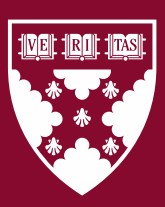
Working Paper 24-064

Greenlighting Innovative Projects: How Evaluation Format Shapes the Perceived Feasibility of Early-Stage Ideas

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



Harvard Business School

Greenlighting Innovative Projects: How Evaluation Format Shapes the Perceived Feasibility of Early-Stage Ideas

Jacqueline N. Lane*
Harvard Business School
D^3 Institute at Harvard

Simon Friis
Harvard Business School
D^3 Institute at Harvard

Tianxi Cai
Harvard TH Chan School of Public Health

Michael Menietti
D^3 Institute at Harvard

Griffin Weber
Harvard Medical School

Eva C. Guinan
Dana-Farber Cancer Institute

Working Paper 24-064

Copyright © 2024 and 2025 by Jacqueline N. Lane, Simon Friis, Tianxi Cai, Michael Menietti, Griffin Weber and Eva C. Guinan.

Working papers are in draft form. This working paper is distributed for purposes of comment and discussion only. It may not be reproduced without permission of the copyright holder. Copies of working papers are available from the author.

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

GREENLIGHTING INNOVATIVE PROJECTS:
HOW EVALUATION FORMAT SHAPES THE PERCEIVED FEASIBILITY OF EARLY-STAGE IDEAS

Jacqueline N. Lane[†*], Simon Friis[†], Tianxi Cai, Michael Menietti,

Griffin Weber and Eva C. Guinan

[†]Co-first authorship, [*]Corresponding author

Jacqueline N. Lane
Harvard Business School
D^3 Institute at Harvard
Boston, MA 02163
jnlane@hbs.edu

Simon Friis
Harvard Business School
D^3 Institute at Harvard
Boston, MA 02163
sfriis@hbs.edu

Tianxi Cai
Harvard T.H. Chan School of Public Health
Boston, MA 02115
tcai@hsph.harvard.edu

Michael Menietti
D^3 Institute at Harvard
Boston, MA 02163
mmenietti@hbs.edu

Griffin Weber
Harvard Medical School
Boston, MA 02115
griffin_weber@hms.harvard.edu

Eva C. Guinan
Dana-Farber Cancer Institute
Boston, MA 02215
eva_guinan@dfci@harvard.edu

GREENLIGHTING INNOVATIVE PROJECTS:
HOW EVALUATION FORMAT SHAPES THE PERCEIVED FEASIBILITY OF EARLY-STAGE IDEAS

ABSTRACT

The evaluation of innovative early-stage projects is essential for allocating limited resources. We investigate how the evaluation format affects the identification of feasibility issues through a field experiment at a leading research university. Experts were randomly assigned to multi-criteria (impact, novelty, feasibility) or single-criterion (feasibility-only) reviews of grant proposals. Feasibility-only reviews yielded lower scores, and substantial reordering of proposal rankings compared to multi-criteria evaluators. Analysis of qualitative comments reveals two mechanisms: focused attention enables discovery of critical flaws in feasibility components, while coherence scrutiny facilitates integrated assessment, exposing inconsistencies across components. Tracking subsequent publications shows that feasibility-only evaluations better predict research productivity than multi-criteria evaluations. These findings suggest value in combining approaches—using feasibility-only reviews to surface critical implementation barriers while leveraging multi-criteria processes to capture interdependencies, ensuring high-potential projects receive both a thorough feasibility and holistic evaluation.

Keywords: innovation evaluation, evaluation criteria, feasibility assessment, attention allocation, cognitive mechanisms, 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 potential impact of new projects, their novelty in advancing the field, and their feasibility given available resources, time constraints, and budgetary considerations (Berg, 2016; Boudreau et al., 2016; Criscuolo et al., 2017; Mount et al., 2021). While novel initiatives drive economic growth and scientific breakthroughs (Schumpeter, 1942; Uzzi et al., 2013) and open new paradigms and fields (Kuhn, 1977), they often face higher costs and failure rates (Fleming, 2001; Leiponen & Helfat, 2010).

Building on the notion that innovative initiatives often yield greater impact but fail more often, this paper investigates the effectiveness of different evaluation formats—specifically, multi-criteria (impact, novelty, feasibility) versus single-criterion (feasibility-only)—in identifying potential feasibility issues of early-stage projects. Here, we define impact as the extent to which a project can create significant value and influence key stakeholders across multiple domains, including markets, industries, and scientific fields (Kuhn, 1977; Uzzi et al., 2013). 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 extent to which the proposed project is achievable, i.e., can be implemented within a specific award's constraints (Franzoni & Stephan, 2023; Lane et al., 2022).

Organizations frequently rely on multi-criteria evaluation frameworks, using these to guide resource allocation decisions for early-stage projects (Bian et al., 2021; Boudreau et al.,

2016; Criscuolo et al., 2017; Knudsen & Levinthal, 2007). For instance, the U.S. National Institutes of Health (NIH)—the primary federal agency supporting biomedical and public health research—uses a multi-criteria framework to evaluate grant proposals based on impact, novelty, and feasibility (*NIH*, 2024).

Evaluating multiple criteria simultaneously can present opportunities and challenges for evaluators' attention. Decision-makers often face attentional constraints on their time (Ocasio, 1997), and the amount of attention on an issue depends on the number of competing demands on decision-makers' time and interrelationships (Eklund et al., 2025). While multi-criteria reviews may help evaluators perceive synergies between dimensions (Baker & Nelson, 2005), the divided attention may dilute focus on individual aspects of a project (Ocasio, 1997). For instance, a project's novel approach might improve feasibility by circumventing traditional implementation barriers, or its feasibility constraints might drive more novel solutions (Baker & Nelson, 2005; Prahalad & Hart, 2002).

The rationale for single-criterion feasibility reviews lies in the critical consequences of inadequate feasibility assessment in staged innovation processes (Azoulay & Li, 2020; Guler, 2007; Kerr et al., 2014). For instance, recent evidence from the pharmaceutical industry indicates that many late-stage failures could have been detected before significant resources were invested, 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 critical implementation issues can enable corrections or timely termination before substantial investment (Åstebro & Elhedhli, 2006; Khanna et al., 2016). By reducing the evaluation to a single criterion like feasibility, evaluators adopt a more focused analytical mindset, enabling a deeper examination of specific

implementation challenges (Joseph & Wilson, 2018). Yet, it is possible that removing specific dimensions from the evaluation may risk filtering out relevant signals about a proposal's bigger-picture implications.

While these evaluations aim to predict project outcomes, studying how evaluation formats affect assessments requires experimental manipulation of established review processes—an opportunity that rarely arises in practice. Hence, our understanding of high-risk or less feasible projects comes from retrospective studies of project outcomes after implementation (Fleming, 2001; Scherer & Harhoff, 2000; Wang et al., 2017) rather than prospective studies.

To address this gap, 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. This setting offered an ideal context to examine how evaluation formats shape assessments of implementation challenges for three reasons: it focused on knowledge creation through scientific peer review of early-stage projects; the medical school sought to enhance its evaluation procedures; and program administrators had observed that standard multi-criteria evaluations frequently failed to identify subtle yet critical feasibility issues in translational research proposals. We randomly assigned expert evaluators to either a "multi-criteria" condition (assessing impact, novelty, and feasibility) or a "single-criterion," feasibility-only condition. Our experiment involved 97 evaluators from seven institutions assessing 47 projects, yielding 641 proposal-evaluator pairs and $500,000 in funding, offering a unique window into resource allocation under contrasting evaluation structures.

We find that when evaluators focus solely on feasibility, they assign feasibility scores about 6–10 percentage points lower than those of multi-criteria evaluators, reflecting a more critical stance on potential implementation challenges. This narrower focus reorders which

proposals appear most feasible: seven of the top ten proposals under multi-criteria assessment drop from the top ten in the feasibility-only condition. Although both sets of evaluators rate the same criterion, the single-criterion approach changes how feasibility concerns are enumerated and prioritized, highlighting how the evaluation format affects resource allocation decisions.

Following an abductive approach (Bamberger, 2019; Behfar & Okhuysen, 2018), we identify plausible mechanisms that explain how the evaluation format shapes the absolute feasibility scores and the relative ranking of proposals. We conduct a hybrid human–LLM analysis of evaluators' qualitative comments, enabling us to examine how different formats shift cognitive attention and judgment. Drawing on the theoretical perspectives of attention allocation (Ocasio, 1997, 2011; Rhee & Leonardi, 2018) and heuristic decision models (Tversky, 1969; Payne, Bettman & Johnson, 1993; Simon, 1955), we propose two complementary mechanisms that could account for the patterns we observe: critical flaw discovery and coherence scrutiny. Critical flaw discovery represents an analytical process of disaggregating feasibility into "sub-criteria" or component parts, expanding the breadth of evaluation to uncover potential vulnerabilities. Coherence scrutiny involves integrating information across components to assess their compatibility and identify inconsistencies that might otherwise remain hidden. Together, these processes help explain why feasibility-only evaluations yield lower scores and meaningfully reorder which proposals appear most viable.

Lastly, we objectively validate feasibility judgments by matching the proposals to subsequent publications. We show that feasibility-only evaluators' assessments demonstrate trends consistent with greater predictive power of post-funding research productivity. This suggests that single-criterion feasibility evaluations capture meaningful implementation challenges that translate to real-world outcomes.

Our research contributes to the innovation evaluation literature by offering new insights into how evaluators allocate their limited time and attention when presented with either a single-criterion or multi-criteria review. Although a rich literature on strategic decision-making suggests that the organizational architecture structurally distributes managerial attention throughout the firm (Joseph & Ocasio, 2012; Joseph & Wilson, 2018; March & Simon, 1958), our research examines the micro-foundations of how attention allocation at the individual level affects evaluators' project assessments and resource mobilization. By systematically omitting specific dimensions from the evaluation process, we demonstrate that changing the rubric directly alters what evaluators notice, discuss, and ultimately value. Specifically, focusing solely on feasibility may allow for more in-depth scrutiny of that single criterion that may be especially warranted for high-potential, innovative projects. In contrast, multi-criteria reviews surface the broader interdependencies between alternative dimensions that ensure alignment with organizational goals or strategic priorities.

Furthermore, our findings have implications for the technology and innovation management literature. They suggest that the choice of evaluation criteria can shape which proposals, innovations, or organizational initiatives ultimately receive support. Shifting attention structures—by broadening or narrowing the criteria under consideration—can transform an organization or funding agency's innovation portfolio to align the format with its objectives for novel, high-impact undertakings.

EVALUATION OF INNOVATIVE PROJECTS AND MULTI-CRITERIA EVALUATION

Schumpeter's theory of creative destruction highlights the integral role of scientists and inventors in driving economic change (Schumpeter, 1942). Technological and scientific innovation 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; Teodoridis, Bikard, & Vakili, 2019). Technical and scientific progress results not from a series of isolated experiments but follows an accretive pattern within a prevailing paradigm (Kuhn, 1977). New research aims 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 is usually 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 early-stage projects before making costly investments in the work. 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 early-stage 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 early-stage projects is typically assessed along three main dimensions: impact, novelty, and feasibility (Boudreau et al., 2016; Mount et al. 2021).

Organizations employ various evaluation formats to assess these dimensions. Standard formats in scientific peer review typically ask evaluators to distribute attention across multiple criteria simultaneously (Azoulay & Li, 2020), while single-criterion formats focus evaluator attention exclusively on one dimension, such as feasibility. Organizations often modify these evaluation formats over time, adjusting various aspects of the review process to improve decision-making. For instance, the NIH recently simplified its evaluation criteria from five to

three, exemplifying how organizations may revise their evaluation frameworks.[1] However, we lack an understanding of how such modifications affect evaluators' assessments.

How Attention Shapes Evaluation

Organizations rely heavily on evaluation processes to make resource allocation decisions, yet these processes involve significant complexity and uncertainty, particularly when assessing early-stage projects. Evaluation is inherently a cognitive process, requiring decision-makers to notice, encode, and interpret information while focusing their limited attention (Ocasio, 1997). In this context, attention represents a scarce resource constrained by bounded rationality (Simon, 1955; March & Simon, 1958), which limits the extent to which evaluators can consider all criteria, trade-offs, and potential outcomes simultaneously.

The allocation of attention has been extensively studied at the firm level, particularly regarding how structural elements of organizational architecture distribute managerial attention throughout the firm (Joseph & Ocasio, 2012; Ocasio, 1997). This literature has illuminated the consequences of specialized versus broad attention in shaping how managers process novel opportunities and how their subsequent actions influence the firm's strategic direction and innovation (Eklund et al., 2025; Rhee & Leonardi, 2018; Tripsas & Gavetti, 2000).

While this body of work has provided important insights into the strategic implications of attention allocation at the firm level, relatively little is known about the micro-foundations of resource allocation at the project level, specifically how evaluation formats influence individual evaluators' attention allocation to alternative criteria. Evaluation rubrics—by presenting specific criteria such as feasibility, novelty, or impact—serve as cognitive scaffolds that direct

[1] https://grants.nih.gov/policy-and-compliance/policy-topics/peer-review/simplifying-review/framework

evaluators' attention toward specific dimensions of a project, potentially altering their prioritization and scrutiny of information (Hsee et al., 1999; Posner & Petersen, 1990). This shift in attention can significantly impact the depth with which evaluators assess some aspects while deprioritizing others (Eklund et al., 2025; Joseph & Wilson, 2018).

Single-criterion feasibility reviews represent an example of attentional specialization, where evaluators focus exclusively on feasibility. This format concentrates evaluators' cognitive resources on a single dimension, potentially intensifying their scrutiny of feasibility sub-criteria such as timelines, resource requirements, and methodological considerations (Levinthal & March, 1993). However, such specialization may come at the cost of neglecting critical interdependencies between feasibility and other dimensions, such as novelty or potential impact.

Despite the importance of these issues, the ways in which multi-criteria and single-criterion evaluation formats influence how feasibility is conceptualized, prioritized, and assessed remain insufficiently understood. Existing research has focused mainly on higher-level decision-making processes and lacks empirical insight into how attentional trade-offs at the project level affect selection decisions. This gap is particularly relevant in the context of early-stage innovation projects, where feasibility assessments are critical to identifying viable opportunities and minimizing downstream risks. In the following sections, we examine these questions through a field experiment comparing multi-criteria and feasibility-only reviews, investigating how the evaluation format shapes the depth and content of feasibility assessments.

RESEARCH SETTING AND DESIGN

We collaborated with the translational science program at a large U.S. medical school to intervene in the evaluation process for a research pilot grant competition. Translation is the "process of turning laboratory, clinic, and community observations 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 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 was open to any member with a university appointment pursuing an investigation related to human sensory systems in health or disease. Between October 19, 2021, and November 10, 2021, the competition received 47 complete proposals covering a variety of topics and awarded $50,000 to each of the top 10 proposals, totaling $500,000 for one year of 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 potential evaluators for each topic area. A subset of the top 100 most relevant evaluators within each topic area was randomly selected based on MeSH term search and invited to participate. In total, 97 evaluators accepted, resulting in 641 evaluator-proposal pairs across the six topics. Evaluators were randomly assigned to either the feasibility-only treatment (52 evaluators) or the multi-criteria control (45 evaluators) review condition. Each evaluator reviewed four to eight proposals within their area to allow multiple observations of evaluator behavior. Each proposal was reviewed by multiple evaluators in both conditions, ensuring that every proposal received independent assessments from evaluators assigned to both

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

the multi-criteria and feasibility-only evaluation formats. 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

We implemented a triple-blind review process where applicants were blinded to evaluators' identities, and evaluators were blinded to applicants' and each other's identities. All evaluations were conducted online using identical formatting and presentation. The full evaluation materials are provided in Online Appendix E.

Multi-Criteria Control Condition

Evaluators in the control condition conducted a multi-criteria review assessing three dimensions, each rated on a four-point scale:

1. *Impact* – is defined as "providing increased understanding of human sensory systems with potential translational benefit to human health assessed by patients, physicians, institutions, or society."
2. *Novelty* – is defined as "likely to lead to new technology or knowledge, an unanticipated application of existing technology, or a novel approach that enhances an established modality."
3. *Feasibility* – is defined as "addressing the proposed specific aims within the year of support using the proposed resources ($50K, direct costs, for 12 months)."

After assigning overall scores for each dimension, multi-criteria evaluators rated specific sub-criteria (i.e., components) related to novelty, impact, and feasibility, as indicated in Table 1. They also indicated their confidence in their feasibility ratings and provided open-ended comments on feasibility concerns and potential improvements.

[Table 1 about here]

Feasibility-Only Treatment Condition

The feasibility-only treatment condition differed from the multi-criteria control in two ways: evaluators received tailored instructions to assess feasibility alone, and only the feasibility rubric was displayed during their review. All other elements of the online review system—including proposal content, formatting, sequence, and platform functionality—were held constant across both conditions. Specifically, feasibility-only evaluators were instructed to focus solely on feasibility before beginning their reviews:

> *"[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."*

Evaluators in the feasibility-only treatment condition 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 multi-criteria control condition 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) multi-criteria control reviews per proposal for 303 evaluator-proposal pairs.

Main Variables

Dependent variables. The primary dependent variable, the *Feasibility score*, is the overall feasibility score evaluators assign 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).

Treatment variable. Our primary 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 multi-criteria control (equal to 0) condition.

Independent variables. Although random assignment mitigates the need for controls, we included several variables to further investigate potential drivers of feasibility assessments. First, we measured evaluator topic area expertise based on the evaluator's self-reported familiarity with the sensory domain they were assigned to review on a scale of 1 ("Not at all familiar") to 4 ("Extremely familiar") and computed intellectual similarity by matching the proposal's MeSH terms (i.e., standardized biomedical indexing concepts used by the U.S. National Library of Medicine) to the MeSH terms in evaluators' first- and last-authored publications, yielding a cosine similarity score. We also tracked early career status (coded 1 for assistant professor or equivalent) and proposal characteristics (e.g., counts of figures, tables, references, and abstract word count). Additional procedural details concerning these variables are provided in the Online Appendix A.

Estimation Approach

We performed ordinary least squares (OLS) regressions to estimate the relationships between evaluators' *Feasibility scores* and the *Feasibility-only treatment* to investigate the effect of the feasibility-only evaluation format on the assigned feasibility scores. Our simplest model regresses *Feasibility scores* on the *Feasibility-only treatment* for each evaluator-proposal pair $\{i, j\}$. We then added evaluator covariates, X_i, (*Early career*, *Topic area expertise*), evaluator-proposal attributes (*Intellectual similarity*), ξ_{ij}, and proposal attributes (*Figures*, *Tables*,

References, Abstract), γ_{ij}. The main model relating feasibility scores to the feasibility-only treatment is presented in Equation (1).

$$Feasibility\ scores_{ij} = \beta_0 + \beta_1 Feasibility_Only\ Treatment_{ij} + \beta_3 X_i + \beta_4 \xi_{ij} + \beta_5 \gamma_{ij} + \varepsilon_{ij}. \quad (1)$$

In certain model specifications, we added proposal fixed effects for all evaluators who reviewed more than one proposal. We performed robustness checks for equation (1) using ordinal logit models, in which we model the dependent variable as a categorical variable with four ordered possible values.

QUANTITATIVE RESULTS FROM FIELD EXPERIMENT

Covariate balance checks indicate that the treatment and control assignments were successfully randomized, with no meaningful associations between the assignment and observed covariates (detailed summary statistics and balance checks are available in Online Appendix B1).

Main Effect: Impact of Feasibility-Only Treatment on Feasibility Scores

In Table 2, we use OLS models to estimate the average treatment effect of the feasibility-only evaluation format on feasibility scores (Table 2). In Table 2, Model 1 begins with the most basic comparison between the feasibility scores and feasibility-only treatment. Model 2 adds evaluator and evaluator-proposal level attributes, Model 3 adds proposal characteristics, and Model 4 adds proposal fixed effects.

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 multi-criteria 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: -0.186, $p = 0.003$). The coefficients for *Feasibility-only treatment* also remain consistent in Models 3 and 4 (Model 3: -0.188, $p = 0.003$; Model 4: -0.197, $p = 0.001$) and align with the randomization of evaluators into the multi-criteria versus feasibility-only conditions. The estimated relationships in Table 2 indicate that the feasibility-only treatment condition leads to lower feasibility scores. Supplementary analyses in Online Appendix B2 confirm that these results are robust to ordinal logit models.

[Table 2 about here]

The results in Table 2 indicate that when given a single-criterion review of feasibility, evaluators assign systematically lower feasibility scores than in a multi-criteria review, suggesting that they are more critical of the project's feasibility.

Impact of Feasibility-Only Treatment on Relative Rankings

Next, we investigate whether these format-induced differences in feasibility scores also produce divergent proposal rankings. Figure 1 illustrates that shifting from a multi-criterion to a feasibility-only review substantially changes proposals' relative standings. A relatively low Spearman rank-order correlation of $\rho = 0.222$ ($p = 0.132$) for feasibility scores in the two formats suggests that evaluation formats do more than merely shift absolute scores: they fundamentally reshape evaluators' judgments about the relative feasibility of projects.

[Figure 1 about here]

In sum, the results in Table 2 and Figure 1 collectively demonstrate that asking evaluators to focus exclusively on feasibility induces both a "feasibility penalty" (scores roughly 6–10 percentage points lower) and a distinct reordering of which proposals appear most feasible.

This divergence has critical real-world implications for funding decisions. In this pilot grant competition, there is a 70% turnover in the highest-rated proposals when comparing the top ten proposals rated by feasibility scores under the feasibility-only and multi-criteria formats.

Impact of Feasibility-Only Treatment on Funded Proposals

Figure 2 compares the cumulative mean scores for feasibility (left), impact (center), and novelty (right) across proposals ranked by feasibility under each experimental condition. This visualization reveals how different evaluation formats fundamentally alter which projects receive funding. To simulate real-world consequences, we examine what would happen if the organization funded each evaluation format's top ten proposals (vertical dashed line). Crucially, all scores come from multi-criteria reviewers' assessments, creating a consistent benchmark regardless of which evaluation format ranked the proposals. This approach isolates how evaluation structure affects selection outcomes without introducing rating scale differences.

[Figure 2 around here]

Our analyses reveal two distinct patterns. First, proposals deemed most feasible under the feasibility-only format score lower across all dimensions—even feasibility itself—when judged by multi-criteria standards. This counterintuitive finding demonstrates how evaluation format alone can transform the perceived quality profile of funded projects.

Second, the relationship between feasibility and high-reward attributes reverses between conditions. In multi-criteria evaluation, feasibility positively correlates with novelty and impact (blue lines slope downward together). However, in feasibility-only reviews, this relationship inverts—projects rated highest on feasibility show the lowest novelty and impact scores (yellow lines slope upward). This reveals a critical trade-off: single-criterion evaluation may

systematically filter out high-potential innovations while being particularly stringent toward projects with transformative possibilities.

MECHANISM ASSESSMENT

Our empirical results suggest that feasibility evaluations may operate within a predominantly non-compensatory framework in this context. Prior research has established that technical assessments often function through non-compensatory processes (Åstebro & Elhedhli, 2006; Lane et al., 2022; Sleesman et al., 2018), particularly in contexts where implementation failure can be costly (Khanna et al., 2016). Under such a 'weakest link' logic, critical sub-criteria must generally be satisfied and coherent; failure in one area is difficult to offset by strengths in another (Holyoak & Simon, 1999; Payne, Bettman & Johnson, 1993). By contrast, novelty and impact evaluations may function within a more compensatory framework, where strong performance in one sub-dimension can redeem weaknesses in another.

The allocation of attention—whether focused on a single criterion or divided across multiple criteria—interacts with these evaluation frameworks to influence how proposals are assessed (Ocasio, 1997). In our study, evaluators in the multi-criteria condition operate under divided attention, simultaneously weighing feasibility, novelty, and impact. Conversely, evaluators in the feasibility-only condition can directly focus attention exclusively on feasibility considerations.

Drawing on these concepts, we propose two complementary mechanisms through which focused attention within a non-compensatory feasibility framework might influence judgment: (1) critical flaw discovery and (2) coherence scrutiny. If these mechanisms are operating, they will produce lower feasibility scores, and a systematic reordering of the top-rated proposals based on different evaluation patterns.

Critical Flaw Discovery: Expanding the Breadth of Evaluation

Critical flaw discovery represents an analytical process of disaggregating feasibility into component parts for individual assessment. When evaluators focus exclusively on feasibility, they free up cognitive bandwidth that would otherwise be divided across multiple criteria, namely feasibility, impact, and novelty. This additional bandwidth allows them to unpack feasibility into multiple sub-criteria, such as budget, technical feasibility, regulatory approvals, and staffing requirements.

This expanded analytical scope systematically elevates the probability of identifying disqualifying constraints, thereby producing lower aggregate feasibility scores. In settings where feasibility is assessed non-compensatorily, identifying even a single serious shortcoming can lower a project's overall feasibility score—there is no counterbalancing strength to offset it. As evaluators examine more potential failure points, they systematically increase the likelihood of encountering a disqualifying flaw. This breadth-oriented mechanism helps explain the lower average feasibility scores observed under feasibility-only evaluation: not because the proposals are weaker, but because broader scrutiny reveals more potential barriers that must each be cleared.

Coherence Scrutiny: Deepening the Integration of Analysis

Coherence scrutiny represents an interdependent process of integrating information across components to assess their compatibility. While critical flaw discovery expands evaluation horizontally across more sub-criteria, coherence scrutiny deepens the vertical integration of these elements. Focused attention provides the cognitive bandwidth necessary to synthesize information across multiple aspects of a proposal, identifying potential inconsistencies or structural misalignments that remain obscure when evaluators' attention is divided across

multiple criteria. For example, an ambitious timeline may conflict with a complex analysis plan, or a large enrollment target may be unsupported by the proposed staffing and budget. Such integrative judgment is combinatorially complex: as the number of sub-criteria increases, the number of potential pairwise (and higher-order) inconsistencies grows rapidly. Detecting these misalignments requires evaluating each sub-criterion individually and reasoning about how they interact—something that becomes feasible only when attention is not fragmented across multiple criteria.

In a non-compensatory regime, these coherence failures are especially consequential: a mismatch between sub-criteria can serve as the critical flaw that undermines the entire proposal. Because these failures are proposal-specific and not systematically related to the surface-level impressions formed under divided attention, their detection can significantly reorder rankings. Some proposals drop sharply in ranking when hidden inconsistencies are surfaced; others rise by withstanding deeper scrutiny. This helps explain why feasibility-only evaluation reorders feasibility proposal rankings, not just lowers feasibility scores.

Scope Conditions and Empirical Assessment

These mechanisms of critical flaw discovery and coherence scrutiny are likely to operate in other settings where (a) evaluation involves criteria that function predominantly in a non-compensatory manner and (b) evaluators' attention is focused on a single criterion rather than divided across multiple criteria. If our proposed mechanisms are valid, they should generate distinct observable patterns in evaluator behavior: critical flaw discovery would manifest as greater breadth in the types of feasibility issues discussed, while coherence scrutiny would be evidenced by more frequent references to relationships between different feasibility components.

The following section tests these proposed mechanisms by analyzing evaluators' open-ended comments using natural language processing techniques. This analysis will help us determine whether evaluators in the feasibility-only condition demonstrate greater attention to feasibility sub-criteria and more intensive assessment of internal coherence compared to their multi-criteria counterparts.

Human-LLM Content Coding Analysis of Qualitative Comments

We analyze evaluators' open-ended comments using a hybrid human-LLM coding approach to gain deeper insight into how the evaluation format shapes the conceptualization of feasibility. This approach begins with human coding to identify relevant codes and then scaling the analysis using OpenAI's GPT-4o —a state-of-the-art large language model capable of analyzing and coding text. This approach allowed us to capture how evaluators conceptualized feasibility in terms of interdependencies among sub-criteria and with novelty and impact.

Data and Context

All evaluators were allowed to leave optional comments in an open-text box after assigning their evaluation scores. The instructions prompted evaluators to describe anticipated feasibility issues ("If you have concerns about project feasibility, please provide brief comments:"). Responses ranged from terse one-liners (e.g., "The scope of the experiment is too ambitious given the budget") to multi-sentence paragraphs offering detailed critiques or advice.

Evaluators left comments on 307 of the 641 evaluations (47.9%). After removing brief, generic remarks such as "None" or "No concerns," we were left with 281 comments, corresponding to 43.8% of all evaluations. Evaluators in the feasibility-only condition were more likely to comment than those in the multi-criteria condition. Specifically, 183 (54.1%) of the 338

feasibility-only evaluators provided concerns, compared to 98 (32.3%) of the 303 multi-criteria evaluators, $\chi^2((1, N = 641) = 29.958, p < 0.001))$.

The length of open-ended comments did not meaningfully differ between conditions. Multi-criteria evaluators averaged 298 characters versus 275 characters for feasibility-only evaluators ($t(184) = 0.800$, $p = 0.425$). These descriptive statistics indicate that the feasibility-only format primarily increased the likelihood of providing feedback, not the length of responses.

Hybrid Human-LLM Coding Approach

Two authors independently analyzed a random sample of 57 comments distributed across treatment conditions (multi-criteria vs. feasibility-only). After excluding non-informative responses, we assessed 969 possible codes across feasibility, impact, and novelty sub-criteria. Through comparative analysis and resolution of coding discrepancies, we established a preliminary coding framework that captured anticipated and emergent themes within the data.

We then iteratively refined this initial framework through structured consultations with two program managers (scientists with more than ten years of grant evaluation experience) from the medical grant organization. These scientists reviewed our coding categories, suggested refinements, and helped identify domain-specific nuances crucial for accurate classification. Through multiple cycles of expert feedback and framework adjustment, we developed a consensus-based codebook incorporating predefined and emergent dimensions.

The finalized codebook then served as the foundation for developing LLM coding prompts, ensuring alignment between human-coded judgments and automated scaling of the full dataset. We used a few-shot prompting approach by providing the model with concrete examples illustrating each of the seven feasibility sub-criteria. Each comment underwent structured coding

via a custom LLM prompt that instructed the model to (1) summarize the comment's main points and (2) identify mentions of the seven feasibility sub-criteria, novelty, and impact. The full prompt and details of the coding procedure are provided in Online Appendix D.

For example, consider the anonymized comment, for which LLM coding classified as mentioning both experimental design and alternative plans:

> The authors have identified numerous potential problems with this group of patients, and they have realistically made proposals for alternative plans. It may be challenging to keep the patients on task long enough to avoid fatigue, which could alter the [sensory] modality of the study since it follows the [task]. Even with breaking up the testing into additional sessions, subjects may forget portions of the prior [training], and repeat [training] may lead to fatigue before the auditory portion begins.

To verify the alignment of LLM-based coding to human coders, two authors independently coded a random sample of 20 comments, stratified by treatment condition. The inter-rater reliability between human coders and the LLM was measured using Cohen's kappa, yielding an average $\kappa = 0.75$ across four human-LLM rating pairs. According to conventional thresholds (Landis & Koch, 1977), this value indicates substantial agreement between human and automated coding approaches. Further validation came from the two scientist-program managers from the grant funding program, who independently reviewed the LLM prompt and a separate sample of LLM-assigned codes and confirmed their accuracy. Detailed inter-rater reliability metrics are provided in Online Appendix D6.

Results from Human-LLM Analysis of Qualitative Data

Figures 3 and 4 visualize how the cognitive mechanisms of critical flaw discovery and coherence scrutiny differ between the feasibility-only and multi-criteria evaluators, as revealed through the language patterns of the evaluators' open-ended comments. The data patterns demonstrate how

structural changes in evaluation format alter evaluators' judgments systematically and predictably.

Critical Flaw Discovery and Lower Feasibility Scores

Figure 3 reveals clear patterns in attention allocation between our two experimental conditions. As predicted in our theoretical framework, critical flaw discovery manifested as expanded breadth in examining the feasibility sub-criteria from the evaluation rubric. Feasibility-only evaluators mentioned feasibility sub-criteria—particularly "analysis," "timeframe," and "regulatory" aspects—more frequently than multi-criteria evaluators in their open-ended comments. This pattern supports our theoretical argument that focused attention under a non-compensatory evaluation framework enables a broader decomposition of feasibility into its constituent sub-criteria, as defined in the evaluation rubric.

Notably, the data also shows that feasibility-only evaluators mention novelty and impact sub-criteria (such as "new technology," "new approach," and impact on "healthcare systems") less frequently than multi-criteria evaluators. This provides evidence of our proposed focused attention allocation mechanism: when instructed to focus exclusively on feasibility, evaluators redirect cognitive resources away from innovation and impact dimensions and toward a more comprehensive feasibility analysis. In contrast, multi-criteria evaluators' selective emphasis on "budget" and "resources" suggests they employ a more superficial assessment strategy when their attention is divided, prioritizing the most immediately visible feasibility factors that can be more readily addressable while potentially overlooking more subtle implementation challenges. These patterns are consistent with the proposed mechanism of critical flaw discovery in the feasibility-only treatment condition, resulting in lower feasibility scores than the multi-criteria control condition.

[Figure 3 about here]

Coherence Scrutiny and Proposal Re-Ranking

Figure 4 provides evidence for our proposed coherence assessment mechanisms. The visualization shows differences in co-occurring mentions of sub-criteria pairs between the feasibility-only treatment and multi-criteria control conditions. The feasibility-only evaluators demonstrate substantially more joint consideration of multiple feasibility sub-criteria. The strongest co-occurrence patterns appear for pairs like "analysis-design," "analysis-timeframe," and "analysis-resources," indicating that feasibility-only evaluators actively examine interdependencies among different sub-criteria more frequently than the multi-criteria evaluators.

This integration pattern supports the cognitive mechanism that focused attention under a non-compensatory framework facilitates intra-criterion coherence assessment. Evaluators scrutinize how various feasibility aspects interact with each other, potentially identifying inconsistencies that might be overlooked in a more divided evaluation. Multi-criteria evaluators, in contrast, show fewer co-occurring mentions of feasibility sub-criteria (indicated by red bars). The substantial differences in coherence scrutiny patterns help explain why proposal rankings would be reordered, not just downgraded, in the feasibility-only condition, as this deeper integration reveals systemic inconsistencies that would remain hidden under divided attention conditions.

The extensive integration of multiple feasibility sub-criteria by feasibility-only evaluators is consistent with a non-compensatory weakest link approach, where proposals must satisfy requirements across all interconnected dimensions. Multi-criteria evaluators' more limited consideration of feasibility sub-criteria co-occurrences suggests they may be employing a more

compensatory approach, possibly at the expense of a more comprehensive feasibility analysis, potentially the systematic re-ranking of proposals between conditions.

[Figure 4 about here]

PREDICTIVE VALIDITY OF FEASIBILITY ASSESSMENTS

We analyze the relationship between feasibility scores and subsequent research output to determine whether feasibility-only evaluations better identify implementable projects. Building on prior work that uses post-funding scientific productivity to validate ex-ante assessments (Li, 2017), we use matched publication activity as an outcome-based proxy for feasibility: projects rated as more feasible should generate more thematically related publications.

We identify all 2023 and 2024 publications authored by the proposal's principal investigator and match them to the proposal title using a keyword-based approach. This extended window captures the earliest plausible outputs from funded work while allowing sufficient time for early-stage, resource-constrained research to yield observable publications. The dependent variable is the number of papers with at least one matched keyword in the title, a proxy for downstream implementation of the proposed project (see Online Appendix C for details).

Table 3 presents the results separately for funded and unfunded proposals. In both subsamples, feasibility-only scores are more predictive of research output than multi-criteria scores. For unfunded proposals, the interaction between evaluation condition and feasibility score is positive (Model 1: 1.279, $p = 0.038$), indicating that feasibility-only reviewers more accurately identified projects that ultimately produced related scientific output, even without receiving any funding. For funded proposals, the interaction effect is positive (Model 2: 1.051, $p = 0.081$), though the smaller magnitude and higher p-value may reflect the dampening effect of

resource allocation: funded projects may use the financial support to overcome feasibility barriers, weakening the link between ex-ante judgments and downstream implementation.

These patterns suggest that feasibility-only evaluations provide a stronger signal of a project's scientific readiness. Multi-criteria evaluators—likely influenced by impact and novelty considerations—produced scores less aligned with actual implementation outcomes. This interpretation is consistent with earlier findings (Figure 2) that multi-criteria evaluations favor more novel or ambitious proposals, which may carry greater feasibility risks despite their scientific potential.

[Table 3 about here]

DISCUSSION AND CONCLUSION

Evaluations of early-stage ideas are pivotal in assessing the feasibility of projects given current resources, skills, and technology. Evaluation formats that amalgamate various criteria, such as a project's novelty, impact, and feasibility, can potentially shape judgment patterns based on how evaluators allocate their limited attention across multiple criteria and how they conceptualize relationships between these criteria (Boudreau et al., 2016; Criscuolo et al., 2017; Li et al., 2017). Since decision-makers often leverage evaluator judgments to allocate resources, these assessment patterns can impact the selection process for innovative projects. This paper explores the role of feasibility-only evaluations in developing more comprehensive assessments of whether a project has limitations that could impede implementation.

We investigated this research question by designing an intervention that exogenously varied the evaluation format to accommodate a comparison of feasibility assessments between multi-criteria evaluators, who assess project novelty, impact, and feasibility, and feasibility-only evaluators, whose focus is exclusively on determining 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 several noteworthy patterns. Our field experiment revealed that the evaluation format meaningfully affects how thoroughly evaluators assess feasibility. Feasibility-only evaluators assigned systematically lower scores (6-10 percentage points). The low correlation ($\rho = 0.22$) between proposal rankings in the two conditions further suggests that the evaluation format fundamentally alters feasibility assessment.

Our analysis reveals two mechanisms driving differences between evaluation formats. First, critical flaw discovery occurs when feasibility-only evaluators expand their analytical breadth across sub-criteria potentially overlooked under divided attention. This expanded scrutiny increases the probability of identifying disqualifying constraints, contributing to lower feasibility scores (Table 2). Second, coherence scrutiny emerges as feasibility-only reviewers demonstrate greater synthesis across proposal elements, examining how sub-criteria interact and identifying inconsistencies obscured under divided attention. Analysis of co-occurring mentions confirms that feasibility-only evaluators more frequently consider relationships between sub-criteria like experimental design, analysis, and resources. Multi-criteria evaluators show fewer interconnected feasibility assessments but link feasibility to novelty and impact. These different integration patterns explain the proposal reordering (Figure 1).

Theoretical and Methodological Contributions

Our work makes several key contributions. First, this work extends our understanding of organizational evaluation processes by connecting the role of assessment format in shaping

attention and evaluation logic. Prior research has highlighted various factors affecting evaluation outcomes, including evaluator characteristics (Boudreau et al., 2016), contextual factors (Bian et al., 2021; Criscuolo et al., 2021), information sharing (Lane et al., 2022), and organizational context (Criscuolo et al., 2017). We show that the structure of the evaluation process itself—particularly whether criteria are assessed jointly or separately—affects how evaluators approach their task and what they conclude. This reveals an essential tension between a broad assessment (multi-criteria review) and a narrow analysis (single-criterion review), with practical implications for how organizations design their evaluation systems. Importantly, our findings bring the importance of attention allocation and cognitive resources (Ocasio, 1997) to the innovation evaluation literature to suggest how evaluation formats can affect assessments of project quality: while multi-criteria evaluation enables evaluators to identify complementarities or trade-offs between dimensions, feasibility-only evaluation promotes a more comprehensive analysis of potential implementation challenges. This suggests that organizations might benefit from incorporating both approaches in their evaluation processes.

While previous research has established the importance of attention in organizational decision-making, we demonstrate how specific structural choices—in this case, evaluation format—causally affect attention allocation and its consequences. Our findings show that focused attention, enabled by single-criterion evaluation, provides novel evidence for how organizational design choices can direct cognitive resources toward specific aspects of assessment, with meaningful consequences for evaluation outcomes. More generally, in an era of increasing information overload and competing demands for attention (Anderson & De Palma, 2012; Leonardi et al., 2012), our findings suggest that organizational structures that deliberately constrain the scope of evaluation tasks can enhance the comprehensiveness of the assessment.

This has implications beyond evaluation contexts to other organizational processes that operate in a non-compensatory framework, such as risk assessment, due diligence, and strategic planning, where the weakest link can undo an otherwise promising initiative.

Second, this study contributes to judgment and decision-making research by demonstrating how evaluation formats fundamentally alter cognitive processes. While prior work has established that decision-makers employ different heuristics and processing strategies depending on effort-accuracy trade-offs (Payne, Bettman & Johnson, 1993; Kahneman, 2011), we provide causal evidence of how structural choices shape these dynamics in evaluation contexts. Our findings reveal that dividing attention across multiple criteria versus focusing on a single dimension transforms how they process information and make judgments. Feasibility-only evaluators employ more comprehensive analytical strategies, disaggregating complex assessments into component parts and scrutinizing connections between them. This extends our understanding of bounded rationality (Simon, 1955) by showing how organizational design choices systematically influence the breadth and depth of information processing.

Third, this study makes a methodological contribution by demonstrating how field experiments combined with human-LLM analysis can reveal cognitive mechanisms underlying organizational processes. While cognitive processes are typically difficult to observe directly, our approach shows how careful research design and novel analytical methods can unpack these mechanisms. This methodological innovation opens new possibilities for studying how organizational structures affect individual cognition and decision-making in practice.

Policy Implications for Evaluation Design and Innovation Funding

Our findings have important policy implications for organizations allocating funding for innovative early-stage projects. Feasibility-only evaluations may be particularly valuable for

identifying and addressing implementation challenges in high-risk, high-reward projects that score highly on novelty and impact but are associated with lower feasibility. This approach ensures that evaluators dedicate sufficient cognitive resources to uncovering potential barriers to implementation, which can inform more strategic decision-making and mitigate downstream risks of failure. However, relying exclusively on feasibility assessments risks prioritizing safer, incremental projects, possibly at the expense of transformative breakthroughs.

Amid the push for evidence-based evaluation (Azoulay & Li, 2020), policymakers and funders could adopt hybrid or staged reviews that blend the depth of single-criterion focus with the breadth of multi-criteria perspective, mitigating cognitive biases, clarifying trade-offs among feasibility, novelty, and impact, and ultimately optimizing portfolios for both near-term implementability and long-term value.

Limitations and Future Research

Although we have made a thorough and concerted effort to analyze how feasibility-only and multi-criteria evaluations shape feasibility assessments, our study has some empirical limitations.

First, our study examines evaluation processes within a relatively established field in the life sciences—specifically, in human sensory systems. Future research could explore feasibility evaluations in other scientific fields with varying degrees of theoretical consolidation and openness to unconventional approaches (Azoulay et al., 2019). Additionally, our sample of evaluators, drawn primarily from a highly selective medical school and affiliated institutions, reflects specific institutional norms, networks, and incentives that shape feasibility perceptions. Research in alternative institutional contexts—including those with differing degrees of interdisciplinary collaboration, industry ties, or funding constraints—could yield further insights into how evaluators weigh feasibility in different decision-making environments. Moreover,

while our study focuses on scientific evaluation, the dynamics of multi-criteria assessment are highly relevant to other domains, such as venture capital investing (Bian et al., 2021) and new product development (Criscuolo et al., 2021; Katila & Ahuja, 2002), offering compelling opportunities for experimentation with alternative evaluation formats. In domains where innovation tends to be driven by feasibility concerns, novelty- or impact-only evaluations may be more informative in filling the innovation pipeline's void of high-impact projects (Krieger et al., 2022).

Second, our study captures how evaluators assess feasibility, novelty, and impact in a real-world setting. To provide objective validation, we implemented Li's (2017) approach to track publication outcomes linked to funded proposals by matching grant titles to subsequent publications by the same authors. This approach offers valuable insight into the predictive validity of feasibility assessments across evaluation formats. However, this validation approach has important limitations. Publications represent just one dimension of feasibility, namely the scientific rigor and technical implementation of the proposed work, but they potentially miss other aspects like adherence to timeline, budget, or resource efficiency. Additionally, the proposed work's inherent quality will likely predict future funding and implementation. While our publication matching approach offers a valuable external validation measure, future research could develop more comprehensive multidimensional measures of implementation success that capture the full spectrum of feasibility dimensions.

Third, another limitation of our study is that we cannot fully separate the effects of evaluation format from differences in cognitive workload between conditions. Multi-criteria evaluators faced inherently higher cognitive demands—they needed to assess three distinct dimensions (impact, novelty, and feasibility) rather than just one, requiring mental task-

switching and distributing limited cognitive resources across multiple evaluation criteria. Consequently, their less thorough feasibility assessments might stem partly from having less time and attention to devote to feasibility after addressing other dimensions rather than from fundamentally different cognitive processing strategies. While our qualitative analyses suggest systematic differences in how evaluators conceptualize feasibility across conditions, future research could more precisely isolate these mechanisms by implementing designs that control for total evaluation time and cognitive load. For instance, researchers could employ matched-effort conditions where single-criterion evaluators complete additional unrelated tasks to equalize total cognitive effort or use eye-tracking and neuroimaging techniques to measure attention allocation directly during evaluation tasks.

Finally, conducting a field experiment ensures high ecological validity by capturing real-world decision-making among expert evaluators. This design strengthens the study's practical relevance but also imposes trade-offs. The constraints of expert recruitment, the demands on evaluators' time, and the real funding stakes limited the experimental conditions we could implement. As a result, we did not test alternative sequences of evaluation criteria, systematically manipulate the proposal load, or include a novelty- or impact-only condition to isolate cognitive mechanisms further. Future studies could complement our findings through controlled lab experiments that systematically vary these factors, testing whether the mechanisms we identify generalize across different evaluation conditions. Additionally, changing the number of evaluation criteria could shed light on how additional cognitive load affects resource allocation in multi-criteria decision-making.

In conclusion, we believe that the findings of the present research pave the way to more informed selections of evaluation formats for identifying high-impact 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. (1983). Brilliant but cruel: Perceptions of negative evaluators. *Journal of Experimental Social Psychology*, *19*(2), 146–156.

Anderson, S. P., & De Palma, A. (2012). Competition for attention in the information (overload) age. The RAND Journal of Economics, 43(1), 1-25.

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., & Li, D. (2020). *Scientific grant funding* (Vol. 26889). Cambridge, MA: 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.

Bamberger, P. A. (2019). On the replicability of abductive research in management and organizations: Internal replication and its alternatives. *Academy of Management Discoveries*, *5*(2), 103-108.

Baker, T. and Nelson, R.E. (2005). Creating something from nothing: Resource construction through entrepreneurial bricolage. Administrative Science Quarterly, 50(3), 329–366.

Behfar, K., & Okhuysen, G. A. (2018). Perspective—Discovery within validation logic: Deliberately surfacing, complementing, and substituting abductive reasoning in hypothetico-deductive inquiry. *Organization Science*, *29*(2), 323-340.

Berg, J. M. (2016). Balancing on the creative highwire: Forecasting the success of novel ideas in organizations. *Administrative Science Quarterly*, *61*(3), 433–468.

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.

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.

Chang, L.W., Kirgios, E.L., Mullainathan, S. and Milkman, K.L., 2024. Does counting change what counts? Quantification fixation biases decision-making. Proceedings of the National Academy of Sciences, 121(46).

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.

Dougherty, D. (1992). A practice-centered model of organizational renewal through product innovation. *Strategic Management Journal*, *13*(S1), 77–92.

Eklund, J., Raj, M., & Eggers, J. P. (2025). Attention focus and new opportunities: The moderating role of managerial attention to alternative issues. *Organization Science*, *36*(1), 21-39.

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.

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.

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.

Holyoak, K. J., & Simon, D. (1999). Bidirectional reasoning in decision making by constraint satisfaction. *Journal of Experimental Psychology: General*, *128*(1), 3.

Hsee, C.K., Loewenstein, G.F., Blount, S. and Bazerman, M.H. (1999). Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. Psychological Bulletin, 125(5), 576–590.

Joseph, J., & Wilson, A. J. (2018). The growth of the firm: An attention-based view. *Strategic Management Journal*, *39*(6), 1779-1800.

Joseph, J., & Ocasio, W. (2012). Architecture, attention, and adaptation in the multibusiness firm: General electric from 1951 to 2001. *Strategic Management Journal*, *33*(6), 633-660.

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., Li, D., & Papanikolaou, D. (2022). Missing novelty in drug development. *The 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: Inside the Curious World of Academic Judgment*. Harvard University Press.
Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, *33*(1), 159–174.
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.
Leiponen, A., & Helfat, C. E. (2010). Innovation objectives, knowledge sources, and the benefits of breadth. *Strategic Management Journal*, *31*(2), 224–236.
Leonardi, P. M., Neeley, T. B., & Gerber, E. M. (2012). How managers use multiple media: Discrepant events, power, and timing in redundant communication. Organization Science, 23(1), 98-117.
Levinthal, D. A., & March, J. G. (1993). The myopia of learning. *Strategic Management Journal*, *14*(2), 95-112.
Li, D., 2017. Expertise versus Bias in Evaluation: Evidence from the NIH. *American Economic Journal: Applied Economics*, *9*(2), 60–92.
March, J. G., & Simon, H. A. (1958). *Organizations*. Wiley.
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.
NIH Research Grants – Digital Media Kit. (2024, December 19). National Institutes of Health; National Institutes of Health. https://www.nih.gov/news-events/nih-research-grants-digital-press-kit.
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.
Posner, M.I. and Petersen, S.E. (1990). The attention system of the human brain. Annual Review of Neuroscience, 13(1), pp.25-42.
Prahalad, C. K., & Hart, S. L. (2002). The fortune at the bottom of the pyramid. Strategy+Business, 26.
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.
Rosenberg, N., & Nelson, R. R. (1994). American universities and technical advance in industry. *Research Policy*, *23*(3), 323–348.

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. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics, 69*(1), 99–118.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.
Tan, D. (2022). The road not taken: Technological uncertainty and the evaluation of innovations. *Organization Science*, *34*(1), 156–175.
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.
Tripsas, M., & Gavetti, G. (2000). Capabilities, cognition, and inertia: Evidence from digital imaging. *Strategic Management Journal*, *21*(10-11), 1147–1161.
Tversky, A. (1969). Intransitivity of preferences. *Psychological Review, 76*(1), 31–48.
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.
Weitzman, M. L. (1998). Recombinant growth. *Quarterly Journal of Economics*, *113*(2), 331–360.Zhou, K. Z., & Wu, F. (2010). Technological capability, strategic flexibility, and product innovation. *Strategic Management Journal, 31*(5), 547–561.

Table 1. Primary Evaluation Criteria and Sub-Criteria

	Criteria		
Sub-Criteria	**Novelty**	**Impact**	**Feasibility**
	Produce new technology or knowledge	Patients	Experimental design
	Show unanticipated application of existing knowledge	Healthcare providers	Proposed analysis plan
	Generate a new approach for an established modality or concept	Healthcare systems	Proposed resources, including access to samples/patients
	Address a translational gap	Society	Alternative plan
	Improve an existing paradigm	Academic field	Timeframe (12 months)
			Scope – for $50K
			Regulatory requirements

Table 2. OLS Regression Models of Solution Feasibility Score on Feasibility-Only Treatment

	Dependent Variable: Feasibility Score			
	(1)	(2)	(3)	(4)
Feasibility-only treatment	-0.184	-0.186	-0.188	-0.197
	(0.063)	(0.063)	(0.063)	(0.061)
Intellectual similarity		-0.267	-0.300	-0.482
		(0.955)	(0.961)	(1.083)
Topic area expertise		-0.009	-0.014	-0.055
		(0.035)	(0.037)	(0.040)
Early career		0.063	0.063	0.089
		(0.073)	(0.074)	(0.073)
Constant	3.116	3.110	3.477	0.934
	(0.046)	(0.139)	(0.335)	(4.355)
Observations	641	641	641	641
Proposal Attributes	N	N	Y	N
Proposal FE	N	N	N	Y
R^2	0.013	0.014	0.017	0.163

Table 3. OLS Models Testing Predictive Validity of Feasibility Scores on Feasibility-Only Treatment

	Dependent Variable: # of Keyword Matched Post-Publications	
	Unfunded Proposals	Funded Proposals
	(1)	(2)
Feasibility-only Treatment	-3.561	-3.866
	(1.872)	(2.086)
Feasibility Score	-1.224	-0.476
	(0.463)	(0.500)
Treatment × Feasibility Score	1.279	1.051
	(0.617)	(0.603)
Constant	7.884	4.136
	(1.427)	(1.780)
Observations	471	170
R^2	0.015	0.028

Note: Standard errors (in parentheses) clustered by evaluator and proposal.

Figure 1. Divergence in Feasibility Rankings Under Multi-Criteria vs. Feasibility-Only Evaluations. Each line represents a single proposal's average feasibility score in the two conditions. Colors indicate whether a proposal's feasibility appears in the top 10 in both conditions, only one condition, or neither. The bounded boxes demarcate proposals in the top 10 in their respective conditions.

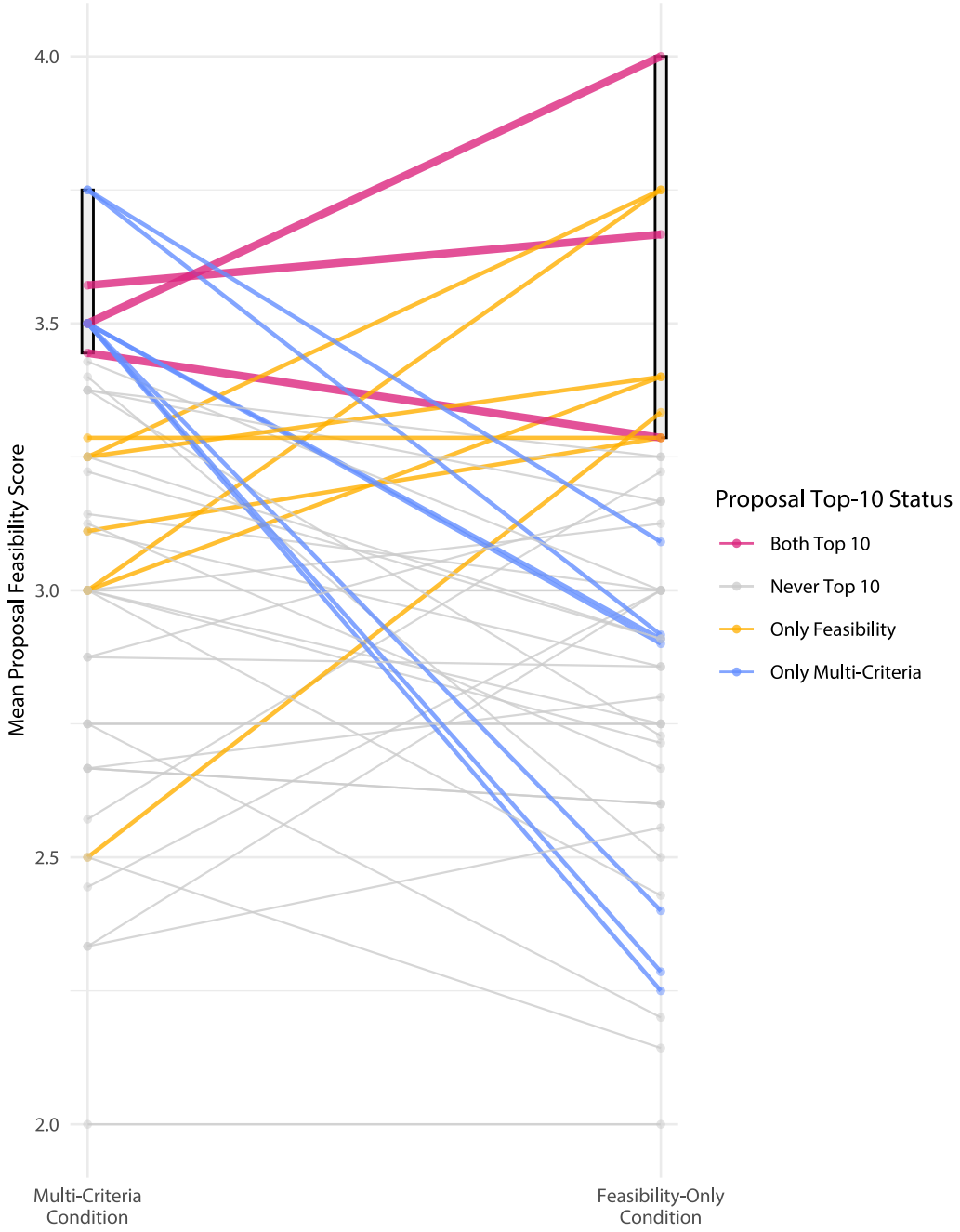


Figure 2. Comparison of Multi-Criteria Scores for Proposals Ranked by Feasibility Under Different Evaluation Formats. This figure displays three metrics—feasibility (left), impact (center), and novelty (right)—for proposals ranked by their feasibility scores under each evaluation condition. The feasibility-only condition is assigned the impact and novelty score of the proposal's multi-criteria review. All scores use the same measurement scale of cumulative mean scores from multi-criteria evaluations to ensure consistent comparison. The vertical line at rank 10 indicates the funding threshold.

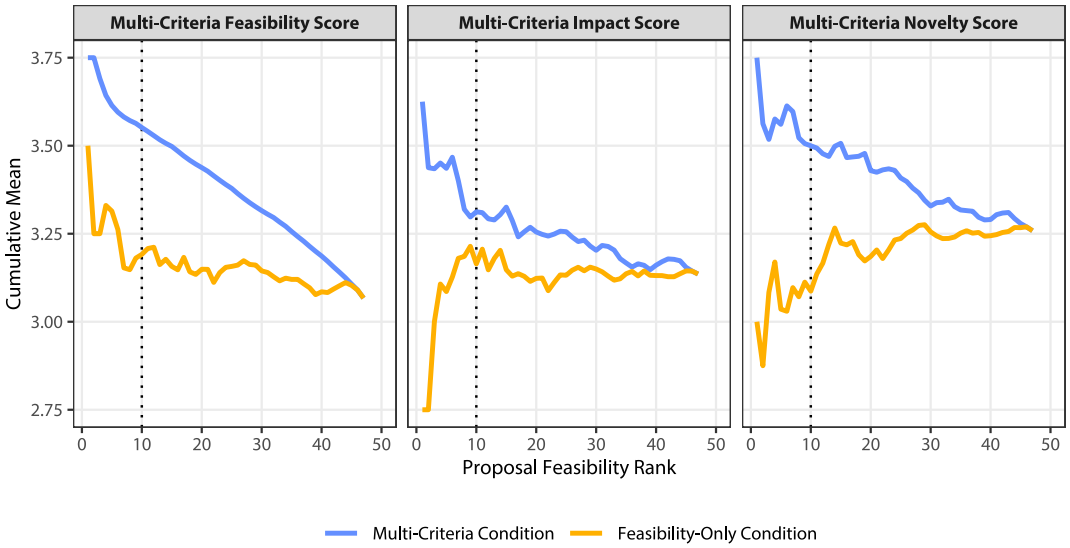


Figure 3. Difference in Sub-criteria Mentions between Feasibility-Only and Multi-Criteria Conditions normalized as percentage points of total mentions in each condition. Positive values (orange) indicate sub-criteria mentioned more frequently by feasibility-only evaluators; negative values (purple) indicate higher mention frequency by multi-criteria evaluators. Based on 183 feasibility-only and 98 multi-criteria evaluator comments.

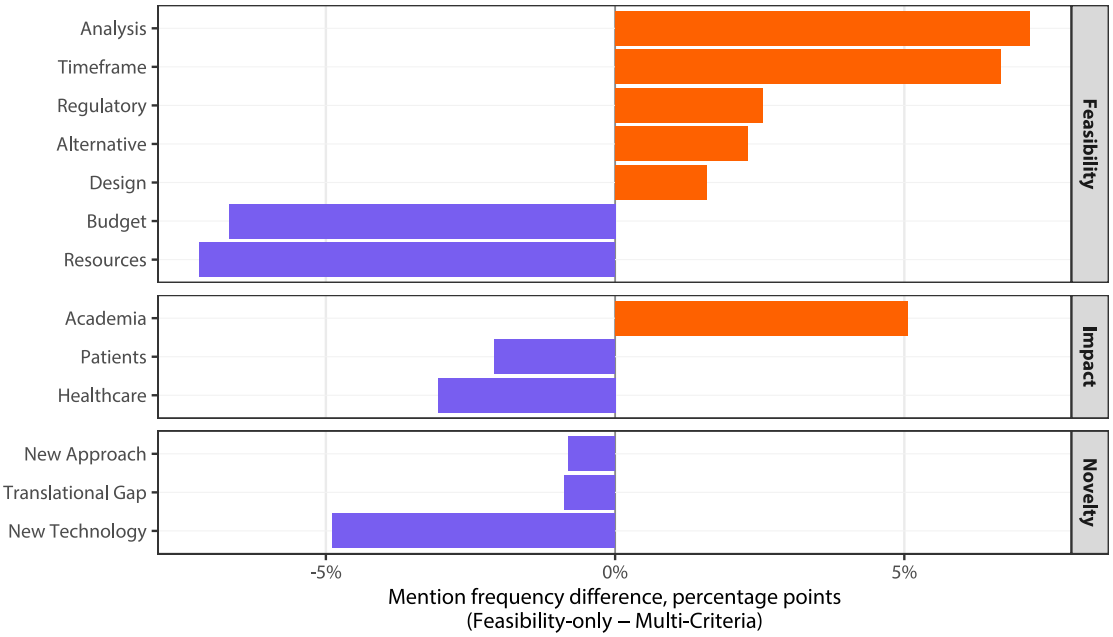


Figure 4. Normalized Difference in Feasibility Sub-criteria Co-occurrence Mentions between Feasibility-Only and Multi-Criteria Conditions. Bars represent the difference in frequency with which pairs of criteria are mentioned together in evaluator comments, normalized as percentage points of total co-occurrences in each condition. Positive values (orange) indicate criterion pairs more frequently co-mentioned by feasibility-only evaluators; negative values (purple) indicate pairs more frequently co-mentioned by multi-criteria evaluators. Based on 183 feasibility-only and 98 multi-criteria evaluator comments.

