Looking Across and Looking Beyond the Knowledge Frontier: Intellectual Distance and Resource Allocation in Science

Kevin J. Boudreau, Eva Guinan, Karim R. Lakhani and Christoph Riedl

Abstract:

Selecting among alternative projects is a core management task in all innovating organizations. In this paper, we focus on the evaluation of frontier scientific research projects. We argue that the intellectual distance between the knowledge embodied in research proposals and an evaluator's own expertise systematically relates to the evaluations given. To estimate relationships, we designed and executed a grant proposal process at a leading research university in which we randomized the assignment of evaluators and proposals to generate 2,130 evaluator-proposal pairs. We find that evaluators systematically give lower scores to research proposals that are closer to their own areas of expertise and to those that are highly novel. The patterns are consistent with biases associated with boundedly rational evaluation of new ideas. The patterns are inconsistent with intellectual distance simply contributing “noise” or being associated with private interests of evaluators. We discuss implications for policy, managerial intervention and allocation of resources in the ongoing accumulation of scientific knowledge.

*Kevin J. Boudreau, London Business School (kboudreau@london.edu); Eva Guinan, Dana Farber Cancer Institute and Harvard Medical School (Eva_Guinan@dfci.harvard.edu); Karim R. Lakhani, Harvard Business School and Institute for Quantitative Social Science (k@hbs.edu); Christoph Riedl, D’Amore-McKim School of Business and College of Computer and Information Science, Northeastern University (c.riedl@neu.edu). We are especially indebted to leaders and administrators at an anonymous granting organization and host university for their invaluable support. We also thank executives at InnoCentive who greatly assisted us with platform and information technology design and support. We would also like to acknowledge Thomas Astebro, Michael Bikard, Amy Edmonson, Dan Elfenbein, Alfonso Gambardella, Yigal Gerchak, Shane Greenstein, Ben Hallen, Danielle Li, Lars Bo Jeppesen, Scott Page, Gary Pishno, Anisa Shyti, Eric von Hippel, Keyvan Valkili, Feng Zhu and participants at presentations at Bocconi University, Cambridge University, INSEAD, London Business School, National Bureau of Economic Research (NBER), the Roundtable for Engineering Entrepreneurship Research (REER), University of Maryland and Washington University at St. Louis for useful comments. We acknowledge support from Deloitte Institute on Innovation and Entrepreneurship, London Business School RAMD, the NASA Tournament Laboratory, German Research Foundation grant RI 2185/1-1, and support form Harvard Catalyst, and NIH UL1TR000170, UL1RR025758-02S4 and UL1TR001102S grants. Eric Lonstein provided excellent support with the execution of the review process and data collection. Griffin Weber provided valuable assistance in selecting evaluators based on their field of expertise. Eric Lin provided access to the PubMed citation data. All errors are our own.
Looking Across and Looking Beyond the Knowledge Frontier:
Intellectual Distance and Resource Allocation in Science

Abstract

 Selecting among alternative projects is a core management task in all innovating organizations. In this paper, we focus on the evaluation of frontier scientific research projects. We argue that the “intellectual distance” between the knowledge embodied in research proposals and an evaluator’s own expertise systematically relates to the evaluations given. To estimate relationships, we designed and executed a grant proposal process at a leading research university in which we randomized the assignment of evaluators and proposals to generate 2,130 evaluator-proposal pairs. We find that evaluators systematically give lower scores to research proposals that are closer to their own areas of expertise and to those that are highly novel. The patterns are consistent with biases associated with boundedly rational evaluation of new ideas. The patterns are inconsistent with intellectual distance simply contributing “noise” or being associated with private interests of evaluators. We discuss implications for policy, managerial intervention and allocation of resources in the ongoing accumulation of scientific knowledge.

1 Introduction

A fundamental challenge that all organizations engaged in scientific and technological innovation face is how to allocate resources across alternative project proposals. Senior managers and scientific researchers alike devote significant time and effort to evaluating and selecting projects. A common approach in evaluating innovative projects is to refer to experts with deep domain knowledge to assess quality of proposed projects (Chubin and Hackett, 1990; Lamont, 2009). In the United States, for example, academic research, which is the feedstock for many subsequent commercial innovations, depends on expert peer review to allocate more than $40 billion of research funds every year in engineering, medicine, science and technology (Xie and Killewald, 2012). Contrary to a popular notion of a “marketplace for ideas,” in which the best ideas simply rise to the top, resource allocation in academic science is shaped in important ways by supporting institutions and processes (Kuhn, 1962; Merton 1968; Dasgupta and David, 1994; Stephan, 2012). In this paper we investigate how “intellectual distance”—the degree of overlap and relatedness between evaluators’ knowledge or expertise and the knowledge embodied in research proposals—plays a role in systematically shaping evaluation outcomes and consequent resource allocation in scientific peer review.

1 Stevens and Burley (1997) find that executives have to manage, on average, more than 3,000 ideas to secure one commercial success and tens of thousands of experts are involved in the annual evaluation of more than 89,000 research applications by the National Institutes of Health (NIH) and National Science Foundation (NSF).
The evaluation and funding process for leading edge scientific and technological projects is highly competitive. In the United States, for example, the National Institutes of Health (NIH) fund fewer than one in six applications and in the National Science Foundation (NSF) it is one in four. Between one-third and one-half of rejected project proposals and their associated research lines are subsequently discontinued by their authors (Chubin and Hackett, 1990). Although rejected proposals might simply be of lower quality and deserve to be stopped, tremendous unexplained variation and seeming “noise” is the single most regular feature of scientific peer evaluations. Interrater reliability in funding decisions is routinely found to be very low (e.g., Rothwell and Martyn, 2000; Bornmann and Daniel, 2008; Jackson et al., 2011), with concordance sometimes “barely beyond chance” (Kravitz et al., 2010; Lee et al., 2013) and “perilously close to rates found for Rorschach inkblot tests” (Lee, 2012). Variance among reviewers is sometimes greater than variance between submissions (Cole et al., 1981). Beyond the fact of low inter-rater reliability, there is yet little agreement about underlying causes. Past research has argued that expert evaluation of research proposals may be shaped by any number of factors beyond the “true” quality of research including researcher and evaluator characteristics, ties between researchers and their evaluators, proposal formats and evaluation procedures. (See Marsh et al. (2007, 2008) and Lee et al. (2013) for comprehensive reviews and syntheses of the relevant findings).

In this paper, we depart from the existing literature that has hypothesized the role of personal characteristics and social structure as determinants of scientific evaluations. Here, we investigate whether the intellectual distance between evaluators’ knowledge and the knowledge embodied in research proposals has systematic effects on evaluations. We consider three theoretical perspectives and associated mechanisms through which intellectual distance might affect evaluations, independent of the true quality of a proposal. First, the evaluation process might simply be understood as a matter of evaluators each discerning a noisy “signal” of true quality, following a classical statistical decision-making under uncertainty perspective. In this case, greater intellectual distance (less expertise, greater ignorance) would lead to less precise evaluations, but no bias in evaluations. By contrast, a bounded rationality perspective predicts that regular heuristics in expert human judgement produce systematic biases in evaluation, and these should vary with intellectual distance. An agency perspective emphasizes the possibility that private interests of evaluators could play a role in evaluations, with plausible positive or negative effects on evaluations, depending on a project’s
bearing on or relationship with the evaluators’ own career. Our discussion of theory develops predictions relating evaluation scores to intellectual distance between the content of research proposals and the knowledge of evaluators, and also to the novelty of proposals in relation to its departure from all past research (itself a kind of distance).

Our key empirical challenge is to precisely observe variation in intellectual distance and relate this to evaluation outcomes, independent of conflating factors—including the (unobserved) true quality of research proposals. To implement a suitable experimental research design, we collaborated with the administrators of a research-intensive U.S. medical school to modify details of a research grant process for endocrine-related disease. We recruited a large number of evaluators, 142 world-class researchers from within the institution that were drawn from fields both inside and outside the disease domain. We randomly assigned each evaluator to 15 proposals from a total of 150 research proposals, yielding 2,130 evaluator-proposal pairs. The process was “triple blinded”, with evaluators and authors blinded to one another, while evaluators were also kept separate and anonymous in relation to one another. Focusing our analysis on the first stage of the grant process, in which ideas and new hypotheses were solicited and evaluated, allowed us to standardize the format and content of proposals and to simplify submission requirements so that we could restrict the process to single-author submissions. This design allowed us to associate each proposal with fine-grained metrics at the level of individual submitters and evaluators.

We find intellectual distance between proposal-evaluator pairs is positively related to evaluation scores; that is, evaluators with more expertise are systematically more critical in their evaluations, consistently assigning lower scores. Evaluation scores are also negatively related to proposal novelty. The relationships are large. The range of intellectual distance observed here (associated with similarly-trained, leading medical researchers evaluating proposals in an endocrine-related disease), accounts for 1.1 points on a 10-point scale. (The standard deviation of evaluation scores is 2.6 or 1.7 points when proposal and evaluator fixed effects are removed.) The difference in evaluation scores between the least novel and the most novel proposals is -2.7 points. As the absolute scores of evaluations might often be less important than the rank-ordering of proposals, we perform simulations to reveal the (very) large impact of effects and evaluation policies on rank-ordering and grant awards.

The systematic relationships with intellectual distance and novelty are consistent with biases introduced as a consequence of expert evaluators’ bounded rationality in evaluating new ideas
(Kahneman et al., 1982; Johnson et al., 1982; Camerer and Johnson, 1991). Specifically, mental processes of highly-trained and experienced experts allow them to perceive and make sense of informational cues that go undetected by less expert, less intellectually proximate evaluators. The patterns in the data are consistent with non-experts being less capable of detecting nuanced problems and limitations that differentiate proposals as compared to experts, than they are in perceiving the intended contributions and virtues of research proposals. This generates an information “sampling” problem leading to bias, rather than just a question of “noise” in evaluations. Patterns are also inconsistent with evaluations being biased by agency problems. The negative relationship between evaluations and proposal novelty, in particular, is consistent with bounded rationality in the form of systematic miscontrual of approaches that deviate from existing mental maps and established knowledge. We are, however, unable to rule out the possibility that novel proposals are discounted on the basis of “ambiguity aversion” (Fox and Tverksy, 1995) or that novel proposals are, by their nature, of systematically lower–expected quality.

The remainder of the paper proceeds as follows. Section 2 reviews past literature and motivates possible links between intellectual distance and evaluations. Section 3 describes the research design. Section 4 presents our main results. These are discussed and interpreted in Section 5, together with a series of supplementary discriminating tests. Section 6 discusses policy implications. Section 7 concludes.

2 Advancing Scientific Knowledge and Evaluations

In this section, we first describe how recurrent patterns of knowledge accumulation in science inevitably lead to some degree of intellectual distance between new research proposals and the knowledge of evaluators. We distinguish intellectual distance between particular pairs of research proposals and evaluators from novelty in relation to the entire existing body of research. We then discuss three distinct theoretical perspectives suggesting intellectual distance might shape and evaluations, independent of the true quality of a proposal. (Note, each of there perspective reflects vast literatures and we only provide a brief overview of arguments as a means of summarizing key differences in their implications.)
## 2.1 Intellectual Distance in the Regular Advance of Scientific Knowledge

Advances in scientific knowledge tend not to be a scattershot of isolated experiments in all directions but rather a series of regular accumulative patterns (Gibbons et al., 1994). Initial progress on the resolution of a scientific problem gives rise to a scientific paradigm (Kuhn, 1962), defined as: common knowledge and consensus on what is to be observed; which questions are legitimate and interesting to ask; what constitutes appropriate and useful approaches to addressing these questions; what methods might be fruitfully employed; and even what legitimate answers might look like. Thus, except in the rare instances in which one paradigm is abandoned for another, the stock of knowledge tends to grow by regular accretion within the prevailing paradigm.

Disclosure and diffusion of scientific knowledge through publication, conferences, seminars, textbooks, graduate training and other means creates something of a common stock of open knowledge (Boudreau and Lakhani, 2015), as well as a commonly perceived knowledge frontier or envelope that demarcates what is currently known from what remains to be investigated. New research, which by definition aims to extend the current state of knowledge, creates intellectual distance between evaluators and proposals if only by requiring evaluators to look *beyond* the existing knowledge frontier. Incrementally novel advances can be made by continuing within existing pathways and paradigms. More novel departures from the existing paradigm might also be pursued, in hope of finding new viable research pathways and “breakthroughs” (Uzzi et al. 2013). Thus, novelty should be considered a matter of degree. Just as incremental advances largely proceed in a cumulative process that draws on existing templates, knowledge and ideas, novel departures themselves do not come from utterly unprecedented work. Rather, as documented in a range of empirical and theoretical considerations (Becker, 1982; Weitzman 1998; Fleming 2001; Uzzi, et al. 2013), novel approaches themselves draw on existing knowledge, but tend to then recombine and reconfigure this knowledge in unprecedented ways (Simonton 1995, 1999; Ben-David, 1960; Mullins, 1972; Law, 1973).

Intellectual distance between a particular evaluator and particular research proposal also arises as a result of growing specialization as scientific research advances. Despite the open and shared knowledge commons, scientific knowledge remains too vast, nuanced and complex to be understood in its entirety by any one scientist (Cowan et al. 2000; Wuchty et al. 2007; Jones 2010). Figure 1 illustrates the growth of scientific knowledge in the life sciences over 60 years (1950-2010) and
the tendency towards specialization into subfields through the increase in the cumulative numbers of journals, articles and research keywords. Even scientists that *prima facie* appear to be working in the same domain will differ in the particulars of their research program and differ in precise experience, training and exposure to phenomena and methods. As a result, evaluation of new research proposals also requires evaluators to look *across* the knowledge frontier to other domains not precisely overlapping with their own expertise, training and experience. Hence the very nature of scientific inquiry and our society’s reliance on experts to evaluate and allocate resources generates intellectual distance between evaluators and new proposals and creates evaluation challenges.

<Figure 1>

2.2 Three Perspectives on Intellectual Distance and the Evaluation of New Projects

Here we review three broad theoretical perspectives, each motivating possible links between intellectual distance and research evaluations, apart from any differences in true research quality. Although these perspectives are not mutually exclusive or entirely independent, it is useful to consider their arguments in turn. Predictions of these perspectives are summarized in Table 1 at the end of this section.

2.2.1 Agency Problems and the Private Interests of Evaluators

Much of the existing research on research evaluations hypothesizes some form of evaluator bias shaping evaluations. Most existing evidence is correlational and associative and not yet directly related to the question of intellectual distance.\(^2\) Nonetheless, we take the more general point emphasized by this work that evaluators’ private interests might lead to systematic deviations between expected quality and reported evaluations. Even just the content of a research proposal may relate to private interests of evaluators. For example, a negative relationship between evaluations and intellectual distance could exist, if evaluators are inclined to be less critical of or to favor “close” research. This is plausible given the nature of institutions and rewards in science (Stephan, 1996). Increased attention can attract additional resources and renown for one’s area of research,

\(^2\)Sources of bias considered include social category, status and prestige, sex, nationality, language and relationships between evaluators and researchers (see, for example, Merton, 1973; Ceci and Williams, 2011; Rees, 2011; etc.).
boosting the prospects of all involved—including the evaluator. Equally, a negative relationship
could exist if evaluators’ have preferences for given “schools of thought” or have a propensity for
“cognitive cronyism” (Travis and Collins, 1991). Alternatively, a positive relationship could exist
where, for example, research in the same domain and in close proximity is perceived to exert a
negative externality on the evaluator, creating incentives to discount evaluations. For example, in
certain instances a close and competitive proposal might be expected to draw resources and attention
away from an evaluator’s own work (Campanario and Acedo, 2007). Similarly, a wish to “protect”
orthodox theories might dispose evaluators to look negatively at research that is both proximate and
proposes a conflicting perspective (Travis and Collins, 1991). These biases, in whichever direction,
might also occur more subtly than simply evaluations in bad faith, as when personal interests affect
how much effort an evaluator is willing to devote to an evaluation (Johnson and Payne, 1985).

Past empirical research with some relevance to these arguments is not conclusive on these points.
For example, several papers have failed to find upward bias in evaluations of research that cites
evaluators (Sandstrom, 2009; Sugimoto and Cronin, 2013). Li (2013) finds clearer evidence of a
positive causal bias towards close researchers in the context of NIH committee evaluation; however,
committee dynamics and non-blinded evaluations make it difficult to interpret results in relation to
intellectual distance per se.

### 2.2.2 Uncertainty, Risk and Decision Theory Perspectives

Another theoretical perspective views proposal evaluation as akin to the problem of classical (stat-
istical) decision-making under uncertainty (e.g., Berger, 1985; Anand, 1993). This might be under-
stood in terms of reported evaluation scores ($V_{reported}$) being understood as reflecting both some
true, unobserved quality ($V_{true}$) and some “error” term (e.g., Blackburn and Hakal, 2006, p. 378),
i.e., $V_{evaluation} = V_{true} + error$. This perspective is implicit in the many references to “luck” and
“noise” in the literature (e.g., Cole et al., 1981; Marsh et al., 2008; Graves et al., 2011). This view
also relates to the common practices of averaging multiple evaluation scores in hopes of cancelling
noise and errors (Lee et al., 2013).

Following this view, greater intellectual distance can be interpreted as being less well-informed–
and therefore having greater uncertainty. Greater intellectual distance and uncertainty might then
manifest, for example, as a larger “error” term. This could produce greater dispersion and variance
of evaluations without necessarily affecting mean evaluations. Alternatively, greater intellectual distance and uncertainty might reduce confidence in assessments, which could plausibly lead to greater risk discounting of more distant evaluations.

Novel research proposals may face an added hurdle. Apart from uncertainty in the form or risk or errors, novelty introduces a form of fundamental uncertainty that can not entirely be resolved without experimentation. It is thus difficult to assign probabilities to outcomes *ex ante*. In cases of such unresolvable uncertainty or “ambiguity”, researchers in the behavioral decision making under uncertainty literature have found individuals tend to discount outcomes on the basis of “ambiguity aversion” (Fox and Tverksy, 1995). This reasoning also predicts a negative relationship between novelty and evaluations.

### 2.2.3 Bounded Rationality and Expert Cognition Perspectives

Research on bounded rationality and expert cognition also suggests links between intellectual distance and evaluations. The literature in this tradition finds that, across a wide range of human endeavor, expert judgment is associated with qualitatively distinct cognitive processes than those of non-experts. Experts, those closest to a particular subject matter, are able to observe and exploit a far broader array of informational cues. They perceive and appreciate more detail, complexity, patterns and meaning when making the very same observations as non-experts (see Bouman, 1980; Kahneman et al., 1982; Johnson et al., 1982; Camerer and Johnson, 1991). These advantages in information-processing are rooted in the development of a richer, more textured library of domain-specific knowledge accumulated through extended periods of training, experience and practice. As a result, experts require the same or less time and effort to generate more discerning judgements (Johnson and Russo, 1984; Johnson, 1988; Bedard, 1989). Expert cognitive processes are even often seemingly automatic, and even instantaneous, as a result of knowledge stored and comprehended in “chunks” and mental maps of hierarchies, relationships, contingencies and “configural rules” (Fitts and Posner, 1967; Newell and Simon, 1972; Chase and Simon, 1973; Ericsson and Smith, 1991).

Therefore, rather than a matter of intellectual distance resulting in more or less “error” in perceiving the same object, these points raise the possibility of information processing and “seeing more” creating differential *sampling* of information. Following this interpretation, the effect of intellectual distance and expertise depends on whether experts disproportionately see (sample)
merits or demerits in relation to those perceived (sampled) by less expert evaluators. It is only when experts differentially sample positive merits as they do negative demerits of a research proposal (and also weight them equally), where we would not expect some effect of expertise on mean evaluations. If merits and contributions are much plainer to see than are more subtle questions of feasibility, implementation and correctness greater expertise could result in more negative evaluations. This suggests the possibility of a positive relationship between intellectual distance and evaluations.

A distinct branch of the research on cognitive biases, studying effects of extrapolating on the basis of one’s existing knowledge into new domains, also suggests implications around questions of novelty. Extrapolation beyond the domain for which knowledge was developed has been documented to result in sharply degraded performance, even to the point that human judgment becomes inferior to naive actuarial models (e.g., Johnson, 1988; Sternberg, 1996; Tetlock, 2005; Chi, 2006). Expert mental maps have thus been described as “brittle” (Camerer and Johnson, 1991) and subject to breakdown when applied to new areas (Levenberg, 1975; Lichtenstein, Fischhoff and Phillips, 1977; Brehmer, 1980; Holland et al., 1986; Meyer, 1987; Camerer and Johnson, 1991; Chi, 2006). These findings suggest that novel approaches might be systematically “misconstrued” if uncertainty surrounding them leads them to be interpreted on the basis of existing knowledge and mental maps. If this leads to discounted evaluations, a negative relationship between evaluations and “novel” research proposals will manifest.

### 2.3 Summary and Research Questions

Intellectual distance is a regular feature of the evaluation process and deserves careful study as a variable that might influence evaluation and resource allocation in science. The theoretical perspectives reviewed above and the mechanisms they suggest, are summarized in Table 1, with predictions in relation to mean evaluations. Several points relate specifically to the case of novel departures from existing research approaches. Our main goal in this study is to test for systematic relationships between evaluation scores and intellectual distance. A secondary goal is to attempt to rule in and rule out alternative theories.

<Table 1>
3 Research Design

In this section, we describe the setting and research design, providing details on proposal generation, evaluator recruitment, random assignment and our key measures.

3.1 A Call for Research Proposals from the “First Phase” of a Grant Process

We carried out our research in the context of a scientific grant solicitation and evaluation process for research on endocrine-related disease, a major economic and health burden on society and a focus of considerable research effort at the host medical school. Working closely with grant administrators, we altered the usual grant procedures to allow us to make precise observations and to derive meaningful inferences. The grant process we studied involved seed grant awards, intended to enable investigators to initiate their research efforts to generate preliminary data (to support later NIH grant applications).

In terms of defining the scope, we deliberately defined the grant solicitation in terms of a disease area rather than making any mention of existing literature, the existing body of scientific knowledge or established research pathways. The articulated aim for the grant was otherwise stated in general terms of directing research attention and financial resources to make progress in endocrine-system related disease research, treatment and care. The content of proposals was otherwise unconstrained; we welcomed submissions related to diagnosis, treatment and prophylaxis. To attempt to draw a variety of submissions, the university president communicated an open call to participate to all members of the medical school and broader university community via email.

A fundamental research design choice was to partition the grant proposal process into two phases. The first, involving solicitation of proposals for approaches and ideas, was essentially a call for research hypotheses. It is this first phase—of defining research goals, approaches and hypotheses—that is most relevant to the questions raised earlier (Section 2). Partitioning the proposal process in this manner also reduced “entry costs” to prospective submitters, making it possible to document submissions in shorter proposals. (Average proposal length in this exercise was roughly six pages.) This design decision also allowed us to require submissions be authored by individual scientists rather than teams. Thus, we could associate each proposal with the attributes of the individual submitter. Shorter and more standardized proposal format also allowed us to minimize the extent
to which submission format shaped evaluations (Langfeldt, 2001, 2006).

Explicit incentives in this process included a $2,500 cash prize awarded to each of the top 12 winners. The process also generated additional incentives as the winning proposals would form the basis for a call for research proposals, the second phase, in which a total of $1M in seed grants would be available. Being in the top of the first phase increased the odds of being able to create a successful second stage proposal. (Indeed, four second phase winners were also first phase winners.) The first phase of the process also served as a platform for high-profile exposure among peers and university leaders, as awards being conferred by the dean of the medical school in a formal public ceremony attended by colleagues, White House staff and members of the media. This process elicited 150 research proposals, with 72 coming from within the host university.

### 3.2 Recruiting Evaluators

Major funding agencies regularly invite researchers with relevant subject knowledge to participate in evaluating research proposals (Langfeldt, 2006). An *ad hoc* evaluation team might include a few, perhaps five to seven (Langfeldt, 2006), specialized researchers whose phenomenological interest, research methods and/or topical focus relate to the research proposal(s) in question (Jayasinghe et al., 2003). More extensive evaluation processes covering large numbers and steady flows of proposals, like those employed by the NIH and NSF, often involve standing committees and subcommittees formed around topic areas to which proposals are directed, as appropriate. Such committees can be as large as 30 to 50 researchers (Li, 2013) and their identities publicly disclosed.

Given our interest in generating variation, and also abundant replication and degrees of freedom, we recruited roughly equal numbers of evaluators from among three distinct groups of host university faculty: (i) those with at least one publication in the disease area, (ii) those without publications in the particular disease area, but with at least one publication with someone with a publication in the disease domain and (iii) those without any publications or links to the disease area. Within each of these groups, we recruited equal numbers of senior and junior faculty (30 of each). We populated these six groups by rank ordering faculty at the medical school according to publication counts and inviting the top-ranked faculty from each of the three groups to participate. Drawing on faculty from the host university assured high-calibre participants, independent of rank. Strong institutional support helped minimize drop out. Of the 180 invitations (ie., six groups, times thirty
invitations per group), 142 individuals accepted and participated in the exercise. This produced roughly equal proportions, balanced across the groups in relation to both the literature and junior and senior scholars. Each group also reflects considerable diversity in gender, age and training (in terms of M.D. or Ph.D.). The group is uniform in including just highly accomplished researchers, with an average publication count of 101. Submitters are themselves accomplished, but clearly more junior, on average, with roughly one-tenth as many publications, on average.

3.3 Evaluator Assignment and the Evaluation Process

Our assignment of evaluators and proposals yielded 2,130 proposal-evaluation pair observations. Ten blocks of 15 research proposals, randomly drawn from 150 total, were randomly assigned to each of the 142 evaluators, giving an average of 14.2 randomly-selected faculty per proposal. Block randomization in this fashion was implemented to ease back office implementation of the procedure by administrators at the institution. Following convention in medical research grant proposal evaluations, the task of evaluators was to score proposals by responding to the question, “On a scale of 1 to 10 (1 Lowest - 10 Highest) please assess the impact on disease care, patients or research.”

Given our interest in having evaluators respond to the content of proposals rather than the identities of submitting researchers, we designed the process to minimize the probability of identities being revealed. Submitters’ names were blinded on proposals and evaluators, whose identities were also blinded, performed their evaluations independently and had access only to the 15 assigned proposals. Evaluators were neither given the names of, nor interacted with, other evaluators. With evaluators thus effectively blinded from one another, the overall evaluation process was “triple blinded.”

3.4 Data Collection and Variables

Our central concerns are to measure the relationship between evaluation scores and intellectual distance and to novelty in relation to existing research. We therefore devised means of measuring these key objects and identified several control variables relevant to our analysis. The data set includes evaluators’ score sheets, submitted proposals, detailed backgrounds and resumes (of those evaluators and submitters at the host university) from the host university’s database, third-party topical

---

3 We tested for and found no evidence of statistical differences across the blocks.
keyword coding of submissions and the PubMed database (an extensive database of research publications in life sciences). An overview of definitions and summary statistics for the main variables are provided in Tables 2 and 3.

| Table 2 |
|---------|

| Table 3 |
|---------|

**Evaluation Scores.** The main dependent variable, \textit{EVALUATION\_SCORE}, is the integer score from 1 to 10 (\textit{mean} = 5.7; \textit{mode} = 7; \textit{std. dev.} = 2.6) given by evaluators in response to the main scoring question. Figure 2 displays all scores assigned to each proposal. Proposals appear in descending order by average score, along the x-axis. (Average score was the basis for conferring awards.) Figure 3 also present the plus and minus of one standard deviation, as a means of highlighting the consistently wide variation in evaluations across each proposal. The patterns are consistent with considerable noise in the evaluation process. For example, dummy variables for individual research proposals explain just 26\% of variation in terms of the $R^2$ statistic; dummy variables for individual evaluators explain 19\% of variation in terms of the $R^2$ statistic.

| Figure 2 |

**Intellectual Distance between Evaluators and Research Proposals.** A first approach to measuring intellectual distance in our set-up is simply to distinguish those evaluators who have previously published within the disease domain versus those who have not, as captured by the indicator variable \textit{OUTSIDE\_DOMAIN}. We also constructed a continuous measure of intellectual distance on the basis of keywords used to describe and categorize the content of research in the life sciences, collectively referred to as “Medical Subject Heading” (MeSH) terms. This is a controlled vocabulary used by the U.S. National Library of Medicine to index articles for PubMed. MeSH keywords are assigned not by authors, but rather by professional science librarians trained specifically to perform this task. Use of this controlled vocabulary is intended to assure global and consistent assignment of keywords across the life sciences (Coletti and Bleich, 2001). We hired a professional librarian trained in standardized procedures for evaluating the content of research according to NIH National Library of Medicine (NLM) guidelines to code the proposals. We used the 2012 edition of the MeSH
set, which contains 26,579 terms. On average, proposals in our sample were assigned 12.42 MeSH terms (std. dev. = 5.42). This enabled us to represent each proposal as a vector of ones and zeroes, depending on relevant MeSH terms. We constructed analogous vectors to reflect evaluators' backgrounds, with counts of numbers of papers referring to MeSH terms. Our continuous measure of intellectual distance is then simply the angular separation or cosine between the vectors for the proposal and the evaluator, expressed as a percentile, EVALUATOR_DISTANCE. The value of 1% reflects the closest and 100% the greatest intellectual distance. We refer to “evaluator” distance in naming this variable to emphasize that distance varies in relation to evaluator-proposal pairs. Formulating the variable as a percentile lead the distribution to be uniform and also eases interpretation; coefficients can be directly read as the effect of moving from the min (1st) to max (100th) percentile. (Alternative formulations of the variable produce similar results, as noted in the analysis.)

Novel Departures of Proposals from Existing Research. Our measure of novelty is also based on the MeSH lexicon. MeSH keywords attributions are intended to capture key aspects of the research, including scientific approach, topic, methods and other key issues. To develop a measure of novelty we therefore simply looked for novelty in MeSH term combinations in relation to the existing literature. We compared the MeSH term combinations of a proposal with combinations that appear in the entire existing scientific literature, as reflected in the PubMed database. We examined all possible pairs of MeSH terms (i.e., for N terms there would be N(N − 1)/2 pairs) and determined what fraction of these pairs for a given proposal had not previously appeared in the accumulated literature. The variable is then expressed as the percentile, PROPOSAL_NOVELTY, with 1% being least and 100% most novel. We refer to “proposal” novelty in the naming of this variable to emphasize its relation to the proposal in relation to the broader stock of research rather than to any one evaluator. (Alternative formulations of the variable produce similar results, as noted in the analysis.)

Other Variables. The analysis relies most heavily on the research design’s randomization and exploitation of multiple observations per proposal and per evaluator, with a series of dummy vari-

---

4PubMed includes approximately 185 million MeSH term combinations (from a body of 26,579 unique terms) assigned to some 21 million articles published between 1855 and 2010.
ables for evaluators and proposals providing controls. We also use a series of proposal covariates as a control vector (number of words, number of references cited, number of figures, presence of an introductory section that provides context in the proposal) where we cannot use proposal dummy variables. We discuss the relevance of these covariates in the analysis to follow.

4 Main Results

Here we present our main results, estimating the relationship between evaluation scores and intellectual distance, and with proposal novelty. We report results in separate subsections, given estimates of relationships with distance and novelty require different econometric approaches.

4.1 Intellectual Distance and Evaluation Scores

The evaluation of proposal $i$ by evaluator $j$ might be shaped by a range of proposal covariates ($X_i$) (including, of course the underlying quality and merit of the proposal), evaluator covariates ($X_j$) and luck or noise, which we describe with a zero-mean error term ($\varepsilon_{ij}$). Proposal-evaluator pair characteristics can also play a role. However, having controlled, via the research design, for any relatedness of evaluators and researchers, we focus here on intellectual distance between evaluators and proposals ($EVALUATOR\_DISTANCE$). These variables relate to evaluation scores through some function $g(\cdot)$, $EVALUATION\_SCORE_{ij} = g(EVALUATOR\_DISTANCE_{ij}, X_i, X_j; \varepsilon_{ij})$. Our empirical models estimate this expression in a series of linearly separable specifications. Coefficients and robust standard error estimates are reported in Table 4.\textsuperscript{5}

We begin with a most straightforward comparison between evaluation scores of those evaluators who have conducted research within the disease domain versus those who have not. As in model (1), evaluation scores of those outside the disease domain are 0.37 points higher (s.e. = 0.12), on average. Given randomized assignment, adding proposal dummy variables, as in model (2), does not change the estimated coefficient, but reduces standard errors.\textsuperscript{6}

Apart from discrete differences, we expect effects of intellectual distance to manifest in more continuous variation. We therefore add our continuous measure, $EVALUATOR\_DISTANCE$ to the

\textsuperscript{5}Alternative specifications allowing for truncation or for the non-negative integer nature of the dependent variable, do not alter the results.

\textsuperscript{6}Of those doing research outside of the disease domain, roughly half had and half did not have, a coauthor publishing within the domain. We find no differences in evaluations between these subgroups.
As reported in model (3), we again find a positive relationship with distance, the estimated coefficient on `EVALUATOR_DISTANCE` being 1.10 (s.e. = 0.19).

<Table 4>

Importantly, using the continuous measure allows us to introduce evaluator dummy variables, as controls, and then to exploit just variation in evaluator-proposal pairs to generate estimates. Our preferred, most stringent specification thus includes dummy variables for both research proposals ($\eta$) and evaluators ($\delta$), with `OUTSIDE_DOMAIN` dropping out of the model, as follows:

\[
EVALUATION\_SCORE_{ij} = \beta \cdot EVALUATOR\_DISTANCE_{ij} + \delta_i + \eta_j + \epsilon_{ij}, \tag{1}
\]

where $\epsilon$ is a zero-mean error term. As reported in model (4), this produces a slightly smaller, but statistically unchanged, coefficient on `EVALUATOR\_DISTANCE` (0.86, s.e. = 0.33).

Therefore, there is a large positive relationship between evaluation scores and intellectual distance. Given randomization, this can be interpreted as a causal relationship. Therefore, not only do specialized experts provide more discerning evaluations, they also provide systematically lower or more critical evaluations in relation to a wider population of highly competent scientific researchers who are less proximate in their knowledge to the proposal at hand. Having defined `EVALUATOR\_DISTANCE` in terms of percentiles, we can interpret the coefficient as indicating a roughly one-point difference in score across the entire population, with varying intellectual distance, in addition to the earlier-reported 0.4 added points for by those outside the research domain. This is a large effect in comparison with the standard deviation of evaluation scores, 2.6 (or 1.7 standard deviation, if proposal and evaluator dummy variables are removed).

### 4.2 Novel Departures from Existing Research and Evaluation Scores

We now examine the relationship between evaluation scores and novel departures from the existing research. Because this reintroduces a proposal covariate, `PROPOSAL\_NOVELTY`, to the model, we can no longer exploit proposal dummy variables (as novelty is a fixed feature of a given proposal). Alternative measures, such as the simple cosine (not expressed as a percentile), Euclidean distance and simple counts of overlapping areas, produce similar results.
and cannot vary). Further, because randomization is inherently not possible in separating variation in novelty from other proposal attributes, we include a vector of precise proposal covariates, $X_j$, as follows:

\[
EVALUATION\_SCORE_{ij} = \beta \cdot EVALUATOR\_DISTANCE_{ij} + \gamma \cdot PROPOSAL\_NOVELTY_j + \delta_i + \zeta \cdot X_j + \epsilon_{ij},
\]

where we continue to control for evaluator characteristics with dummy variables, $\delta_i$. $\zeta$ is the vector of parameters to be estimated on control variables. The error term is redefined accordingly.

We control for differences in scores related to different specific fields and topics with the series of dummy variables of individual MeSH terms. We control for differences in quality with numbers of author publications and citations. We also control for a series of descriptive features of proposals. Exploiting this control vector requires that we study just the subsample of 689 proposal-evaluator pairs for which we have these control variables (i.e., submissions from within the host university) rather than our full sample of 2,130 evaluator-proposal pairs. This leaves ample degrees of freedom and the mean and variance of $EVALUATION\_SCORE$ are statistically the same in the subsample. Results are reported in Table 5.\(^8\)

Model (1) regresses evaluation scores on $PROPOSAL\_NOVELTY$ together with evaluator dummy variables and the control vector of proposal covariates. The estimated coefficient on $PROPOSAL\_NOVELTY$ is large and negative, at -2.67 (s.e. = 0.64). Given the deliberate selection of control variables, it is not surprising that most are statistically significantly related to evaluation scores. The exception is the number of words per proposal, which becomes insignificant when included with other proposal variables (but is otherwise positive and significant). The control vector is highly effective at accounting for proposal characteristics; variation explained (unadjusted $R^2 = 0.428$) is almost the same as when proposal dummy variables are included (unadjusted $R^2 = 0.475$). Therefore, introducing the long list of controls leaves little room for lingering omitted variable bias, if only because there is little omitted variation, whatsoever.

---

\(^8\)The results do not depend on whether novelty is measured as a share or absolute number of new keyword pairs, triplets or quadruplets; whether the variable measures departures from the last ten years or the entire history of the literature on the Pub Med database; or whether the model controls for the absolute numbers of keywords.
If the model is well-controlled, and there is little scope for explaining added variation in scores associated with proposal characteristics, and to the extent these characteristics omitted in the earlier model are not somehow correlated with novelty, then introducing more controls should have no effect on estimates. Model (2) re-estimates the model, adding to controls the author citations for the past seven years (in case recency of citations is salient), counts of first author publications and maximum number of citations of any one of an author’s publications. Indeed, this leaves the coefficient on \textit{PROPOSAL}_NOVELTY statistically unchanged.

As additional assessments of the possibility that omitted variable bias is the reason for the negative relationship with novelty, we also considered effects of removing controls. When examining all possible combinations of control variables (including the main control vector and in the added controls introduced in model (2)), we find that progressively adding more controls to the model generally produces more negative estimates, not less. For example, dropping control variables altogether produces a far less negative coefficient (-0.25; s.e. = 0.20). Model (3) then also introduces \textit{EVALUATOR}_DISTANCE into the model at the same time. The coefficient on proposal novelty is again unchanged despite this inclusion. Further, the coefficient on \textit{EVALUATOR}_DISTANCE is itself statistically unchanged from earlier estimates in Table 4 that used proposal fixed effects, again affirming the effectiveness of our control vector.\textsuperscript{9} (As discussed in Sections 5.3 and 5.4, there are no significant interactions between distance and novelty.) Therefore these patterns indicate that the relationship between evaluator scores and our measure of novelty is negative. It remains possible that novel proposals simply are intrinsically of lower quality, however we all but ruled out the possibility of omitted proposal characteristics producing biased estimates.

Having established the meaningfulness and stability of our specification, we also investigated whether there existed non-linearities in these relationships. In Figure 3, we present results in which we allow second-order polynomial specifications of the relationship between evaluation scores and both intellectual distance and novelty. Panel I shows no evidence of nonlinearity in the relationship with intellectual distance. Panel II shows a relationship between scores and novelty that is first positive with increasing novelty at low levels of novelty, but becomes negative at increasing levels of novelty. Alternative specifications (including higher order polynomials, nonparametric estimation

\textsuperscript{9}This is also the case despite these latter estimates being based on just the subset of authors from the host institution for whom we observe control variables.
and dividing the domain with dummy variables) are consistent with these results in Panels I and II.

5 Discussion and Interpretation

Here we discuss our results in light of the three theoretical perspectives and associated mechanisms, described earlier in Section 2.2 and summarized in Table 1.

5.1 Agency Problems and Private Interests

Recall, if agency problems play a role, then evaluators may shade assessments up or down, depending on private interests (Section 2.2.1). This view is very difficult to reconcile with patterns in the data.

Mean Responses and Bias. In contrast to this view of bias based on private interests, the coefficients on \textit{PROPOSAL\_NOVELTY} and \textit{EVALUATOR\_DISTANCE} are opposite in sign. Thus, the effect of “close” research are opposite in either case. It remains plausible that research that is both proximate and novel is perceived to be competitive and therefore a threat. However, there is no evidence of an interaction between \textit{PROPOSAL\_NOVELTY} and \textit{EVALUATOR\_DISTANCE}. Also in contrast to this view, we might expect that effects of bias would become more severe or discretely “kick-in” only among closest proposals. However, we instead documented smooth, linear relationships between scores and distance (Figure 3, Panel I).

Heterogeneous Responses across the Distribution of Evaluators. Further, if private interests and agency problems were to play a role, we might expect biases to vary both in strength and direction or sign across different evaluators. To investigate this possibility, we re-estimate the model described in expression (3), but allow the coefficient on \textit{EVALUATOR\_DISTANCE} to be heterogeneous across evaluators $\beta_i \sim N\left(\bar{\beta}, \sigma_\beta^2\right)$. Estimating this random coefficient specification, we find the estimated mean, $\bar{\beta}$, at 1.48 (s.e. = 0.41), is considerably larger than the standard deviation across the population of evaluators, 0.61 (s.e. = 0.32), indicating that the response to intellectual distance is the same in sign across the mainstream of evaluators. Figure 4 plots for each
evaluator, individual relationships with intellectual distance underlining this behavior hardly varies in strength and direction; it is nearly universal.

<Figure 4>

**Interactions and Evaluator Types.** If private interests play a role, we might also expect the effects of `EVALUATOR_DISTANCE` to vary systematically with factors associated with strength of interests or behavioral orientations. However, as reported in Table 6, we find no significant interactions with evaluator seniority (model 2), years since graduating (model 3) or gender (model 4). Neither do we detect any interactions when introducing all interaction terms at once, as in model (5).

<Table 6>

5.2 Uncertainty, Risk and Decision Theory Perspectives

If intellectual distance has the effect of producing greater uncertainty, this could plausibly product a noiser “signal” and greater dispersion and variance or, alternative, lead to greater risk discounting (Section 2.2.2). This view is also difficult to reconcile with patterns in the data.

**Dispersion and Variance.** To investigate the possibility of greater dispersion and noise with intellectual distance, we re-estimate the earlier model, while allowing the model error term to vary with a multiplier, `m`, i.e., `m_{ij} \cdot \epsilon_{ij}`, where this multiplier is allowed to vary with our key explanatory measures: `m_{ij} = 1 + \beta^e \cdot EXPERT\_DISTANCE + \gamma^e \cdot PROPOSAL\_NOVELTY_j`. We simultaneously estimate all model coefficients with maximum likelihood. We find coefficients in the conditional mean model to be unchanged and coefficients in the multiplier to be statistically indistinguishable from zero (`\beta^e = -0.21`, s.e. = 0.17; `\gamma^e = -0.19` s.e. = 0.16). Therefore, we find no evidence consistent with growing dispersion with distance.

**Uncertainty Discounting.** It is also difficult to reconcile the results with greater uncertainty discounting. The relationship with `EVALUATOR\_DISTANCE` is positive, rather than negative.  

---

10 Assessing all other possible combinations of second order polynomials in both conditional mean and error term models and exploiting either proposal fixed effects or the proposal control vector, does not alter this result.
Only the relationship with \textit{PROPOSAL\_NOVELTY} is negative. If there is no general discounting with risk and uncertainty, it remains plausible that ambiguity aversion plays a role in accounting for the negative relationship with novelty (Section 2.2.2).

5.3 Bounded Rationality and Expert Cognition Perspectives

A bounded rationality characterization of the evaluation process implies that those with most relevant knowledge—experts—will be better able to discern informational cues, not just “seeing” a research proposal with less noise, but seeing more informational cues. This should lead to systematic differences in evaluations, depending on whether experts are differentially better than non-experts at recognizing additional merits or additional demerits (Section 2.2.3). (It is only in that case that added expertise leads to “sampling” and weighting equally on good and bad informational cues that experts will produce the same evaluation.) A bounded rationality perspective also suggests that established knowledge and mental models are “brittle” and lead to systematic errors and miscontrual of new knowledge and ideas, when those new knowledge and ideas are construed through extrapolation from old knowledge and mental models beyond the domain for which they were developed (Section 2.2.3). Here, we find the data to be consistent with these ideas.

The positive relationship between evaluations and \textit{EVALUATOR\_DISTANCE} is consistent with experts perceiving and processing a wider set of informational cues and for this wider set to include more critical assessments, on average. The pattern of more critical assessments of closer proposals is consistent with the merits and intended contributions of a research proposal being more easy to perceive and recognize than the range of pitfalls and problems that could exist in the details of a research design. This might also be thought of as experts applying more extensive “tests,” and, on balance, uncovering more errors, problems and limitations. Among explanations for a relationship between evaluation scores and evaluator distance, as in Section 2, it is only this bounded rationality explanation of the role of experts’ more discerning inspection that is consistent with observed patterns.

The negative relationship between evaluations and \textit{PROPOSAL\_NOVELTY} is consistent with departures from the existing literature being discounted, possibly on the basis of being misconstrued given they involve extrapolations beyond existing knowledge. This is also consistent with the finding that the relationship is reversed—positive—for low levels of novelty. Some level of novelty should be
required for research to be regarded as promising, accounting for the positive relationship between evaluations and novelty, at least at very low levels of the latter. Further, we might expect any misconstrual and extrapolation to not play a major role at such low levels of novelty, and for the brittleness of mental models and existing paradigms to become more relevant only with more significant departures.

The “smoothness” and gradual changes (in the case of novelty) and linearity (in the case of intellectual distance), as in Figure 3, are themselves consistent with the role of bounded rationality, given that bounded rationality effects might be expected to manifest progressively and as a matter of degree with growing distance rather than to appear very sharply and suddenly. Further, constant variance or error with intellectual distance (Section 5.2) and the similarity of responses across the broad cross-section of evaluators (Figure 4) are consistent with the universality of bounded rationality-related effects.

Therefore, among explanations for a relationship between evaluation scores and evaluator distance, as in Section 2, this bounded rationality explanation of the problems of extrapolating from old knowledge to interpret new knowledge is consistent with observed patterns, as is the possibility of discounting on the basis of ambiguity aversion (see Section 5.2). Further, while we all but certainly ruled out omitted variable bias in measuring the relationship between evaluation scores and novelty (Section 4.2), it remains possible that novel proposals are inherently of lower expected quality, by the very nature of being novel. The combination of these explanations–miscontrual (Sections 2.2.3 and 5.3), fundamental uncertainty and ambiguity aversion (Sections 2.2.2 and 5.2) and lower expected quality (Section 5.1)–all plausibly even coexist as mechanisms explaining the negative relationship between evaluation scores and novelty. We are not able to discern among them with these data. Bounded rationality perspectives, including the possibility of misconstrual of novelty and the differential assessment by experts of proximate research–have the advantage of being able to account for both key relationships documented in Sections 5.1 and 5.2.

6 Policy Implications of Managing Intellectual Distance

We now turn to policy implications of our findings and the challenge of selecting among competing innovation projects. We consider a number of counterfactuals in an effort to understand the impact
of varying expertise and novelty and evaluation policies.

6.1 Managing Expertise Distance

Experts provide more discriminating evaluations (i.e., on average they will provide more meaningful rank-orderings). Because of this, it is inherently attractive to seek the evaluation of the closest expert. However, relying on just one evaluator’s evaluation can lead to considerable idiosyncratic “noise”. Moreover, relying on individual evaluators may lead to different evaluators being assigned to different proposals. Averaging multiple evaluations from larger sets of evaluators would seem to be a remedy; errors and differences across individuals can be averaged out. This, however, introduces a distinct set of challenges. A larger group of evaluators with varying distance and expertise to a proposal will contain individuals with varying abilities to discern true rank order. Further, evaluators of varying distances will also be inherently more and less critical—another source of variation in scores that is unrelated to underlying true quality of proposals.

To better calibrate and compare the problems created by groups of evaluators versus those of closest experts, we compare group scores with those of closest experts within groups of 15 evaluating each proposal—simulating outcomes under expert evaluation versus group evaluation. Panel I of Figure 5 presents the differences between the scores given by the closest experts and the group average. We plot these differences in relation to the final (average) rank order used in evaluations in this exercise. The fitted mean of differences between expert and averaged scores is negative, with expert evaluators assigning 7% lower scores on average—consistent with more critical evaluations by experts. Far more striking than differences in means are differences in in the scoring of individual proposals by experts and groups and resulting changes in rank order. Panel II of Figure 6 plots ranks that would be given by experts versus the group average. Individual research proposal ranks would change by a staggering 31.8 rank positions (std. dev. = 26.0)—on average—if ranked by closest experts. Even among the top 25 ranks, the mean absolute change in rank order would be 23.8 positions (std. dev. = 27.0).

To provide an indication of whether these large differences between experts and group evaluations are more reflective of idiosyncratic errors of experts or garbling of groups, we re-plot the comparisons after taking out individual fixed effects for experts (i.e., means across each of their assessments)

\[11\text{Mean EVALUATOR\_DISTANCE is 0.13 for the closest experts, 0.50 for groups.}\]
and also linearly “correcting” for differences in intellectual distance between these expert evaluators and proposals, as in Panel II of Figure 5 (i.e., using the earlier regression model results to virtually set intellectual distance to be the same among closest experts). As can be seen in the figure, these attempts to correct for idiosyncratic errors of closest experts lead to widening differences in assessments among higher quality proposals (i.e., the left side of the dashed line being shifted away from the 45 degree line). Among lower quality proposals, these corrections lead to closest experts and group evaluations to become more similar (i.e., the right side of the dashed line being shifted towards the 45 degree line). That closest experts diverge increasingly from group averaged evaluations once their own idiosyncratic errors (i.e., noise) are reduced is consistent with closest experts providing more discriminating evaluations among top proposals where they are able to detect more subtle differences that distinguish ostensibly high quality proposals. That closest experts converge to group averaged assessments once their own idiosyncratic errors are reduced is consistent with expert judgment being less of an advantage when judging coarser, larger differences in quality.

<Figure 5>

Given these conditions, it appears that the tradeoff between idiosyncratic errors of closest experts and garbling of information in group averaged evaluations is especially unproductive among top, highest-quality proposals where differences in quality and potential are subtle and the need for expert judgement and small errors is higher. One possible approach to addressing these problems, given ample data, is to apply adjustments and algorithms akin to those used in this study as a means of better extracting signals of quality. Where such data is not available, senior evaluators who review the evaluations of other evaluators might play a special role in discerning evaluator “fixed effects” (perhaps informally or instinctively) when aggregating insights across evaluators. Equally important, senior evaluators will process evaluations in a manner far more sophisticated than the simple averaging. An effective senior evaluator therefore does not simply average evaluations or “tally votes” from evaluators, but rather he or she effectively uses experience, judgement and assimilation of signals to discern meaningful assessments.
6.2 Managing Proposal Novelty

Whatever the reason for novelty discounting (miscontrual and bounded rationality, ambiguity aversion or lower expected quality or some combination of these things) there may be instances in which more novel proposals may be sought to initiate a wider “search” of the knowledge frontier (e.g., Levinthal, 1997; Fleming, 2001; Uzzi et al., 2013). Unfortunately, here are are less clear remedies to countervailing a systematic tendency to lower evaluations for novel proposals. Our results and analysis suggest that these are not issues that double blinding, averaging of multiple opinions or expert discernment can clearly address. Avenues perhaps deserving further consideration include priming of and coaching of evaluators to create greater metacognition and awareness of resource allocation goals and cognitive limits or behavioral biases that can enter into assessments. This might be supplemented with more analytically informed evaluation processes (such as reporting measures of departures from the existing body of research, as we have done here). Such approaches could at least lead to more explicit consideration of the question of novelty. Programs geared to providing researchers with less stringent constraints in allocating resources might also play a role in fostering novel innovation.

7 Summary and Conclusions

Analyzing data from a medical research grant evaluation process, we found evaluators to systematically assign lower scores to research closer to their own area of expertise. The effects are established through random assignment and therefore can be interpreted as causal. In the range of variation observed here, effects are quite large, a one point or more difference (on a 10-point scale). Effects of varying intellectual distance affect the wide cross-section of evaluators. We found no evidence of changing variance or magnitude of “errors” with varying intellectual distance. We emphasize that these effects of proximity of expertise are not the result of an “expert” versus a lay person, but rather of differences in precise areas of specialization within a group of world-leading medical researchers. These patterns are consistent with a bounded rationality perspective in which experts see more and “sample” greater informational cues than non-experts, and added informational cues are disproportionately related to subtle demerits and limitations of proposals that are more difficult for non-experts to recognize than are the intended merits and contributions of a research study.
The patterns are inconsistent with agency problems in which evaluators bias evaluations in relation to their private interests. The patterns are also inconsistent with evaluations merely becoming more noisy with greater distance.

When comparing evaluations of closest experts and (less expert) group averages, we find this produces a tradeoff between the idiosyncratic errors of the assessment of a lone expert versus the “garbling” of relevant rank ordering signals from groups of evaluators. We found that this tradeoff and challenges in evaluation were especially challenging in establishing an appropriate rank-ordering of proposals among the highest quality proposals. We suggested a number of potential policies to reduce the problems.

We found proposals with large novel departures from the existing body of research are associated with lower evaluations. The size of relationships is large and comparable, in these data, to that of varying proximity of expertise. The patterns are again consistent with a bounded rationality perspective, however in this case, one related to the “brittle” nature of mental models and existing knowledge when extrapolating to evaluate new knowledge and approaches. The patterns are also plausibly consistent with novel proposals being discounted on the basis of ambiguity aversion. Although our estimates of the relationship between evaluation scores and novelty all but rule out the possibility of omitted variable bias, it also remains possible that novel proposals are inherently of lower expected quality. We cannot discern among these explanations in these data. We speculate they may coexist. There are greater challenges in finding policy remedies for promoting novel research proposals (where this is indeed the policy objective), on account of the nature of the relationship of evaluations with novelty; we listed several possibilities deserving investigation.

This work complements several decades of research that has theorized a range of influences that might operate within scientific evaluation and peer review processes (e.g., Cole, Cole and Simon, 1981; Chubin and Hackett, 1990; Lee et al., 2013). Our paper relates, in particular, to a handful of studies that have attempted to establish causal inferences (e.g., McNutt et al., 1990; van Rooyen et al. 1999; Li, 2013). This research might also more broadly relate to wider traditions of research on project evaluation (e.g., Link and Long, 1981; Astebro and Elhedhli, 2006; Hallen, 2008; Stephan, 2012; Xie and Killewald, 2012; Jeter and Albar, 2013; Piezunka and Dahlander, 2014). The present work, however, differs in focusing on how the “structure of knowledge” and “positions in intellectual space” systematically shape variations and consequent resource allocation in science. Such effects are
of particular interest, given that they may be insensitive to customary double blinding procedures or averaging of large numbers of evaluations. Further, given intellectual distance is a stable feature of scientific and innovation evaluation processes, these effects help explain the recurring patterns of resource allocation and knowledge advance and the overall trajectory of scientific developments (e.g., Kuhn, 1962).
REFERENCES

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

Anand, P. 1993. Foundations of Rational Choice Under Risk. Oxford: Oxford University Press.

Bedard, J. 1989. Expertise in auditing: Myth or reality? Accounting, Organizations and Society, 14(1), 113–131.

Ben-David, J. 1960. Roles and innovations in medicine. American Journal of Sociology 65(6), 557–558.

Berger, J. 1985. Statistical Decision Theory and Bayesian Analysis (2nd ed.). New York: Springer-Verlag.

Blackburn, J.L., & Hakel, M.D. (2006). An examination of sources of peer-review bias. Psychological Science, 17(5), 378–382.

Bormann, L., & Daniel, H.-D. 2008. The effectiveness of the peer review process: Inter-referee agreement and predictive validity of manuscript refereeing at Angewandte Chemie. Angewandte Chemie International Edition, 47(38), 7173-7178.

Boudreau, K.J., & Lakhani K.R. Forthcoming. “Open” Disclosure of Innovations, Incentives and Follow-on Reuse: Theory on Processes of Cumulative Innovation and a Field Experiment in Computational Biology. Research Policy.

Bouman, M. J. 1980. Application of information-processing and decision-making research. In G. R. Ungson and D. N. Braunstein (Eds.), Decision Making: An Interdisciplinary Inquiry. Boston, MA: Kent, 129-167.

Brehmer, B. 1980. In one word: Not from experience. Acta Psychologica, 45(1), 223–241.

Campanario, J. M., & Acedo, E. (2007). Rejecting highly cited papers: The views of scientists who encounter resistance to their discoveries from other scientists. Journal of the American Society for Information Science and Technology, 58(5), 734–743.

Camerer, C. F., & Johnson, E. J. 1991. The process-performance paradox in expert judgment: How can experts know so much and predict so badly? In K. A. Ericsson and J. Smith (Eds.), Toward a General Theory of Expertise: Prospects and Limits. Cambridge: Cambridge University Press, 195-217.

Chase, W. G., & Simon, H. A. 1973. Perception in chess. Cognitive Psychology, 4(1), 55-81.

Chi, M. T. H. 2006. Two approaches to the study of experts characteristics. In Ericsson, K. A., Charness, N., Felstovich, P. J., & Hoffman, R. R. (Eds), The Cambridge Handbook of Expertise and Expert Performance. New York, NY: Cambridge University Press, 21-30.

Chi, M. T. H., Felstovich, P., & Glaser, R. 1981. Categorization and representation of physics problems by experts and novices. Cognitive Science, 5, 121-152.

Chubin, D. E., & Hackett, E. J. 1990. Peerless Science: Peer Review and U.S. Science Policy. Stony Brook, NY: State University of New York Press.

Cole, S., J. R. Cole, G. A. Simon. 1981. Chance and consensus in peer review. Science, 214(4523), 881-886.

Coletti, M. H., & Bleich, H. L. 2001. Medical subject headings used to search the biomedical literature. Journal of the American Medical Informatics Association, 8(4), 317-323.

Cowan, R., David, P. A., & Foray, D. 2000. The explicit economics of knowledge codification and tacitness. Industrial and Corporate Change, 9(2), 211-253.

Dasgupta, P., & David, P. A. 1994. Toward a new economics of science. Research Policy, 23(5), 487–521.

Dogan, M., R. Pahe. 1990. Creative Marginality: Innovation at the Intersection of Social Sciences. Westview Press, Boulder, CO.

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: evidence on maximal adaptations on task constraints. Annual Review of Psychology, 47(7), 273–305.
Ericsson, K. A., & Smith, J. (Eds.). 1991. Toward a General Theory of Expertise: Prospects and Limits. Cambridge University Press.

Fitts, P. M., & Posner, M. I. 1967. Learning and Skilled Performance in Human Performance. Belmont, CA: Brock-Cole.

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.

Fox, C. R., & Tversky, A. 1995. Ambiguity aversion and comparative ignorance. Quarterly Journal of Economics, 110(3), 585-603.

Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., & Trow, M. 1994. The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies. London, UK: Sage.

Graves, N., Barnett, A. G., & Clarke, P. 2011. Funding grant proposals for scientific research: retrospective analysis of scores by members of grant review panel. British Medical Journal, 343: d4797.

Hallen, B. L. 2008. The causes and consequences of the initial network positions of new organizations: From whom do entrepreneurs receive investments? Administrative Science Quarterly, 53(4), 685-718.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. 1986. Induction: Processes of Inference, Learning and Discovery. Cambridge, MA: MIT Press.

Jackson, J. L., Srinivasan, M., Rea, J., Fletcher, K. E., & Kravitz, R. L. 2011. The validity of peer review in a general medicine journal. PLoS ONE, 6(7), e22475.

Jayasinghe, U. W., Marsh, H. W., & Bond, N. 2003. A multilevel cross-classified modeling approach to peer review of grant proposals: The effects of assessor and researcher attributes on assessor ratings. Journal of the Royal Statistical Society. Series A (Statistics in Society), 166(3), 279-300.

Jetter, A., & Albar, F. M. 2013. Fast and Frugal Heuristics for New Product Screening—is managerial judgment good enough? International Journal of Management and Decision Making, 12(2), 165-189.

Johnson, E. J. 1988. Expertise and decision under uncertainty: Performance and process. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), The Nature of Expertise. Hillsdale, NJ: Erlbaum, 209-228.

Johnson, E. J., & Payne, J. W. (1985). Effort and Accuracy in Choice. Management Science, 31(4), 395-414.

Johnson, E. J., & Russo, J. E. 1984. Product familiarity and learning new information. Journal of Consumer Research, 11(1), 542–550.

Johnson, P. E., Hassebrock, F., Duran, A. S., & Moller, J. H. 1982. Multimethod study of clinical judgment. Organizational Behavior and Human Performance, 30(2), 201-230.

Jones, B. F. 2009. The burden of knowledge and the death of the renaissance man: Is innovation getting harder? Review of Economic Studies, 76(1), 283-317.

Jones, B. F. (2010). Age and great invention. Review of Economics and Statistics, 92(1), 1-14.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). 1982. Judgment under uncertainty: Heuristics and biases. Cambridge University Press.

Knight, F. 1921 Risk, Uncertainty, and Profit. Boston, MA: Hart, Schaffner & Marx.

Kravitz, R. L., Franks, P., Feldman, M. D., Gerrity, M., Byrne, C., & Tierney, W. M. 2010. Editorial peer reviewers? recommendations at a general medical journal: Are they reliable and do editors care? PLoS ONE, 5(4), e10072.

Kuhn, T. S. 1962. The Structure of Scientific Revolutions: 50th Anniversary Edition. Chicago,
Lamont, M. 2009. How Professors Think: Inside the Curious World of Academic Judgment. Cambridge, MA: Harvard University Press.

Langfeldt, L. 2001. The decision-making constraints and processes of grant peer review, and their effects on the review outcome. Social Studies of Science, 31(6), 820-841.

Langfeldt, L. 2006. The policy challenges of peer review: Managing bias, conflict of interests and interdisciplinary assessments. Research Evaluation, 15(1), 31-41.

Larkin, J., McDermott, J., Simon, D. P., & Simon, H. A. 1980. Expert and novice performance in solving physics problems. Science, 208(4450), 1335-1342.

Law, J. 1973. The development of specialties in science: The case of x-ray protein crystallography. Social Studies of Science, 3(3), 275–303.

Lee, C. J. 2012. A Kuhnian critique of psychometric research on peer review. Philosophy of Science, 79(5), 859–870.

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.

Levenberg, S. B. 1975. Professional training, psychodiagnostic skill, and kinetic family drawings. Journal of Personality Assessment, 39(4), 389–393.

Levinthal, D. 1997. Adaptation on Rugged Landscapes. Management Science, 43(7): 934-950.

Li, D. 2013. Information, bias, and efficiency in expert evaluation: Evidence from the NIH. Unpublished manuscript.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. 1977. Calibration of probabilities: The state of the art. In Jungerman, H., & De Zeeuw, G. Decision Making and Change in Human Affairs. Springer Netherlands (275–324).

Link, A. N., & Long, J. E. 1981. The simple economics of basic scientific research: a test of Nelson’s diversification hypothesis. Journal of Industrial Economics, 30(1), 105-109.

Marsh, H. W., Bond, N., & Jayasinghe, U. W. 2007. Peer review process: Assessments by applicant-nominated referees are biased, inflated, unreliable and invalid. Australian Psychologist, 42, 33-38.

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), 160-168.

McNutt, R., A. Evans, R. Fletcher, S. Fletcher. 1990. The Effects of Blinding on the Quality of Peer Review: A Randomized Trial. Journal of the American Medical Association. 263(10), 1371-1376.

Merton, R. K. 1968. The Matthew Effect in science. Science, 159(3810), 56-63.

Meyer, R. J. 1987. The learning of multiattribute judgment policies. Journal of Consumer Research, 14(2), 155-173.

Mollick, Ethan R. and Nanda, Ramana, Wisdom or Madness? Comparing Crowds with Expert Evaluation in Funding the Arts (May 29, 2014). Harvard Business School Entrepreneurial Management Working Paper No. 14-116.

Mullins, N. 1972. The development of a scientific specialty: The phage group and the origins of molecular biology. Minerva 10(1) 51–82.

Nelson, R. R. 1961. Uncertainty, learning, and the economics of parallel research and development efforts. Review of Economics and Statistics, 43(4), 351-364.

Nelson, R. R., & Winter, S. G. (1982). An Evolutionary Theory of Economic Change. Boston, MA: Belknap.

Newell, A., H. A. Simon. 1972. Human Problem Solving. Engelwood Cliffs, NJ: Prentice-Hall.

van Rooyen, S., Godlee, F., Evans, S., Black, N., & Smith, R. 1999. Effect of open peer review on quality of reviews and on reviewers? recommendations: A randomized trial. British Medical
Journal, 318(7175), 23-27.

Rothwell, P. M., & Martyn, C. N. 2000. Reproducibility of peer review in clinical neuroscience — Is agreement between reviewers any greater than would be expected by chance alone? Brain, 123(9), 1964–1969.

Schilling, M.A. & Green, E. 2011. Recombinant search and breakthrough idea generation: An analysis of high impact papers in social sciences. Research Policy, 40(10), 1321-1331.

Simonton, D.K. 1995. Foresight in insight? A Darwinian answer, in R.J. Sternberg, J.E. Davidson (Eds) The Nature of Insight. MIT Press, Cambridge, MA, 465-494.

Simonton, D. K. (1999). Origins of Genius: Darwinian Perspectives on Creativity. New York City, NY: Oxford University Press.

Stephan, P. E. 1996. The economics of science. Journal of Economic Literature, 34(3), 1199–1235.

Stephan, P. E. 2012. How Economics Shapes Science. Cambridge, MA: Harvard University Press.

Sternberg, R. J. 1996. Costs of expertise. In K. A. Ericsson (Ed.), The Road to Excellence: The Acquisition of Expert Performance in the Arts and Sciences, Sports, and Games. Hillsdale, NJ: Erlbaum, 347-354.

Stevens, G. A., & Burley, J. 1997. 3,000 raw ideas = 1 commercial success! Research Technology Management, 40(3), 16-27.

Sugimoto, C. and B. Cronin. 2013. Citation gamesmanship: testing for evidence of ego bias in peer review. Scientometrics 95(3): 851-862.

Taatgen, N. A., Huss, D., Dickison, D., & Anderson, J. R. 2008. The acquisition of robust and flexible cognitive skills. Journal of Experimental Psychology: General, 137(3), 548-565.

Tetlock, P.E. 2005. Expert Political Judgement. Princeton, NJ: Princeton University Press.

Travis, G. D. L., & Collins, H. M. 1991. New light on old boys: Cognitive and institutional particularism in the peer review system. Science, Technology, & Human Values, 16(3), 322-341.

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

Warren, R.J., & Marshall, B. 1983. Unidentified curved bacilli on gastric epithelium in active chronic gastritis. The Lancet, 321(8336), 1273-1275.

Weitzman, M. L. (1998). Recombinant Growth. Quarterly Journal of Economics, 113(2), 331–360.

Wuchty, S., Jones, B. F., & Uzzi, B. 2007. The increasing dominance of teams in production of knowledge. Science, 316(5827), 1036-1039.

Xie, Y. & Killewald, A. A. 2012. Is American Science in Decline? Cambridge, MA: Harvard University Press.
FIGURES

Figure 1. Time Trend of Cumulative Numbers of Publications, Unique Journals, and Unique Pairs of Keyword Topics and Article Counts.

Note. Based on data from PubMed database. Keywords are based on standardized lexicon (MeSH terms).

Figure 2. Evaluation Scores for Each Proposal, Ordered By Mean Scores (mean and +/-1 standard deviation shown).

Note. Individual integer scores are vertically randomly “jittered” to avoid overlap.
Figure 3. First- and Second-Order Polynomial Relationships

Note. 90% confidence intervals shown.
Figure 4. Fitted Linear Relationships for Individual Evaluators

Note. Quantile and mean fitted lines are also shown to provide additional perspective on the distribution of data; each is regressed as a second-order polynomial.

Figure 5. Counterfactual Evaluations based on Closest Expert versus Average Group Assessments
## TABLES

Table 1 Alternative Theoretical Mechanisms Possibly Relating Intellectual Distance to Evaluations

| Theoretical Perspective | Mechanism                                                                 | Predicted Relationship of Intellectual Distance to Mean Evaluation | Predicted Relationship of Novelty to Mean Evaluation | Predicted Relationship with Variance of Evaluations |
|--------------------------|---------------------------------------------------------------------------|-------------------------------------------------------------------|-----------------------------------------------------|-----------------------------------------------------|
| 2.2.1 Uncertainty, Risk and Decision Theory | • Distance, Uncertainty and Dispersion  
• Discounting and risk adjustments  
• Discounting and ambiguity aversion | (-)  
(-) | (-) | (+) |
| 2.2.2 Agency Problems and Private Interests | • Promotion of one’s own work or “schools of thought” or “protecting” existing approaches  
• Discounting competing research | (-) | (+) | |
| 2.2.3 Bounded Rationality and Expert Cognition | • More discerning and extensive assessments and tests by experts  
• Systematic errors when using existing models to extrapolate to new domains | (+) | | (-) |
Table 2 Definitions of Main Variables

| Variable         | Description                                                                                                                                 |
|------------------|----------------------------------------------------------------------------------------------------------------------------------------------|
| (1) EVALUATION_SCORE | Main integer score from 1 to 10 given by an evaluator to a research proposal                                                                 |
| (2) OUTSIDE_DOMAIN       | Indicator switched to one for those evaluators who have not previously published on endocrine-related disease                                 |
| (3) EVALUATOR_DISTANCE   | With evaluators and research proposals each represented as vectors of (MeSH term) keywords, this variable is the cosine of the angle between the vectors, expressed as a percentile (1% to 100%) |
| (4) PROPOSAL_NOVELTY     | Of all the keywords (MeSH term) used to describe a research proposal, the share of these terms not yet observed in prior published research, expressed as a percentile (1% to 100%) |
| (5) WORDS              | Total number of words in main text of each proposal                                                                                         |
| (6) NUM_REFS           | Total number of references listed in each proposal                                                                                           |
| (7) NUM_FIGS           | Total number of figures shown in each proposal                                                                                               |
| (8) INTRO_SECTION      | Indicator switched to one for those proposals which begin with an overview or introduction section                                             |
| (9) AUTHOR_PUBS        | Count of all publications of the researcher submitting the research proposal                                                                |
| (10) AUTHOR_CITES      | Count of all citations of prior publications of the researcher submitting the research proposal                                               |

Table 3 Means, Standard Deviations and Correlations of Main Variables

| Variable         | Mean | Std. Dev. | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|------------------|------|-----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| (1) EVALUATION_SCORE | 5.69 | 2.58      |     |     |     |     |     |     |     |     |     |
| (2) OUTSIDE_DOMAIN       | .65  | .48       | .03 |     |     |     |     |     |     |     |     |
| (3) EVALUATOR_DISTANCE   | .50  | .29       | .15 | .00 |     |     |     |     |     |     |     |
| (4) PROPOSAL_NOVELTY     | .50  | .29       | -.03| -.01| .10 |     |     |     |     |     |     |
| (5) WORDS              | 1366 | 2489      | .02 | .01 | .03 | .12 |     |     |     |     |     |
| (6) NUM_REFS           | 5.61 | 9.76      | .11 | .00 | .08 | .02 | .40 |     |     |     |     |
| (7) NUM_FIGS           | .28  | .84       | .04 | -.01| .05 | .01 | .35 | .55 |     |     |     |
| (8) INTRO_SECTION      | .21  | .41       | .05 | -.06| .04 | .16 | .07 | .07 |     |     |     |
| (9) AUTHOR_PUBS        | 9.13 | 24.01     | .03 | -.01| .12 | -.07| .00 | .00 | -.07| .23 |     |
| (10) AUTHOR_CITES      | 99   | 521       | .07 | -.01| .16 | .08 | .05 | -.03| -.06| .31 | .90 |
Table 4 Estimated Relationship between Evaluations (EVALUATOR_SCORE) and Intellectual Distance between Evaluators and Research Proposals (EVALUATOR_DISTANCE)

|                      | 1                  | 2                  | 3                  | 4                  |
|----------------------|--------------------|--------------------|--------------------|--------------------|
|                      | Outside of Disease Domain | Control Evaluator Chars. | Continuous Measure of Distance | Control Evaluator & Proposal Chars. |
| OUTSIDE_DOMAIN       | .37*** (.12)       | .37*** (.10)       | .36*** (.10)       |                    |
| EVALUATOR_DISTANCE   |                    | 1.10*** (.19)      | .86*** (.33)       |                    |
| Evaluator Dummies    | Y                  | Y                  | Y                  |                    |
| Research Proposal Dummies |                  |                    |                    | Y                  |
| Adj-R^2              | .004               | .263               | .275               | .475               |

Notes. *, **, and *** indicate statistical significance at the 10%, 5% and 1% levels, respectively; heteroskedasticity-autocorrelation robust standard errors are reported; number of observations = 2,130 research proposal-evaluator pairs.
| Table 5 Estimated Relationships between Evaluations and Proposal Novelty |
|-------------------------------------------------------------|
| **Dependent Variable:** EVALUATION_SCORE                     |
| Evaluator Dummies & Proposal Control Vector | Extended Proposal Controls | Distance and Novelty |
| PROPOSAL_NOVELTY | -.267*** | -.310*** | -.280*** |
|                  | (.64)    | (.89)    | (.64)    |
| EVALUATOR_DISTANCE | 1.48**  |           | (.59)    |
| Evaluator Dummies | Y        | Y        | Y        |
| Researcher Quality |           |           |           |
| AUTHOR_PUBS | -.15*** | -.14*** | -.15*** |
|               | (.03)    | (.03)    | (.03)    |
| AUTHOR_CITES | .005*** | -.02     | 0.006*** |
|               | (.00)    | (.01)    | (.00)    |
| Extended Set of Controls(1) | Y        |           |           |
| Research Type |           |           |           |
| Keyword (Topic) Dummies | Y        | Y        | Y        |
| Number of Keywords | Y        | Y        | Y        |
| Proposal Characteristics |           |           |           |
| WORDS | .00 | .00 | .00 |
|       | (.00) | (.00) | (.00) |
| NUM_REFS | .10*** | .04 | .10*** |
|        | (.04) | (.04) | (.04) |
| NUM_FIGS | -1.12** | -1.14** | -1.18** |
|        | (.45) | (.57) | (.45) |
| INTRO_SECTION | 1.85*** | 1.35*** | 1.94*** |
|        | (.41) | (.50) | (.41) |
| Adj-R² | .423 | .459 | .428 |

Notes. *, **, and *** indicate statistical significance at the 10%, 5% and 1% levels, respectively; heteroskedasticity-autocorrelation robust standard errors are reported; Number of observations = 689 proposal-evaluator pairs and pertain only to submitting researchers from within the host university.
Table 6 Interactions between Evaluator Distance and Factors Plausibly Influencing Incentives and Behaviors

|                          | Dependent Variable: EVALUATION_SCORE | 1       | 2       | 3       | 4       | 5       |
|--------------------------|--------------------------------------|---------|---------|---------|---------|---------|
| EVALUATOR_DISTANCE       |                                      | 1.79**  | 1.88*** | 1.64*** | 1.41**  | 2.09**  |
|                          |                                      | (.78)   | (.72)   | (.58)   | (.64)   | (.98)   |
| DISTANCE X NOVELTY       |                                      | -.38    |         |         |         | -.55    |
|                          |                                      | (1.06)  |         |         |         | (1.04)  |
| DISTANCE X SENIOR        |                                      | -.49    |         |         |         | -.59    |
|                          |                                      | (.74)   |         |         |         | (.74)   |
| DISTANCE X YEARS SINCE GRAD |                                  | .00     |         |         | .00     |         |
|                          |                                      | (.00)   |         |         | (.00)   |         |
| DISTANCE X FEMALE        |                                      |         |         |         | .35     | .44     |
|                          |                                      |         |         |         | (.75)   | (.74)   |
| Evaluator Dummies        | Y                                    | Y       | Y       | Y       | Y       | Y       |
| Research Proposal Dummies| Y                                    | Y       | Y       | Y       | Y       | Y       |
| Adj-R²                   | .482                                 | .482    | .482    | .482    | .480    |

Notes. *, **, and *** indicate statistical significance at the 10%, 5% and 1% levels, respectively; heteroskedasticity-autocorrelation robust standard errors are reported; Number of observations = 689 proposal-evaluator pairs and pertain only to submitting researchers from within the host university.