., 2016; Criscuolo et al., 2017; Uzzi et al., 2013). This pattern suggests that combined-standard evaluators may overlook the feasibility issues of novel projects.

Our study enriches understanding of perceived risk and reward in innovation (Ferguson & Carnabuci, 2017; Fleming, 2001; Khanna et al., 2016; Scherer & Harhoff, 2000) and underscores the critical importance of prospective evaluations in shaping perspectives on the potential success or failure of novel projects. Unlike previous studies that have relied mostly on retrospective analyses of innovative outcomes and concluded that novel projects bear greater risk and are more likely to fail, our approach has led us to propose that outcomes might be influenced by the evaluation process itself, single-criterion evaluation formats that focus on feasibility, for instance, affording better management and possibly even the elimination of some of the perceived risks and failures associated with novel projects.

We note that adopting alternative scoring schemes in standard evaluation formats—selecting projects with the fewest weaknesses or highest consensus in feasibility scores, for instance—may not effectively identify potential biases. This is largely due to evaluators' non-independent assessments of project novelty *and* feasibility (Cooper, 1981; Huang, 2018; Nisbett & Wilson, 1977). Assessments focused solely on feasibility may enable evaluators to discern subtler aspects of a project's feasibility likely to be surfaced only by thorough information processing and concentrated attention (Boudreau et al., 2016).

Our findings further reveal that the potential biases of combined-standard evaluation formats are likely to have different implications for evaluation outcomes for moderately and highly novel proposals. Our finding that the largest differences in feasibility scores occur among moderately novel proposals is consistent with previous research documenting evaluators' tendency to prefer moderately novel over highly novel projects (Boudreau et al., 2016; Chai & Menon, 2019; Criscuolo et al., 2017; Uzzi et al., 2013). Hence, feasibility-only assessments may be particularly relevant for assessing the shortfalls of moderately novel projects for which evaluators are more inclined to overlook potential feasibility concerns. For highly novel projects, the wider range of feasibility scores between feasibility-only and combined-standard evaluators may reflect the inherent uncertainty of such projects. Because highly novel projects often involve new or unique elements not necessarily amenable to direct extrapolation from prior experiences and expertise (Camerer & Johnson, 1991), evaluators' interpretations of their feasibility are likely to vary more widely (Garud & Rappa, 1994). Further contributing to more varied assessments of their feasibility is the tendency of highly novel projects to depart more substantially from the existing knowledge frontier (Kuhn, 1977; Teodoridis et al., 2019) and the consequent absence of clear precedents, scarcity of prior information and benchmarks, and difficulty finding comparable projects. These arguments are strengthened by the fact that evaluators focused solely on feasibility did not, on average, rate the most novel projects less feasible than low or moderately novel projects. Hence, it may be that the feasibility or path to execution of highly novel projects is not so much lower as more uncertain.

Relatedly, our results suggest that feasibility-only assessments may be a plausible mechanism for identifying highly novel proposals. In other words, to identify and reward highly novel research may require different regimes, such as alternative funding thresholds or rules (Azoulay, Fuchs et al., 2019; Franzoni & Stephan, 2023; Piezunka & Schilke, 2023), more time (Criscuolo et al., 2014), or stage-gate based funding (Franzoni & Stephan, 2023; Lerner, 1994; Tyebjee & Bruno, 1984; Zacharakis & Shepherd, 2001). The question then arises of how to reconcile these divergent viewpoints to improve the likelihood of implementation, independent of whether the sought-after outcome is likely to be realized (Franzoni & Stephan, 2023). One potential alternative for improving the odds of successfully completing the most novel initiatives is to offer applicants an opportunity to address feasibility critiques.

Although we have made a thorough and concerted effort to analyze how feasibility-only and combined-standard evaluations shape feasibility evaluations in science, our study admits some empirical limitations. First, whereas we focused on a relatively established field within the life sciences, namely, the Five Senses and Perception, it is important to examine the role of feasibility-only evaluations in other domains with and without substantial existing bodies of work from which to draw and incorporating varying levels of openness to unconventional theories, approaches, and conjectures (Azoulay, Fons-Rosen et al., 2019). Second, that our population of evaluators was drawn primarily from a highly selective medical school and affiliated institutions acknowledges a boundary around the institutional norms, networks, and incentives, including evaluators' perceptions of feasibility, under consideration. Third, although our study is centered on science, other settings, like venture capital investing (Bian et al., 2021; Huang, 2018) and new product development (Criscuolo et al., 2021; Katila & Ahuja, 2002; Leonard-Barton, 1992), also offer compelling opportunities for experimentation with alternative evaluation formats. In domains in which firm innovation tends to be driven by feasibility concerns, it may be that novelty-only evaluations may be more informative in filling the innovation pipeline's void of "missing novelty" (Ettlie, Bridges, & O'Keefe, 1984; Henderson & Clark, 1990; Krieger et al., 2022; Zhou & Wu, 2010; Zhou & Li, 2012).

We believe that the findings of the present research potentially pave the way to more informed selections of evaluation formats for identifying novel yet feasible projects with the potential to advance the knowledge frontier.

REFERENCES

Abernathy, W. J., & Clark, K. B. (1985). Innovation: Mapping the winds of creative destruction. *Research Policy*, *14*(1), 3–22.

Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology*, *43*(5), 997–1013.

Amabile, T. M. (1983). Brilliant but cruel: Perceptions of negative evaluators. *Journal of Experimental Social Psychology*, *19*(2), 146–156.

Amabile, T. M., Conti, R., Coon, H., Lazenby, J., & Herron, M. (1996). Assessing the work environment for creativity. *Academy of Management Journal*, *39*(5), 1154–1184.

Amore, M. D., Garofalo, O., & Martin-Sanchez, V. (2021). Failing to learn from failure: How optimism impedes entrepreneurial innovation. *Organization Science*, *32*(4), 940–964.

Åstebro, T., & Elhedhli, S. (2006). The effectiveness of simple decision heuristics: Forecasting commercial success for early-stage ventures. *Management Science*, *52*(3), 395–409.

Azoulay, P., Fons-Rosen, C., & Graff Zivin, J. S. (2019). Does science advance one funeral at a time? *American Economic Review*, *109*(8), 2889–2920.
Azoulay, P., Fuchs, E., Goldstein, A. P., & Kearney, M. (2019). Funding breakthrough research: Promises and challenges of the "ARPA Model." *Innovation Policy and the Economy*, *19*(1), 69–96.
Azoulay, P., Graff Zivin, J. S., & Manso, G. (2011). Incentives and creativity: Evidence from the academic life sciences. *RAND Journal of Economics*, *42*(3), 527–554.
Azoulay, P., & Li, D. (2020). *Scientific Grant Funding*. National Bureau of Economic Research.
Baer, M. (2012). Putting creativity to work: The implementation of creative ideas in organizations. *Academy of Management Journal*, *55*(5), 1102–1119.
Baer, M., Dirks, K. T., & Nickerson, J. A. (2013). Microfoundations of strategic problem formulation. *Strategic Management Journal*, *34*(2), 197–214.
Baldwin, C. Y., & Clark, K. B. (2000). *Design Rules: The Power of Modularity* (Vol. 1). MIT Press.
Baucells, M., & Heukamp, F. H. (2012). Probability and time trade-off. *Management Science*, *58*(4), 831–842.
Benner, M. J., & Tushman, M. L. (2003). Exploitation, exploration, and process management: The productivity dilemma revisited. *Academy of Management Review*, *28*(2), 238–256.
Berg, J. M. (2016). Balancing on the creative highwire: Forecasting the success of novel ideas in organizations. *Administrative Science Quarterly*, *61*(3), 433–468.
Berlyne, D. E. (1954). A theory of human curiosity. *British Journal of Psychology 45*(3), 180–191.
Bian, J., Greenberg, J., Li, J., & Wang, Y. (2021). Good to go first? Position effects in expert evaluation of early-stage ventures. *Management Science*, *68*(1), 300–315.
Bikard, M., & Marx, M. (2020). Bridging academia and industry: How geographic hubs connect university science and corporate technology. *Management Science*, *66*(8), 3425–3443.
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.
Bourdieu, P. (1975). The specificity of the scientific field and the social conditions of the progress of reason. *Social Science Information*, *14*(6), 19–47.
Bracha, A., & Brown, D. J. (2012). Affective decision making: A theory of optimism bias. *Games and Economic Behavior*, *75*(1), 67–80.
Briel, M., Olu, K. K., Von Elm, E., Kasenda, B., Alturki, R., Agarwal, A., Bhatnagar, N., & Schandelmaier, S. (2016). A systematic review of discontinued trials suggested that most reasons for recruitment failure were preventable. *Journal of Clinical Epidemiology*, *80*, 8–15.
Camerer, C. F., & Johnson, E. J. (1991). *The Process-Performance Paradox in Expert Judgment*. Cambridge, MA: Cambridge University Press.
Carson, R. T., Graff Zivin, J., Louviere, J. J., Sadoff, S., & Shrader, J. G. (2022). The Risk of Caution: Evidence from an Experiment. *Management Science*, *68*(12), 9042–9060.
Catalini, C., Fons-Rosen, C., & Gaulé, P. (2020). How do travel costs shape collaboration? *Management Science*, *66*(8), 3340–3360.
Chai, S. (2017). Near misses in the breakthrough discovery process. *Organization Science*, *28*(3), 411–428.

Chai, S., & Menon, A. (2019). Breakthrough recognition: Bias against novelty and competition for attention. *Research Policy*, *48*(3), 733–747.
Cole, S., & Simon, G. A. (1981). Chance and consensus in peer review. *Science*, *214*(4523), 881–886.
Cooper, W. H. (1981). Ubiquitous halo. *Psychological Bulletin*, *90*(2), 218.
Criscuolo, P., Dahlander, L., Grohsjean, T., & Salter, A. (2017). Evaluating novelty: The role of panels in the selection of R&D projects. *Academy of Management Journal*, *60*(2), 433–460.
Criscuolo, P., Dahlander, L., Grohsjean, T., & Salter, A. (2021). The sequence effect in panel decisions: Evidence from the evaluation of research and development projects. *Organization Science*, *32*(4), 987–1008.
Criscuolo, P., Salter, A., & Ter Wal, A. L. (2014). Going underground: Bootlegging and individual innovative performance. *Organization Science*, *25*(5), 1287–1305.
Dahan, E., & Mendelson, H. (2001). An extreme-value model of concept testing. *Management Science*, *47*(1), 102–116.
Dahlander, L., & McFarland, D. A. (2013). Ties that last: Tie formation and persistence in research collaborations over time. *Administrative Science Quarterly*, *58*(1), 69–110.
Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and Social Psychology*, *53*(3), 497.
Dougherty, D. (1992). A practice-centered model of organizational renewal through product innovation. *Strategic Management Journal*, *13*(S1), 77–92.
Ernst, D., Becker, S., & Horstmann, G. (2020). Novelty competes with saliency for attention. *Vision Research*, *168*, 42–52.
Ettlie, J. E., Bridges, W. P., & O'Keefe, R. D. (1984). Organization strategy and structural differences for radical versus incremental innovation. *Management Science*, *30*(6), 682–695.
Falchetti, D., Cattani, G., & Ferriani, S. (2022). Start with "Why," but only if you have to: The strategic framing of novel ideas across different audiences. *Strategic Management Journal*, *43*(1), 130–159.
Fantz, R. L. (1964). Visual experience in infants: Decreased attention to familiar patterns relative to novel ones. *Science*, *146*(3644), 668–670.
Ferguson, J.-P., & Carnabuci, G. (2017). Risky recombinations: Institutional gatekeeping in the innovation process. *Organization Science*, *28*(1), 133–151.
Fiske, S. T. (1980). Attention and weight in person perception: The impact of negative and extreme behavior. *Journal of Personality and Social Psychology*, *38*(6), 889.
Fleming, L. (2001). Recombinant uncertainty in technological search. *Management Science*, *47*(1), 117–132.
Fleming, L., & Sorenson, O. (2004). Science as a map in technological search. *Strategic Management Journal*, *25*(8–9), 909–928.
Fogel, D. B. (2018). Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review. *Contemporary Clinical Trials Communications*, *11*, 156–164.
Foster, J. G., Rzhetsky, A., & Evans, J. A. (2015). Tradition and innovation in scientists' research strategies. *American Sociological Review*, *80*(5), 875–908.
Fox, C. R., & Tversky, A. (1995). Ambiguity aversion and comparative ignorance. *Quarterly Journal of Economics*, *110*(3), 585–603.

Franzoni, C., & Stephan, P. (2023). Uncertainty and risk-taking in science: Meaning, measurement and management in peer review of research proposals. *Research Policy*, *Research Policy 52*(3), 104706.
Galasso, A., & Simcoe, T. S. (2011). CEO overconfidence and innovation. *Management Science*, *57*(8), 1469–1484.
Gallo, S. A., Sullivan, J. H., & Glisson, S. R. (2016). The influence of peer reviewer expertise on the evaluation of research funding applications. *PloS One*, *11*(10), e0165147.
Garud, R., & Rappa, M. A. (1994). A socio-cognitive model of technology evolution: The case of cochlear implants. *Organization Science*, *5*(3), 344–362.
Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*, 451–482.
Girotra, K., Terwiesch, C., & Ulrich, K. T. (2010). Idea generation and the quality of the best idea. *Management Science*, *56*(4), 591–605.
Gross, K., & Bergstrom, C. T. (2021). Why ex post peer review encourages high-risk research while ex ante review discourages it. *Proceedings of the National Academy of Sciences*, *118*(51), e2111615118.
Guler, I. (2007). Throwing good money after bad? Political and institutional influences on sequential decision making in the venture capital industry. *Administrative Science Quarterly*, *52*(2), 248–285.
Gupta, A. K., Smith, K. G., & Shalley, C. E. (2006). The interplay between exploration and exploitation. *Academy of Management Journal*, *49*(4), 693–706.
Haas, M. R., Criscuolo, P., & George, G. (2015). Which problems to solve? Online knowledge sharing and attention allocation in organizations. *Academy of Management Journal*, *58*(3), 680–711.
Haas, M. R., & Park, S. (2010). To share or not to share? Professional norms, reference groups, and information withholding among life scientists. *Organization Science*, *21*(4), 873–891.
Hasan, I., & Tucci, C. L. (2010). The innovation–economic growth nexus: Global evidence. *Research Policy*, *39*(10), 1264–1276.
Henderson, R., & Clark, K. B. (1990). Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. *Administrative Science Quarterly*, 9–30.
Hirshleifer, D., Low, A., & Teoh, S. H. (2012). Are overconfident CEOs better innovators? *Journal of Finance*, *67*(4), 1457–1498.
Hoetker, G. (2005). How much you know versus how well I know you: Selecting a supplier for a technically innovative component. *Strategic Management Journal*, *26*(1), 75–96.
Howell, J. M., & Higgins, C. A. (1990). Champions of technological innovation. *Administrative Science Quarterly*, 317–341.
Huang, L. (2018). The role of investor gut feel in managing complexity and extreme risk. *Academy of Management Journal*, *61*(5), 1821–1847.
Johnston, W. A., Hawley, K. J., Plewe, S. H., Elliott, J. M., & DeWitt, M. J. (1990). Attention capture by novel stimuli. *Journal of Experimental Psychology: General*, *119*(4), 397.
Justis, R. Y., & Kreigsmann, B. (1979). The feasibility study as a tool for venture analysis. *Journal of Small Business Management*, *17*(1), 35–42.
Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kahneman, D., & Lovallo, D. (1993). Timid choices and bold forecasts: A cognitive perspective on risk taking. *Management Science*, *39*(1), 17–31.
Kahneman, D., Slovic, S. P., & Tversky, A. (1982). *Judgment under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.
Katila, R., & Ahuja, G. (2002). Something old, something new: A longitudinal study of search behavior and new product introduction. *Academy of Management Journal*, *45*(6), 1183–1194.
Kerr, W. R., Nanda, R., & Rhodes-Kropf, M. (2014). Entrepreneurship as experimentation. *Journal of Economic Perspectives*, *28*(3), 25–48.
Khanna, R., Guler, I., & Nerkar, A. (2016). Fail often, fail big, and fail fast? Learning from small failures and R&D performance in the pharmaceutical industry. *Academy of Management Journal*, *59*(2), 436–459.
Knudsen, T., & Levinthal, D. A. (2007). Two faces of search: Alternative generation and alternative evaluation. *Organization Science*, *18*(1), 39–54.
Krieger, J. L., Li, D., & Papanikolaou, D. (2022). Missing novelty in drug development. *Review of Financial Studies*, *35*(2), 636–679.
Kuhn, T. S. (1977). *The Essential Tension: Selected Studies in Scientific Tradition and Change*. Chicago: University of Chicago Press.
Kuhn, T. S. (2003). Objectivity, value judgment, and theory choice. In A. Bird, & J. Ladyman (Eds.), *Arguing about Science* (pp. 74–86). New York: Routledge.
Lamont, M. (2009). *How Professors Think*. Cambridge, MA: Harvard University Press.
Lane, J., Ganguli, I., Gaule, P., Guinan, E., & Lakhani, K. R. (2021). Engineering serendipity: When does knowledge sharing lead to knowledge production? *Strategic Management Journal*, *42*(6).
Lane, J., Teplitskiy, M., Gray, G., Ranu, H., Menietti, M., Guinan, E., & Lakhani, K. R. (2022). Conservatism Gets Funded? A Field Experiment on the Role of Negative Information in Novel Project Evaluation. *Management Science*, *68*(6), 4478–4495.
Lant, T. K. (1992). Aspiration level adaptation: An empirical exploration. *Management Science*, *38*(5), 623–644.
Lee, C. J. (2015). Commensuration bias in peer review. *Philosophy of Science*, *82*(5), 1272–1283.
Lee, C. J., Sugimoto, C. R., Zhang, G., & Cronin, B. (2013). Bias in peer review. *Journal of the American Society for Information Science and Technology*, *64*(1), 2–17.
Leiponen, A., & Helfat, C. E. (2010). Innovation objectives, knowledge sources, and the benefits of breadth. *Strategic Management Journal*, *31*(2), 224–236.
Leonard-Barton, D. (1992). Core capabilities and core rigidities: A paradox in managing new product development. *Strategic Management Journal*, *13*(S1), 111–125.
Lerner, J. (1994). The syndication of venture capital investments. *Financial Management*, 16–27.
Li, D., Azoulay, P., & Sampat, B. N. (2017). The applied value of public investments in biomedical research. *Science*, *356*(6333), 78–81.
Luukkonen, T. (2012). Conservatism and risk-taking in peer review: Emerging ERC practices. *Research Evaluation*, *21*(1), 48–60.
Maitland, E., & Sammartino, A. (2015). Decision making and uncertainty: The role of heuristics and experience in assessing a politically hazardous environment. *Strategic Management Journal*, *36*(10), 1554–1578.

Manso, G. (2011). Motivating innovation. *Journal of Finance*, *66*(5), 1823–1860.
March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science*, *2*(1), 71–87.
Marsh, H. W., Jayasinghe, U. W., & Bond, N. W. (2008). Improving the peer-review process for grant applications: Reliability, validity, bias, and generalizability. *American Psychologist*, *63*(3), 160.
Merton, R. K. (1973). *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago: University of Chicago Press.
Mount, M. P., Baer, M., & Lupoli, M. J. (2021). Quantum leaps or baby steps? Expertise distance, construal level, and the propensity to invest in novel technological ideas. *Strategic Management Journal*, *42*(8), 1490–1515.
Mowery, D. C., & Rosenberg, N. (1999). *Paths of Innovation: Technological Change in 20th-Century America*. Cambridge: Cambridge University Press.
Mueller, J. S., Melwani, S., & Goncalo, J. A. (2012). The bias against creativity: Why people desire but reject creative ideas. *Psychological Science*, *23*(1), 13–17.
Muller, R. A. (1980). Innovation and scientific funding. *Science*, *209*(4459), 880–883.
Murphy, K. R., Jako, R. A., & Anhalt, R. L. (1993). Nature and consequences of halo error: A critical analysis. *Journal of Applied Psychology*, *78*(2), 218.
Nelson, R. R. (1961). Uncertainty, learning, and the economics of parallel research and development efforts. *Review of Economics and Statistics*, 351–364.
Nelson, R. R., & Winter, S. G. (1982). The Schumpeterian tradeoff revisited. *American Economic Review*, *72*(1), 114–132.
Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, *35*(4), 250.
Ocasio, W. (1997). Towards an attention-based view of the firm. *Strategic Management Journal*, *18*(S1), 187–206.
Ocasio, W. (2011). Attention to attention. *Organization Science*, *22*(5), 1286–1296.
Ocasio, W., Rhee, L., & Milner, D. (2020). Attention, knowledge, and organizational learning. In L. Argote, & J. M. Levine (Eds.), *The Oxford Handbook of Group and Organizational Learning* (pp. 81–94). Oxford: Oxford University Press.
Page, S. E. (2019). *The Diversity Bonus: How Great Teams Pay off in the Knowledge Economy*. Princeton: Princeton University Press.
Pain, E., Levine, A. S., Matias, J. N., Mikal, J., Lewis Jr, N. A., Gruber, J., Van Bavel, J. J., & Somerville, L. H. (2021). How to navigate authorship of scientific manuscripts. *Science*, *10*.
Park, H. D., & Tzabbar, D. (2016). Venture capital, CEOs' sources of power, and innovation novelty at different life stages of a new venture. *Organization Science*, *27*(2), 336–353.
Pier, E. L., Brauer, M., Filut, A., Kaatz, A., Raclaw, J., Nathan, M. J., Ford, C. E., & Carnes, M. (2018). Low agreement among reviewers evaluating the same NIH grant applications. *Proceedings of the National Academy of Sciences*, *115*(12), 2952–2957.
Piezunka, H., & Dahlander, L. (2015). Distant search, narrow attention: How crowding alters organizations' filtering of suggestions in crowdsourcing. *Academy of Management Journal*, *58*(3), 856–880.
Piezunka, H., & Schilke, O. (2023). The dual function of organizational structure: Aggregating and shaping individuals' votes. *Organization Science*.

Resch, K. I., Ernst, E., & Garrow, J. (2000). A randomized controlled study of reviewer bias against an unconventional therapy. *Journal of the Royal Society of Medicine*, *93*(4), 164-167.
Rhee, L., & Leonardi, P. M. (2018). Which pathway to good ideas? An attention-based view of innovation in social networks. *Strategic Management Journal*, *39*(4), 1188–1215.
Rindova, V. P., & Petkova, A. P. (2007). When is a new thing a good thing? Technological change, product form design, and perceptions of value for product innovations. *Organization Science*, *18*(2), 217–232.
Rosenberg, N. (1998). Uncertainty and technological change. In T. Siesfeld, J. Cefola, & D. Neef (Eds.), *The Economic Impact of Knowledge* (pp. 17–34). London: Routledge.
Rosenberg, N., & Nelson, R. R. (1994). American universities and technical advance in industry. *Research Policy*, *23*(3), 323–348.
Rosner, M. M. (1968). Economic determinants of organizational innovation. *Administrative Science Quarterly*, *12*(4), 614–625.
Sahoo, N., Krishnan, R., Duncan, G., & Callan, J. (2012). Research note—The halo effect in multicomponent ratings and its implications for recommender systems: The case of Yahoo! Movies. *Information Systems Research*, *23*(1), 231–246.
Sauermann, H., & Stephan, P. (2013). Conflicting logics? A multidimensional view of industrial and academic science. *Organization Science*, *24*(3), 889–909.
Scherer, F. M., & Harhoff, D. (2000). Technology policy for a world of skew-distributed outcomes. *Research Policy*, *29*(4–5), 559–566.
Schumpeter, J. A. (1942). *Capitalism, Socialism and Democracy*. Routledge.
Simon, H. A. (1991). Bounded rationality and organizational learning. *Organization Science*, *2*(1), 125–134.
Sleesman, D. J., Lennard, A. C., McNamara, G., & Conlon, D. E. (2018). Putting escalation of commitment in context: A multilevel review and analysis. *Academy of Management Annals*, *12*(1), 178–207.
Stephan, P. (2012). *How Economics Shapes Science*. Cambridge, MA: Harvard University Press.
Stephan, P., Veugelers, R., & Wang, J. (2017). Reviewers are blinkered by bibliometrics. *Nature*, *544*(7651), 411–412.
Sullivan, B. N. (2010). Competition and beyond: Problems and attention allocation in the organizational rulemaking process. *Organization Science*, *21*(2), 432–450.
Tan, D. (2022). The Road Not Taken: Technological Uncertainty and the Evaluation of Innovations. *Organization Science*, *34*(1), 156–175.
Taylor, S.E., & Fiske, S. T. (1978). Salience, attention, and attribution: Top of the head phenomena. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology*, vol. 11 (pp. 249–288). Academic Press.
Teodoridis, F., Bikard, M., & Vakili, K. (2019). Creativity at the knowledge frontier: The impact of specialization in fast-and slow-paced domains. *Administrative Science Quarterly*, *64*(4), 894–927.
Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, *4*(1), 25–29.
Tripsas, M., & Gavetti, G. (2000). Capabilities, cognition, and inertia: Evidence from digital imaging. *Strategic Management Journal*, *21*(10-11), 1147–1161.
Tushman, M. L., & Anderson, P. (1986). Technological discontinuities and organizational environments. *Administrative Science Quarterly*, 439–465.

Tyebjee, T. T., & Bruno, A. V. (1984). A model of venture capitalist investment activity. *Management Science*, *30*(9), 1051–1066.
Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical combinations and scientific impact. *Science*, *342*(6157), 468–472.
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.
Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology*, *39*(5), 806.
Weitzman, M. L. (1998). Recombinant growth. *Quarterly Journal of Economics*, *113*(2), 331–360.
Wells, J. D., Campbell, D. E., Valacich, J. S., & Featherman, M. (2010). The effect of perceived novelty on the adoption of information technology innovations: A risk/reward perspective. *Decision Sciences*, *41*(4), 813–843.
Whitley, R. (2000). *The Intellectual and Social Organization of the Sciences*. Oxford: Oxford University Press.
Zacharakis, A. L., & Shepherd, D. A. (2001). The nature of information and overconfidence on venture capitalists' decision making. *Journal of Business Venturing*, *16*(4), 311–332.
Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, *35*(2), 151–175.
Zhou, K. Z., & Li, C. B. (2012). How knowledge affects radical innovation: Knowledge base, market knowledge acquisition, and internal knowledge sharing. *Strategic Management Journal*, *33*(9), 1090–1102.
Zhou, K. Z., & Wu, F. (2010). Technological capability, strategic flexibility, and product innovation. *Strategic Management Journal*, *31*(5), 547–561.

Table 1. Covariate Balance Check for Feasibility-Only Treatment

Variable	Treatment (N = 338) Mean (s.d.)	Control (N = 303) Mean (s.d.)	Difference (two-tailed t-test)
Field novelty score	7.201 (0.412)	6.983 (0.414)	0.219, $p = 0.71$
Subfield novelty score	7.293 (0.492)	7.036 (0.515)	0.256, $p = 0.72$
Early career	0.76 (0.02)	0.75 (0.02)	0.020, $p = 0.61$
Topic area expertise	3.27 (0.05)	3.32 (0.05)	-0.056, $p = 0.44$
Intellectual similarity	4.05 (0.17)	4.22 (0.20)	-0.16, $p = 0.61$
Figures	2.65 (0.09)	2.60 (0.10)	0.050, $p = 0.71$
Tables	0.21 (0.03)	0.23 (0.04)	-0.02, $p = 0.72$
References	21.41 (0.40)	21.63 (0.43)	-0.22, $p = 0.71$
Abstract	234.96 (1.51)	236.07 (1.50)	-1.10, $p = 0.61$

Note: Novelty measures are reported as continuous variables.

Table 2. Regression Models of Feasibility Score and Nb. of Feasibility Sub-criteria for Feasibility-Only Treatment

	Dependent Variable: Feasibility Score					Dependent Variable: Nb. of Feasibility Sub-Criteria Reviewed				
VARIABLES	Model 1	Model 3	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10
Feasibility-only treatment	-0.184	-0.199	-0.201	-0.210	-0.476	0.232	0.210	0.212	0.219	0.0517
	(0.0628)	(0.0555)	(0.0556)	(0.0591)	(0.316)	(0.0704)	(0.0649)	(0.0647)	(0.0629)	(0.378)
Intellectual similarity		-0.0102	-0.0103	-0.0143	-0.0178		-0.00421	-0.00456	-0.00144	-0.0239
		(0.00898)	(0.00908)	(0.0117)	(0.0151)		(0.00823)	(0.00829)	(0.00996)	(0.0165)
Confidence		0.484	0.483	0.462	0.461		0.440	0.438	0.412	0.480
		(0.0366)	(0.0368)	(0.0354)	(0.0460)		(0.0470)	(0.0470)	(0.0479)	(0.0777)
Topic area expertise		-0.0641	-0.0692	-0.0811	-0.0984		-0.0763	-0.0611	-0.100	-0.128
		(0.0311)	(0.0318)	(0.0299)	(0.0424)		(0.0322)	(0.0340)	(0.0360)	(0.0568)
Early career		0.0223	0.0229	0.0285	-0.0289		-0.126	-0.144	-0.158	-0.281
		(0.0666)	(0.0670)	(0.0614)	(0.0739)		(0.0700)	(0.0697)	(0.0691)	(0.0987)
Figures			0.00423					-0.00876		
			(0.0166)					(0.0193)		
Tables			0.00960					0.00222		
			(0.0485)					(0.0487)		
References			-0.00116					0.00866		
			(0.00380)					(0.00384)		
Abstract			-0.00123					-0.00104		
			(0.000872)					(0.000981)		
Treatment x Intellect sim.					0.00723					0.0403
					(0.0206)					(0.0177)
Treatment x Confidence					-0.000553					-0.124
					(0.0794)					(0.0996)
Treatment x Topic exp.					0.0414					0.0626
					(0.0661)					(0.0674)
Treatment x Early career					0.134					0.254
					(0.107)					(0.139)
Constant	3.116	1.955	2.276	2.092	2.209	0.856	-0.105	-0.0560	-0.00496	0.0308
	(0.0435)	(0.150)	(0.274)	(0.157)	(0.175)	(0.0568)	(0.185)	(0.306)	(0.282)	(0.370)
Proposal FE	N	N	N	Y	Y	N	N	N	Y	Y
Observations	641	638	638	637	637	641	641	638	637	637
R-squared	0.013	0.240	0.242	0.238	0.241	--	--	--	--	--

Note: We use OLS regression models in Models 1–5 and Poisson regression models in Models 6–10. The number of observations drops to N = 638 in Models 2 and 7 due to three missing *Confidence* ratings, and to N = 637 in Models 4 and 9 due to one evaluator reviewing a single proposal dropping out of the sample. Robust standard errors in parentheses.

Table 3. OLS Models of Feasibility Scores for Feasibility-Only Treatment and Proposal Novelty Scores

	Dependent Variable: Feasibility Score									
		Proposal new-to-field novelty					*Proposal new-to-subfield novelty*			
VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10
Feasibility-only treatment	-0.188	-0.0587	-0.0833	-0.0852	-0.0779	-0.187	-0.0400	-0.0981	-0.0995	-0.0889
	(0.0628)	(0.111)	(0.0991)	(0.0989)	(0.0911)	(0.0628)	(0.0925)	(0.0814)	(0.0817)	(0.0763)
Moderate novelty	-0.146	-0.00921	0.0120	0.0196		-0.121	0.102	0.0328	0.0393	
	(0.0767)	(0.104)	(0.0896)	(0.0922)		(0.0754)	(0.102)	(0.0920)	(0.0960)	
High novelty	0.00243	0.0662	0.0539	0.0719		-0.00103	0.0541	0.0103	0.0298	
	(0.0783)	(0.112)	(0.0970)	(0.0985)		(0.0771)	(0.109)	(0.0981)	(0.101)	
Treatment x Mod novelty		-0.264	-0.294	-0.293	-0.281		-0.440	-0.328	-0.332	-0.369
		(0.153)	(0.133)	(0.134)	(0.126)		(0.149)	(0.134)	(0.133)	(0.127)
Treatment x High novelty		-0.119	-0.0681	-0.0666	-0.112		-0.107	-0.0605	-0.0580	-0.0782
		(0.156)	(0.139)	(0.138)	(0.130)		(0.154)	(0.135)	(0.137)	(0.127)
Intellectual similarity			-0.00959	-0.00921	-0.0138			-0.0105	-0.0101	-0.0122
			(0.00902)	(0.00913)	(0.0107)			(0.00904)	(0.00910)	(0.0107)
Confidence			0.488	0.488	0.464			0.479	0.479	0.454
			(0.0366)	(0.0369)	(0.0399)			(0.0366)	(0.0369)	(0.0398)
Topic area expertise			-0.0699	-0.0738	-0.0822			-0.0654	-0.0698	-0.0858
			(0.0307)	(0.0314)	(0.0330)			(0.0308)	(0.0315)	(0.0325)
Early career			0.00279	0.00321	0.0233			0.0272	0.0264	0.0356
			(0.0664)	(0.0668)	(0.0639)			(0.0671)	(0.0678)	(0.0640)
Figures				0.00478					-0.00485	
				(0.0166)					(0.0177)	
Tables				0.0347					0.0221	
				(0.0473)					(0.0460)	
References				-0.00237					-0.00244	
				(0.00396)					(0.00384)	
Abstract				-0.000829					-0.000934	
				(0.000934)					(0.000933)	
Constant	3.165	3.096	1.952	2.181	2.091	3.150	3.072	1.960	2.247	2.112
	(0.0641)	(0.0786)	(0.157)	(0.281)	(0.159)	(0.0554)	(0.0640)	(0.152)	(0.284)	(0.157)
Proposal FE	N	N	N	N	Y	N	N	N	N	Y
Observations	641	641	638	638	637	641	641	638	638	637
R-squared	0.021	0.025	0.254	0.255	0.353	0.018	0.030	0.252	0.254	0.357

Note: The number of observations drops to N = 638 in Models 3 and 8 due to 3 missing *Confidence* ratings, and to N = 637 in Models 5 and 10 due to one evaluator reviewing a single proposal dropping out of the sample. Robust standard errors in parentheses.

Figure 1. Margins Plot with 95% CIs of Feasibility Scores, Feasibility-Only Treatment, and Proposal Novelty

(A) New-to-field novelty (B) New-to-subfield novelty

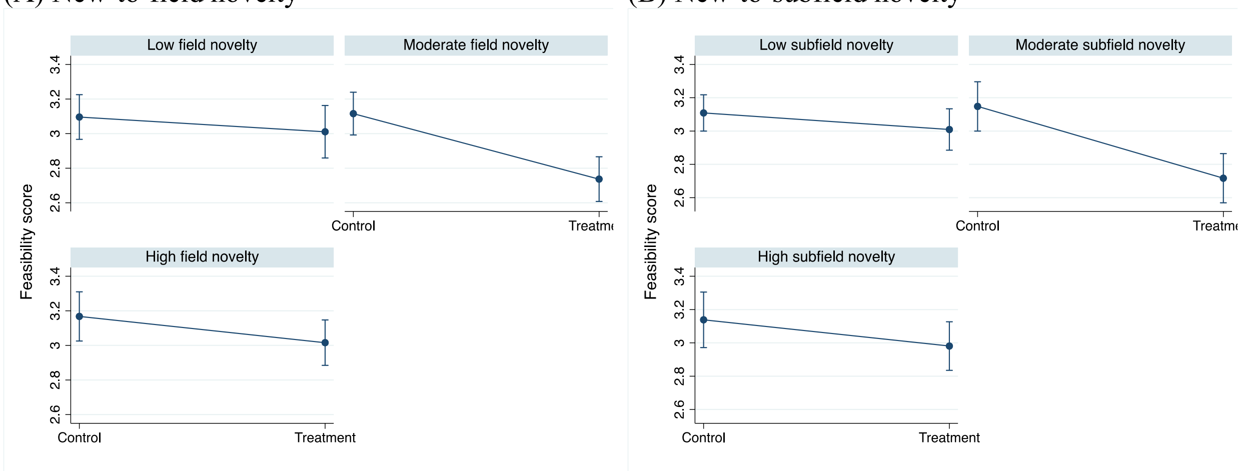

Figure 2. Standard Deviation of Proposal Scores for Feasibility-Only Treatment

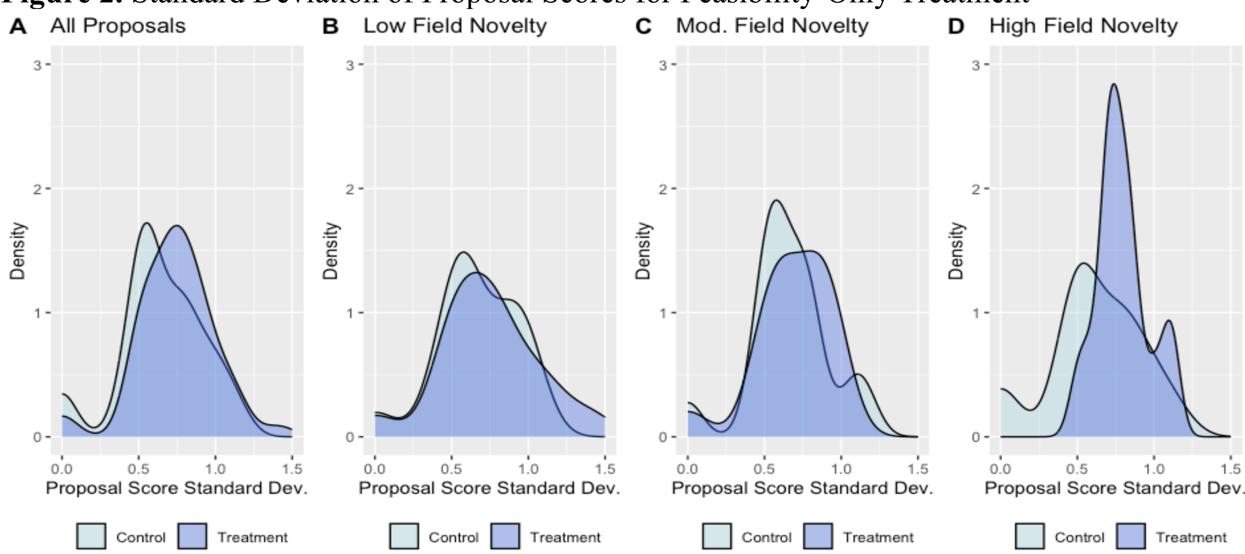

Figure 3. Comparison of Actual and Imputed Merit Score Rank Correlations