http://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Articles in Advance, pp. 1–19
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

The Risk of Caution: Evidence from an Experiment

Richard T. Carson,[a] Joshua Graff Zivin,[a,b] Jordan J. Louviere,[c] Sally Sadoff,[d] Jeffrey G. Shrader[e,*]

[a] Department of Economics, University of California, San Diego, La Jolla, California 92093; [b] School of Global Policy and Strategy, University of California, San Diego, La Jolla, California 92093; [c] School of Marketing, University of South Australia, Adelaide, South Australia 5000, Australia; [d] Rady School of Management, University of California, San Diego, La Jolla, California 92093; [e] School of International and Public Affairs, Columbia University, New York, New York 10027
*Corresponding author

Contact: rcarson@ucsd.edu (RTC); jgraffzivin@ucsd.edu, https://orcid.org/0000-0002-0820-4900 (JGZ); jordan.louviere@unisa.edu.au (JJL); ssadoff@ucsd.edu (SS); jgs2103@columbia.edu, https://orcid.org/0000-0002-8981-0450 (JGS)

Received: March 4, 2020
Revised: June 10, 2021
Accepted: October 5, 2021
Published Online in Articles in Advance: February 8, 2022

https://doi.org/10.1287/mnsc.2021.4292

Abstract. Innovation is important for firm performance and broader economic growth. However, breakthrough innovations necessarily require greater risk taking than more incremental approaches. To understand how managers respond to uncertainty when making research and development decisions, we conducted experiments with master's degree students in a program focused on the intersection of business and technology. Study participants were asked to choose whether to fund hypothetical research projects using a process that mirrors real-world research and development funding decisions. The experiments provided financial rewards that disproportionately encouraged the choice of higher-risk projects. Despite these incentives, most participants chose lower-risk projects at the expense of projects more likely to generate a large payoff. Heterogeneity analysis and additional experimental treatments show that individual risk preferences predict greater tolerance of high-risk projects and suggest that more appropriate decision making can be learned. Thus, for firms seeking to fund breakthrough research and development, appropriate screening and training of employees may play important roles in increasing the likelihood of success.

History: Accepted by Gustavo Manso, finance.
Funding: This work was supported by the National Science Foundation [Grant SBE-1561257].
Supplemental Material: The data files and online appendix are available at https://doi.org/10.1287/mnsc.2021.4292.

Keywords: research and development • innovation • decision analysis • risk • organizational studies • motivation incentives • finance • investment

1. Introduction

Research anddevelopment (R&D) are important determinants of firm growth and performance (Porter 1985, Amit and Zott 2001, Stephan 2010, Teece 2010, Keupp et al. 2012). Innovation is also thought to be a fundamental driver of long-run economic growth (for instance, in the Schumpeterian growth model of Aghion and Howitt 1992). However, although R&D is important for the success of companies in many sectors, it is generally an expensive and complex undertaking. Deciding which elements of prior knowledge are important for current projects, what knowledge should be drawn from, and the particular form in which knowledge should be combined is often shrouded in uncertainty (Boudreau et al. 2016). Appropriate risk taking is important because projects with greater uncertainty have a lower probability of bearing fruit but may also generate more pathbreaking innovations if successful (Azoulay et al. 2011). In this paper, we study the effect of uncertainty on research funding decisions by asking how project risk affects project choice.

One ingredient to a successful R&D program is its ability to encourage appropriate risk taking—tolerating failure in pursuit of reward (March 1991, Manso 2011). This is consistent with recent empirical evidence on research grants (Azoulay et al. 2011) as well as for venture-backed funding of start-up firms (Tian and Wang 2014). Although the importance of appropriate risk taking may be widely recognized, it is often challenging in practice. For example, the decline in new drugs and breakthrough therapeutics—despite increased R&D spending—has been attributed in part to lack of risk taking by pharmaceutical and biotech companies (Munos and Chin 2011, Krieger et al. 2019). Similar concerns exist in private sector areas, including semiconductor manufacturing (Bloom et al. 2017), as well as in academic research. For example, Marks (2011, p. 2) writes that "everyone familiar with NIH operations knows that it is extremely difficult to obtain funding for groundbreaking, high-risk research." The amplification of those challenges in recent decades may, in part, reflect the greater need for risk taking in a crowded scientific arena in which the burden of knowledge required to reach the scientific frontier is ever expanding (Jones 2009, Chu and Evans 2021).

To inform our understanding of risk taking in R&D, we focus on individual decision makers who often serve as gatekeepers in selecting which ideas to invest in and commercialize. Such individuals include R&D managers, external review board members, and investor analysts. Prior work has examined how incentive structures affect risk taking. In this study, we highlight that even if incentive structures are aligned with risk taking, the way individuals respond to uncertainty may shape R&D investment decisions. We examine several potential barriers to risk taking that stem from individual decision makers, including cognitive limitations in processing variance, a desire for diversification, loss aversion, sensitivity to ambiguous payoffs, and personal risk preferences.

We do so using a discrete choice experiment (DCE) designed to highlight some aspects of the role of uncertainty in shaping the decision to fund R&D projects. Experimental participants were asked to rank a series of uncertain research projects. We instructed participants to assume the role of the director of the R&D group at a private company, and they were asked to choose their preferred research projects from a series of hypothetical proposals that had been judged and scored by an objective, third-party science advisory panel. The experimental design was a stylized version of ratings procedures that are used as inputs to allocate internal funding at firms, attract external investors, and award government research grants.[1]

Compensation was determined by a competitive "tournament" structure. Participants were compensated for the performance of the R&D projects that they chose to fund relative to the choices of their peers in the experiment. The highest-scoring participants received a substantially larger monetary reward than their peers. There was no penalty for low performance; the bottom 75% of scorers all received the same compensation. Because there were large rewards for high performance and no downside risk for poor performance, the incentive structure disproportionately rewarded participants for choosing higher-variance (i.e., riskier) projects. That is, projects with greater disagreement in ratings (i.e., some high ratings and some low ratings) had a higher chance of success than projects with the same average rating but greater agreement (e.g., project ratings of 5, 5, 3, 1, 1 versus 3, 3, 3, 3, 3, which has the same mean but lower variance).[2]

The experiment was conducted with 290 Master of Business Administration (MBA) and Master of Finance (MFin) students at a major research university in a program focused on the intersection of business and technology. Many of these students come from an R&D background and will go on to work at investment firms or serve as managers making R&D decisions at companies in the health and technology sectors.

The experiment took place in two phases. In the first phase, 150 research subjects were asked to evaluate projects under three distinct sets of choice scenarios. The second phase of the study was designed to explore some of the potential mechanisms driving our first-phase results. It included three experimental treatments, where 140 subjects were randomized into (1) a replication of the baseline experiment from phase 1, (2) a version of the experiment created to test the role of loss aversion, or (3) a version designed to test for the effects of ambiguity. After the choice experiments, we elicited participants' personal risk preference parameters (both phases) and loss aversion parameters (second phase only).

In phase 1 of the experiment, the first set of choice scenarios assessed whether the incentives to choose high-variance projects in fact led to such choices among participants. Each participant was presented with 10 scenarios, where they were asked to rank four potential projects based on their preferences for funding. For each project, the participant was shown the individual scores from the advisory panel members and the average of those scores.

We find that most participants acted in an excessively risk-averse manner when selecting projects. Because of the competitive incentives, when offered two otherwise identical options, choosing a higher-variance project first-order stochastically dominated choosing a lower-variance project. Despite this, participants were more likely than not to choose dominated projects. In other words, holding average score constant, participants were, on average, significantly *less* likely to choose a project as variance in ratings increased. Even in ideal cases where the participants were choosing between two projects that had identical mean scores, they chose the dominated project—the one with lower variance—three-quarters of the time. Because no risk aversion parameter can rationalize this behavior, we refer to the strong distaste for high-variance projects exhibited by the participants as variance aversion.

Why did the participants behave this way? What might a manager do to overcome or circumvent this behavior in employees overseeing R&D funding? Our subsequent choice scenarios, preference parameter elicitations, and the second phase of the experiment examined potential mechanisms for variance aversion and tested interventions to address it.

We first examined heterogeneity in behavior across elicited preference parameters and participant demographics. The analysis reveals strong correlation between variance aversion and multiple dimensions of heterogeneity. Participants who were more risk loving and had more R&D experience exhibited a greater taste for variance. We also find less robust evidence of greater taste for variance among participants who faced more competition for a high-reward payment (because of discreteness in the number of higher rewards issued) and participants in the MBA program (compared with the MFin program). In contrast, college coursework, elicited discount rates, and a measure that checked for understanding of the experimental tasks instructions were not correlated with variance preferences.

Second, we assessed whether cognitive limitations in effort or attention drive the response to uncertainty. To do so, the second set of choice scenarios tested a simple informational intervention in which we additionally showed participants the variance of project scores. The variance is straightforward to infer from the individual scores in the first set of choice scenarios, so the second set of choice scenarios measures the effect of lowering the cognitive calculation costs and increasing the salience of score variance. Showing the variance backfired, leading subjects to engage in even more variance-averse behavior than in the first set of choice scenarios.

Third, we examined whether variance aversion can be explained by a desire to diversify risk. The third set of choice scenarios asked participants to construct portfolios of research projects rather than pick single projects to assess whether allowing for diverse portfolios would encourage more risk taking. We find, consistent with the first two sets of choice scenarios, that individuals continued to make variance-averse choices. By randomizing the budget that was provided to the subjects for each portfolio problem, we are also able to assess the effect of budgetary pressure. We find that tighter budgets exacerbated the problem, leading to more variance aversion.

Fourth, we examined answers that participants gave during a debriefing that followed the choice scenarios in which they were given free space to tell us why they had made their choices. The majority of participants stated that they treated the choice as a simple mean-variance trade-off. These participants were, unsurprisingly, more likely to exhibit variance aversion in their project choices. A minority of participants were more sophisticated in their decision-making process and looked at individual project scores, explicitly discussed the idea that successful R&D requires one to embrace uncertainty, or gave other answers that suggested they were willing to be variance loving in this setting. Such answers were significantly positively correlated with a greater preference for high-variance projects, being risk loving in the risk preference elicitation, and prior work experience in the R&D sector.

Finally, in the second phase of the study, we investigated whether variance aversion is in part because of loss aversion (i.e., avoiding high-variance projects that could result in losses) or ambiguity aversion. To address the former, we replicated the first choice scenario of the baseline phase 1 experiment but removed all mentions of potential losses. Participant behavior was largely indistinguishable from the behavior in the baseline experiment. In addition, elicited loss aversion preferences were associated, if anything, with greater taste for variance, and considerable variance aversion remained after accounting for loss-averse preferences.

To examine whether participants' response to variance in ratings partially reflects a distaste for ambiguity—in particular, ambiguity in how ratings map onto expected financial returns—we tested a version of the baseline experiments in which the ratings of projects were explicitly denominated in financial terms. The average subject was slightly more averse to choosing high-variance projects than in the baseline experiment, indicating that ambiguity aversion does not explain our core findings.

Our results suggest that explicit risk-taking incentives might not be enough to encourage optimal R&D within a firm and that excessive risk aversion could lead to suboptimal R&D investment. To be more concrete, consider an example based on the empirical results that highlights the effect this behavior could have on breakthrough advances. In the experiments, subjects were shown hypothetical projects with ratings on a one to five scale. Consider two stylized examples of projects with identical average ratings but different variances. The first project is rated a four of five by all seven panelists on the advisory committee. The second project is more divisive: receiving three ratings of three, one rating of four, and three ratings of five. The first project has a variance of zero, whereas the second project has a variance of one. Based on the findings from our experiment, subjects would be six percentage points less likely to choose the second project, despite the fact that the first project has no chance of producing an outcome of the highest-possible quality and the second project has a 43% chance of doing so.

Our examination of participant characteristics points to potential solutions. Risk-loving participants performed better, on average, on the experimental tasks and chose projects more in line with optimal theory. Performance was hampered by treating the choices as "standard" portfolio optimization problems, an impulse that appears to have been tempered, in part, by training and work experience. These findings suggest that firms aiming to encourage more innovation may want to include the risk preferences of those workers in charge of research and development as a factor in their hiring and promotion decisions as well as emphasize the convex nature of returns to R&D as part of their training.

The experiment provides evidence that incentives alone might not be enough to induce appropriate risk taking by individuals. The internal validity of this result and its generalizability to real-world R&D both come with important caveats. First, the use of project score rankings in the first phase of the study prioritizes external validity by capturing one important aspect of the way R&D decisions are made in practice. The scores could, however, introduce ambiguity in how participants interpreted the decision. The objective lottery treatment in phase 2 of the experiment was conducted to address this potential ambiguity. The participants faced only upside risk, in contrast to real-world settings where losses are also potentially important. The use of alternative loss priming and elicitation of loss-averse preferences across the phase 1 and 2 experiments reveals little effect of potential losses in this setting, but the larger stakes involved in real-world R&D settings could lead to different conclusions. Finally, the results show that participants who have substantial experience with real-world R&D do take on more appropriate risk. Thus, the issues we highlight, although of potential concern as the need for risk taking in R&D expands and the existing workforce turns over, might be handled well by current R&D-intensive firms.

2. Literature

Theoretical models of optimal R&D argue that both firms and public funders should invest in high-variance research projects. An important early contribution to this literature, Dasgupta and Maskin (1987) argue that from the perspectives of the individual scientist, competitive firms, and society as a whole, the spoils from R&D are skewed toward novel, high-quality discoveries. Given the disproportionate benefits from producing the highest-quality discoveries, investing in riskier R&D projects is optimal from both social and private perspectives.[3]

More recently, in a theoretical setting similar to our experiment, Tishler (2008) shows that competition among firms or research groups should lead them to adopt high-variance R&D portfolios. Given two projects with the same expected discovery quality, a firm should choose the higher-variance project to capture convex returns. The incentives in our experiment are meant to replicate the competitive compensation scheme modeled by Tishler (2008) and observed in real-world R&D. Participants were paid substantially more if their research projects and portfolios performed well relative to the other participants.

Despite models showing that optimal R&D entails investment in high-variance research, many observers have documented low rates of risk taking by agencies that disburse research funds (Azoulay et al. 2011, Marks 2011) and firms that conduct R&D (Munos and Chin 2011). These papers leave open the question of how the preferences of individual decision makers help drive suboptimal risk taking in R&D, even when explicit incentives for innovation are in place.

A separate strand of research highlights the potential link between individual preferences and innovation. Prior work demonstrates that scientists have important nonpecuniary motivations (Merton 1973, Dasgupta and David 1994). In particular, scientists are willing to accept a lower salary to work in organizations that allow them to pursue independent research (Stern 2004), and such preferences are positively correlated with innovative performance, as measured by patent applications (Sauermann and Cohen 2010). Related work across a range of industries finds that willingness to take risks is positively correlated with assessments of innovative creativity in the workplace at both the organizational (Amabile et al. 2017) and individual levels (Madjar et al. 2011). Bringing these strands together, recent evidence suggests that less risk-averse individuals generate more novel inventions by pursuing riskier innovation strategies (Graff Zivin and Lyons 2020).

Finally, related to our suggestion that firms may want to take into account the risk preferences of their R&D managers, prior work explores the relationship between preferences and selection into innovative sectors. A large literature examines the relationship between risk preferences and both selection and performance of entrepreneurs with mixed findings (Astebro et al. 2014 provide a review). Goel and Thankor (2008) show theoretically that firms might value overconfident chief executive officers (CEOs) if that overconfidence helps counteract risk aversion. Overconfident CEOs are also more likely to invest in risky projects, leading to higher innovation if the firm is in an innovative sector (Hirshleifer et al. 2012). Related work by Kagan et al. (2020) makes a similar point about equity contracts in entrepreneurial teams. Traditionally, researchers have argued that contract structure matters for team performance, but Kagan et al. (2020) show that individual preferences determine which types of contracts are taken up by workers. This selection confounds estimates of the effects of contract type on firm performance and means that individuals in charge of hiring should pay close attention to the preferences of potential employees.

3. Experimental Design

3.1. Experimental Setup

The experiments were implemented among master's degree students enrolled in a program focused on the intersection of business and technology. The typical student has three to four years of work experience with a background either in research-intensive firms in science and technology sectors or in finance, banking, and economics. All have formal academic training in assessing risky trade-offs and portfolio analysis. Many of the graduates will work for investment firms or will assume management positions within research divisions of corporations across a wide spectrum of science and technology spaces. Thus, studying the decisions of this group is particularly germane for our understanding of R&D investment choices within the private sector. Summary statistics for the study participants are discussed in Section 3.1.4.

Participants were asked to assume the role of the head of a research division at an organization considering whether to fund project proposals based on ratings from a third-party scientific advisory panel (see Online Appendix A for the instructions). They were then tasked with ranking research projects in a series of choice scenarios. Participants were allowed to take as much time as they wanted to complete the experiment. Empirically, the average participant spent 58 minutes on the experiment.

The experiment was conducted in two phases. Participants in the baseline phase 1 experiment were asked to rank research projects in three sets of choice scenarios. In the second phase, participants were randomized into one of three distinct experimental treatments to help elucidate the mechanisms driving our baseline results. The two phases of the experiment were distinct. Participants were not explicitly randomized across the first and second phases, but they were drawn from a demographically similar subject pool.

3.1.1. Baseline Phase 1 Experiment. In the baseline experiment, each participant took part in three sets of choice scenarios. In the first set, they were presented with a list of four research projects rated by seven reviewers (on a scale of one to five) along with the average reviewer score for each of the projects.[4] The subjects ranked projects based on the likelihood that they would fund them. The ranking was carried out by first choosing the most and least preferred project and then by ranking the remaining two projects. This process was repeated for 10 different groups of research projects, with each group characterized by different reviewer score profiles.

In the second set of choice scenarios, the same procedure was repeated for 10 more groups of projects, but the subjects were also shown the variance of reviewer scores. Because participants could calculate the variance themselves based on the individual ratings, the second set of choice scenarios did not provide more information than the first one. It was designed to address concerns about cognitive calculation costs, computation errors, or misunderstandings, but it also made that feature more prominent. An example of the initial project choice screen is shown in the online appendix.

The third set of choice scenarios presented each subject with eight portfolio choices. For each portfolio choice, subjects were presented with 10 different projects rated by seven reviewers. As in the second set of scenarios, each project was rated by seven reviewers, and participants saw the individual ratings as well as each project's average rating and variance of ratings. In addition, each project was assigned a cost of either \$1, \$4, \$7, or \$10 million. Subjects were provided a randomized budget that they could use to fund the projects in the portfolio. One of eight possible budgets (\$12, \$13, \$14, \$15, \$16, \$17, \$18, or \$19 million) was chosen without replacement for each portfolio choice, so each subject saw the full set of possible budgets. Participants could select and deselect projects from their portfolio. We displayed the remaining funds in their budget for their chosen portfolio until they finalized their choices. An example portfolio choice question is shown in the online appendix.

At the end of the experiment—after participants made their decisions but before learning of their performance—subjects completed a debriefing about why they had made their decisions as well as a survey that included questions about demographics and their risk preferences. We utilized a multiple price list to elicit risk preferences, a standard technique in the experimental economics literature (Charness et al. 2013). Subjects were provided with a list comparing a guaranteed payment with gambles with progressively lower variance and expected values. The subjects were then asked to make hypothetical choices between the gambles and the guaranteed payment. Based on their choices, we classified participants as risk averse, risk neutral, or risk loving, and we calculated each subject's coefficient of relative risk aversion (details of this calculation can be found in Online Appendix A, Section A.5).

3.1.2. Phase 2 Experiment. In order to tease out the mechanisms underlying the results from our baseline experiment, we conducted a second round of experiments with a new cohort of students drawn from the same academic program. Because of the coronavirus disease 2019 pandemic, these experiments were administered to students taking their classes remotely. In this second phase, participants were randomized into one of three distinct experimental arms: (1) one that replicated the first and second sets of choice scenarios in the baseline experiment, (2) an identical experiment that removed all loss-framing language, and (3) an identical experiment that replaced reviewer rating scores with an objective payoff matrix. In addition, we also gathered information on loss aversion preferences from the subjects using the elicitation from Imas et al. (2017), based on the design of Abdellaoui et al. (2008). Details on the calculation of the loss aversion parameter can be found in Online Appendix A, Section A.6.

3.1.2.1. Arm 1: Replication Experiment. The replication experiment was designed to create a bridge between the two experimental phases by allowing for a direct comparison between the behavior of subjects in each. The randomization across experiments within the second phase also ensures that comparisons across these follow-up experiments can be interpreted in a causal framework.

3.1.2.2. Arm 2: Experiment Without Loss Framing. Participants randomized into the experiment without loss framing were presented with the same choice scenarios as in the replication experiment (choice scenario sets 1 and 2). The only difference was in the instructions that introduced the experiment and choice scenarios. The no loss framing experiment removed all language that stated or implied that losses were possible when investing in R&D projects. The complete instructions are available in Online Appendix A, Section A.2. In all of the experiments, subjects could only gain money, so the change in the instructions only affected the framing of the experiment rather than the true underlying incentives.

3.1.2.3. Arm 3: Experiment with Objective Returns. Participants randomized into the objective returns version of the experiment also engaged in the same choice scenarios as in the replication experiment but with a different framing. In this case, the possible value of a project was presented not in terms of scores generated by an outside scientific advisory panel but as objective financial returns. The returns were consistent with the no loss experiment in that all of the projects were shown to have strictly positive net returns. The returns were displayed in units of millions of dollars. The values for the returns had the same distribution as the advisory panel ratings from the other three experiments. The instructions are shown in Online Appendix A, Section A.3, and example choice scenarios can be seen in Online Appendix A, Figure S6.

3.1.3. Incentives. By design, participants in all experimental treatments and phases were incentivized to choose riskier (i.e., higher-variance) projects. At the beginning of the experiment, subjects were told that they would receive a score based on the projects and portfolios that they chose. The realized value for each project was generated by an independent draw from a normal distribution with mean and variance of the reviewer scores. To maintain incentive compatibility throughout the ranking, final scores were affected by the full ranking of all project choices that the subject made. For each project choice question in the first and second experiments, the final score for each individual project was equal to the full realized value for the first-choice project, 0.75 times the value drawn for the second-choice project, 0.5 times the value drawn for the third-choice project, and 0.25 times the value drawn for the fourth-choice project. The value of the portfolio questions was similarly drawn from a normal distribution with mean equal to the sum of each individual project's mean weighted by cost and with variance equal to the sum of each project's variance weighted by cost. The project and portfolio scores (where applicable) were summed to create the total score for the participant.

We then publicly awarded prizes to the top performers in each session; the top 10%–25% of scores received $25, and the top 10% of subjects received $100. All subjects received a $15 participation fee. Because we offered large rewards for performance in the right tail of the distribution and offered no additional rewards for performance in the bottom three-fourths of the distribution, there was a large potential upside and no downside risk from choosing higher-variance projects. Thus, subjects had a strong incentive to choose higher-variance projects to maximize their probability of winning the largest prizes. For two projects with the same average rating, choosing the higher-variance project first-order stochastically dominated choosing a lower-variance project, meaning that all subjects, regardless of risk preferences, should have chosen higher-variance projects on the margin. We assessed participant understanding of the incentives through actions taken during the portfolio choice section of the experiment and by debriefing the participants after they had completed all project-ranking scenarios. We report results for these two assessments.

3.1.4. Recruitment, Sample Size, and Sample Summary Statistics. All baseline experimental sessions were implemented during regularly scheduled class sessions of the MBA and MFin programs. Participants in the other three experiments were also recruited through their classes. Each professor chose whether to field the experiment during class time or outside of class. The randomization was stratified by class for all experiments.

All students in the class were eligible to take part, and participation was voluntary. After obtaining informed consent from all participants, they completed the experiment on their own computers. On average, the experiment took subjects about one hour to complete. The baseline experimental sessions were conducted in person, and subjects were paid at the end of the session. The three other experiments were conducted remotely, and subjects were paid after all participants in their class had completed the experiment—at most one week after the initial distribution of the experiment.

For the phase 1 experiment, a total of 196 students were recruited in six experimental sessions. One subject started the experiment but had to leave before completing it, and four subjects failed to provide us with answers sufficient to calculate risk preferences. They were excluded from the analysis. In the first session, the order of projects was not randomized because of a coding error, so we exclude the 36 students from that session in the baseline analysis (results with all participants are shown in Online Appendix B, Table B4). Five additional participants exhibited multiple switching on the risk preference elicitation and were also excluded from the base sample. All results are robust to the inclusion or exclusion of these participants. The final sample, therefore, contains 150 subjects.

Each subject faced 10 choice scenarios in choice scenario sets 1 and 2. Each scenario involved choosing between nine potential options, yielding 13,500 total observations for each set. The options in choice scenario set 3 varied by budget, which was randomized across subject. The average subject had 1,399 options, leading to a total sample size for set 3 of 219,310 observations.[5] The standard errors for all analyses are clustered at the subject level to account for correlation within subject across choice scenario.

For the three additional experimental arms, we recruited 140 subjects across four sessions. We randomized within session at the individual level; 46 subjects completed experimental arm 1 (phase 1 replication), 47 subjects completed arm 2 (no loss framing), and 47 subjects completed arm 3 (objective payoffs). Each of the additional experiments involved two sets of 10 choice scenarios. As in the first and second sets of choice scenarios from the phase 1 experiment, each scenario involved ranking four potential projects.

Summary statistics for all study participants in the main estimation sample are shown in Tables 1 and 2. The statistics show that the typical participant has multiple years of work experience, and across the two phases of the experiment, 41% of the participants reported R&D sector work experience. The average participant was risk averse, and in phase 2, the typical subject was also loss averse according to the preference elicitation. The summary statistics provide an initial indication that there are no gross imbalances across observables, which is further validated by formal tests of balance across the treatment arms in phase 2, as shown in Online Appendix B, Table B3.

Table 1. Summary Statistics: Phase 1 Experiment

Variable	Mean	Standard deviation
Age	26.68	5.29
Years of work experience	2.62	4.35
Has worked in R&D	0.30	0.46
Coefficient of relative risk aversion	1.26	0.27
Discount rate	0.25	0.21
Math classes	4.64	1.46
Decision science classes	4.03	1.56
Observations	150	

Notes. The table shows summary statistics for the participants in the experiments. It shows statistics for the 150 participants in the phase 1 experiment.

3.2. Design of the Discrete Choice Experiments

The design for the choice scenarios presented to the subjects builds on models of random utility theory to estimate discrete choice models using decisions from discrete choice experiments.[6] These designs allow discrete choice models to be applied to situations where individuals are making choices that are not currently observed in real markets. We followed this tradition by developing experiments to simulate hypothetical but potentially real proposals and projects and asking individuals to evaluate them and make choices. The design allows us to estimate statistical models using the experimental choices as data to approximate the individuals' choice processes.

DCEs are based on traditional experimental design concepts for fractional factorial designs widely used in applied statistical work.[7] To construct the choice sets in our experiment, we first enumerated all possible combinations of seven hypothetical raters using a five-category rating scale. We then calculated the mean and variance of each combination and sorted them from highest to lowest, and we identified 16 orthogonal combinations of means and associated variances. Using these combinations, we constructed the choice sets for the 20 individual project-ranking questions and then constructed the choice sets for the eight portfolio questions.

To construct the choice sets for the project ranking task, we used a balanced incomplete block design (BIBD)—see Louviere et al. (2015)—to create 20 sets of four project proposals. Each proposal was described by seven ratings. The mean and the variance of these ratings were the two primary attributes associated with each proposal. The 20 sets of projects were divided into two groups of 10 to create sets of choice scenarios.

To ensure that the models we estimated were not saturated and to enhance the degrees of freedom, we made two versions of the DCE by randomly rearranging the original DCE attributes (mean and variance) and again making 20 sets of four proposals using the same BIBD. Again, these 20 sets of proposals were divided into two subsets of 10.

Table 2. Summary Statistics: Phase 2 Experiment

	Arm 1		Arm 2		Arm 3	
Variable	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Age	31.41	6.09	31.70	5.44	32.04	7.47
Years of work experience	7.34	6.49	6.41	5.76	7.95	6.96
Has worked in R&D	0.54	0.50	0.53	0.50	0.53	0.50
Coefficient of relative risk aversion	1.26	0.27	1.14	0.29	1.15	0.33
Discount rate	0.29	0.27	0.27	0.27	0.24	0.21
Loss averse	0.67	0.47	0.77	0.43	0.74	0.44
Math classes	4.15	1.56	4.11	1.75	3.79	1.74
Decision science classes	2.74	1.73	2.77	1.81	2.81	1.50
Observations (total = 140)	46		47		47	

Notes. The table shows summary statistics for the participants in the experiments. It shows statistics for the 140 participants in the phase 2 experiment.

We then randomly blocked each of the two versions of the DCE—Version I and Version II—and the two subsets within version—Subset A and Subset B. This produced four treatment groups: Version I.AB, Version I.BA, Version II.AB, and Version II.BA, where the first letter refers to the subset used in choice set 1 and the second letter refers to the subset used in choice set 2. By showing identical choice scenarios to different participants in the same treatment group, we can identify the effect of changes in attributes (score mean and variance) conditional on choice scenario and participant fixed or random effects. Balance across the experiment versions is shown in Online Appendix B, Table B1.

To construct the choice sets for the portfolio selection task, we used the complement of the BIBD used to construct the choice sets for the project-ranking task (the complement contains all combinations not included in the first BIBD). Costs were also added as an additional attribute for the proposals, with costs randomly assigned following the same procedure for mean and variance used in the project selection tasks. Costs were blocked so that subjects would routinely face choices between two projects with identical expected value (same cost and same mean) but different variance. We exploit this feature to study risk-taking behavior as a function of portfolio budget in the results section. We arrayed the 16 combinations into 16 sets of 10 proposals. We then created four blocks of eight choice sets using the method discussed to make two versions of the DCE and two subsets within each DCE. We randomly assigned each block of eight portfolio selection questions—Block 1, Block 2, Block 3, Block 4—to one of the four experimental versions discussed (i.e., Version I.AB.1, Version I.BA.2, Version II.AB.3, Version II.BA.4).

In the phase 1 experiment, we randomized individuals to one of the four versions, stratified on session. In the phase 2 experiment, within each of the three experimental arms, participants were randomly assigned to one of the four versions, stratified on session. The order in which projects were presented within each version was also randomized across sessions. The experimental instrument was programed and delivered using the Sawtooth Software platform.

4. Empirical Specification

We estimate the relationship between project attributes and subject choice using a generalized multinomial logit (G-MNL) model. The estimating equation models the probability that subject i chose alternative j in choice scenario t as

$$Pr(choice_{it} = j \mid \boldsymbol{\beta}_i) = \frac{exp\left(\boldsymbol{\beta}_i' \boldsymbol{x}_{itj}\right)}{\sum_{k=1}^{J} exp\left(\boldsymbol{\beta}_i' \boldsymbol{x}_{itk}\right)}, \tag{1}$$

where $\boldsymbol{x}_{itj}$ is a vector of attributes (mean and variance of the projects in the baseline models and interactions with subject demographics in the models exploring heterogeneity) and $\boldsymbol{\beta}_i$ is the vector of individual-specific coefficients on the vector of attributes. These coefficients can be interpreted as utility weights placed on the attributes by each individual and are defined by

$$\boldsymbol{\beta}_i = \sigma_i \boldsymbol{\beta} + \boldsymbol{\eta}_i. \tag{2}$$

The coefficients in Equation (2) are a vector $\boldsymbol{\beta}$ that is constant across individuals and measures the average utility weights across the sample for the different variables in $\boldsymbol{x}$; a single parameter for the scale of the individual-level idiosyncratic error σ_i, which captures overall scaling of an individual's tastes; and a random vector $\boldsymbol{\eta}_i$ distributed multivariate normal with mean zero and variance-covariance matrix $\boldsymbol{\Sigma}$, which captures taste heterogeneity. We follow Fiebig et al. (2010) and assume that σ_i is distributed log normal with mean $\overline{\sigma} + \boldsymbol{\theta}' \boldsymbol{z}_i$ and standard deviation τ. The parameter $\overline{\sigma}$ is a normalizing constant, and $\boldsymbol{z}_i$ is a vector of subject characteristics that explain differences in σ_i across individuals. In our application, we focus on project and portfolio attributes and limit our attention to subject indicators in z_i.[8]

The workhorse model in applied microeconomic studies of discrete choice is the multinomial (conditional) logit model. We prefer estimates based on the more flexible G-MNL model because of the strong restrictions imposed by the standard conditional logit. The restrictions effectively rule out many kinds of heterogeneity that are of potential interest when studying behavior and that can lead to confounding. The random coefficient (mixed) logit model relaxes assumptions about preference heterogeneity but restricts all agents to have their error component drawn from the same distribution, such that differences in these "scale" parameters could easily be misidentified as differences in preference parameters. This occurs because preference and scale parameters are not separately identified in choice models and behavior is governed by their ratio, so that an upward shift in the scale parameter must shift the magnitude of the vector of preference parameters upward to maintain the same ratio. The G-MNL model nests both the simple mixed logit model that allows heterogeneity in preference parameters and models that allow for scale heterogeneity by allowing both to vary in a reasonably flexible but statistically identified way.

For ease of interpretation, however, we also present corresponding conditional logit models for the main results in the paper. The analysis of the third set of choice scenarios (portfolio choices) is also carried out using standard conditional logit and fixed effects linear regression specifications. We estimate these specifications because we are interested in the effect of budget constraints on choice, and budget was randomly varied within subject, across choice scenario. As such, we rely on between-subject comparisons that preclude the use of individual and choice scenario-specific heterogeneity parameters. We verify that the budget randomization was balanced on observable characteristics in Online Appendix B, Table B2.

5. Results

5.1. Initial Evidence for Variance Aversion

Our primary question of interest is whether subjects responded to the incentives we gave them by choosing higher-variance projects when faced with a choice between two otherwise similar research proposals. We formally test this by estimating statistical models that control for the average score, allowing us to isolate the effect of variance on the likelihood that a subject would choose a given project. As discussed, the repeated within-subject sampling of the experimental design allows us to estimate G-MNL models that further account for latent subject-specific heterogeneity while relaxing strong assumptions that underly the estimation of conditional logit models.

Table 3 shows results from the first set of choice scenarios in the phase 1 experiment. In all columns, the dependent variable is an indicator equal to one if the subject chose the project.[9] The explanatory variables are the project mean and variance, and they are standardized to have an average value of zero and standard deviation of one.

The coefficients in the top portion of the table (labeled "Average Utility Weight") are the estimates of the utility weight that subjects placed on average project score and score variance (the $\boldsymbol{\beta}$ terms in Equation (2)). The second section of the table (labeled "Utility Weight Heterogeneity") reports estimates of the heterogeneity in preference (the σ_i terms in Equation (2)). The third section of the table reports the estimate of the standard deviation of individual-level scale heterogeneity, which we estimate to be small in this case.

The results show that, on average, participants had strong preference for projects with higher average scores and lower score variance. This behavior is at odds with the incentives the participants faced and provides our first evidence of variance aversion. The result holds in the G-MNL model, in a traditional conditional logit model (column (2)), and across

Table 3. Project Choice as a Function of Mean and Variance

Dependent variable	(1) Phase 1, choice scenario, Set 1, project choice	(2) Phase 1, choice scenario, Set 1, project choice
Average Utility Weight		
Average Project Score	5.09***	2.48***
	(0.51)	(0.18)
Project Score Variance	−0.63***	−0.45***
	(0.075)	(0.051)
Utility Weight Heterogeneity		
Average Project Score	2.35***	
	(0.30)	
Project Score Variance	0.75***	
	(0.076)	
τ	0.029	
	(0.090)	
Model	G-MNL	C-Logit
Observations	13,500	13,500
Subjects	150	150

Notes. The table shows results estimated using choice scenario-level data from choice scenario set 1 in the phase 1 experiment. Column (1) is estimated using a G-MNL model (Equation (1)). Column (2) is estimated using a conditional logit model. The outcome variable is an indicator for whether the project was chosen. "*Average project score*" is the average of the five scores for the project. "*Project score variance*" is the variance of the scores. Both explanatory variables are standardized. All models contain subject and choice scenario random effects in addition to the variables shown in the table. Standard errors, clustered at the subject level, are in parentheses.

$^{***}p < 0.01$.

each choice occasion when analyzed with a rank-based multinomial logit model (Online Appendix B, Table B6).

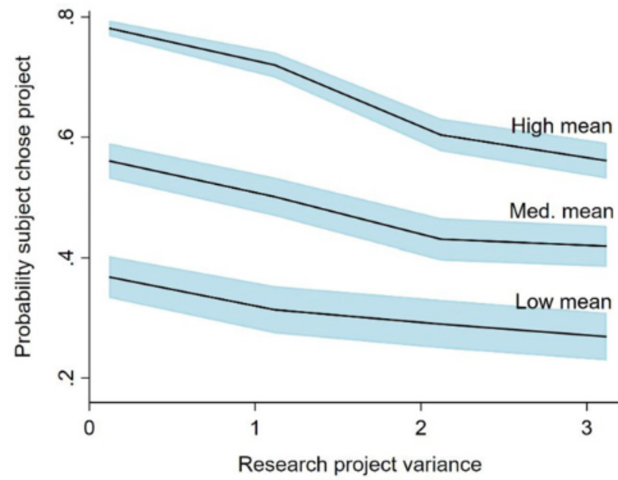
The G-MNL estimates suggest that the simpler conditional logit model, however, is inconsistent. The significant utility weight heterogeneity parameters indicate that there are nonnegligible preference scale differences across participants. The conditional logit estimates will be inconsistent when the independent and identically distributed (i.i.d.) specification of its error component does not hold. One way to compare across the models is to consider the ratio of average utility weights for the two project attributes. Dividing the two parameters from the first panel of column (1) yields a ratio of −8.08, whereas in column (2), the ratio is −5.51. Thus, the estimated preference for higher average project score relative to project variance is stronger in the G-MNL model. The conditional logit model estimates are lower and are confounded by scale heterogeneity.

The full distributions of estimated preference parameters from the G-MNL model are shown in Online Appendix C, Figure C1. The results demonstrate that 1/6th of the sample exhibited variance-loving (rather than variance-averse) preferences. It is this heterogeneity—the fact that the average subject, despite the incentives, exhibited variance aversion, whereas some subjects behaved in accordance with the incentives to seek variance—that we look to explain in the next section.

Figure 1 summarizes the G-MNL results from Table 3 for a range of different project attributes. The figure shows the estimated probability that the average subject would choose a project with a high, medium, or low average score[10] and score variances that span the set shown to study participants. The figure again shows that subjects strongly preferred to choose projects with higher mean scores and lower score variance. The figure also reveals that the effect of variance was stronger for projects with a higher mean. For a project with a high mean score, an increase in variance by one reduced the probability that a typical subject chose that project by 11.9 percentage points (95% confidence interval of 9.1 to 14.7). In contrast, for a low-mean score project, a one-unit increase in variance decreased the probability of selection by only 2.2 percentage points (95% confidence interval of −1.0 to 5.3).

In other words, participants were particularly variance averse when choosing between high mean projects—so much so that they were willing to frequently forgo selection of projects with the very highest average scores if it meant reducing their exposure to variance. When assessing projects with middling average scores, participants were less reluctant to choose high-variance projects, potentially because neither project

Figure 1. (Color online) Likelihood of Choosing a Project with a Given Mean and Variance



Notes. The figure shows the average probability of a subject choosing a project with a given mean and variance. Each solid line shows how the choice probability changes as variance increases for three different mean project scores: a high mean score of 4, a medium mean score of 3.5, and a low mean score of 3. The lighter shaded areas are 95% confidence intervals. The estimates are based on a G-MNL model fit to data from the first set of choice scenarios from the phase 1 experiment. The underlying G-MNL model results are presented in Table 3.

looked particularly attractive. We report results from a debriefing with participants that shed further light on this behavior in the next section. Irrespective of the underlying drivers of this behavior, it has potentially important implications for research project selection if otherwise stellar research projects are being rejected because of disagreement in reviewer scores, whereas more mediocre, if less controversial, projects are being favored instead.

As described in Section 3.1, we replicated the phase 1 experiment with a new set of subjects. The participants were drawn from a similar population as those who completed the baseline experiment. We find that the participants in the replication experiment behaved similarly to those in the baseline; both groups preferred projects with higher mean scores and lower score variance, on average. There is no significant difference in the variance preferences across the two groups, as shown in Online Appendix B, Table B5.

5.2. Mechanisms Underlying Project Choice and Variance Aversion

We now investigate the factors that might explain why subjects preferred projects with lower variance. We first assess whether a variety of subject characteristics, demographics, and other covariates do or do not correlate with variance aversion in the first set of choice scenarios from the phase 1 experiment. We then present results from four additional sets of choice scenarios and experimental interventions to determine whether differences in the experiment itself could lead participants to choose higher-variance projects.

5.2.1. Heterogeneity in Variance Aversion. The results from the G-MNL estimates in Table 3 showed that there was substantial subject-specific heterogeneity in variance preferences. In Table 4 and Figure 2, we examine correlation between variance aversion and observable, subject-level covariates. Table 4 presents G-MNL estimates that include interactions between project attributes and eight dimensions of heterogeneity. Figure 2 shows heterogeneity in the effect of a one-unit increase in project score variance on the probability that a project gets chosen. The measures are all based on observational data rather than explicit randomization, so we emphasize that the results are suggestive and correlational.

The figure shows a wide range in the correlation between observables and variance preferences. On one side, there is essentially no correlation between variance preferences and the number of college math courses taken, the number of decision science courses taken, and the elicited discount rate.

Some of the measures help us assess whether simple misunderstanding is driving the results. We cannot definitively rule out that study participants misunderstood the task. We do observe, however, that behavior is consistent between participants who exhibited some clear indicators of misunderstanding and those that did not. The fourth measure, titled budget spending, is an indicator equal to one if the participant left any of their allotted budget unspent during the third set of choice scenarios. The incentives were to spend all of the budget, so this indicator is a measure of whether the participants understood the instructions and incentives they faced. Although many participants left some budget unspent, this behavior also does not strongly correlate with variance aversion.

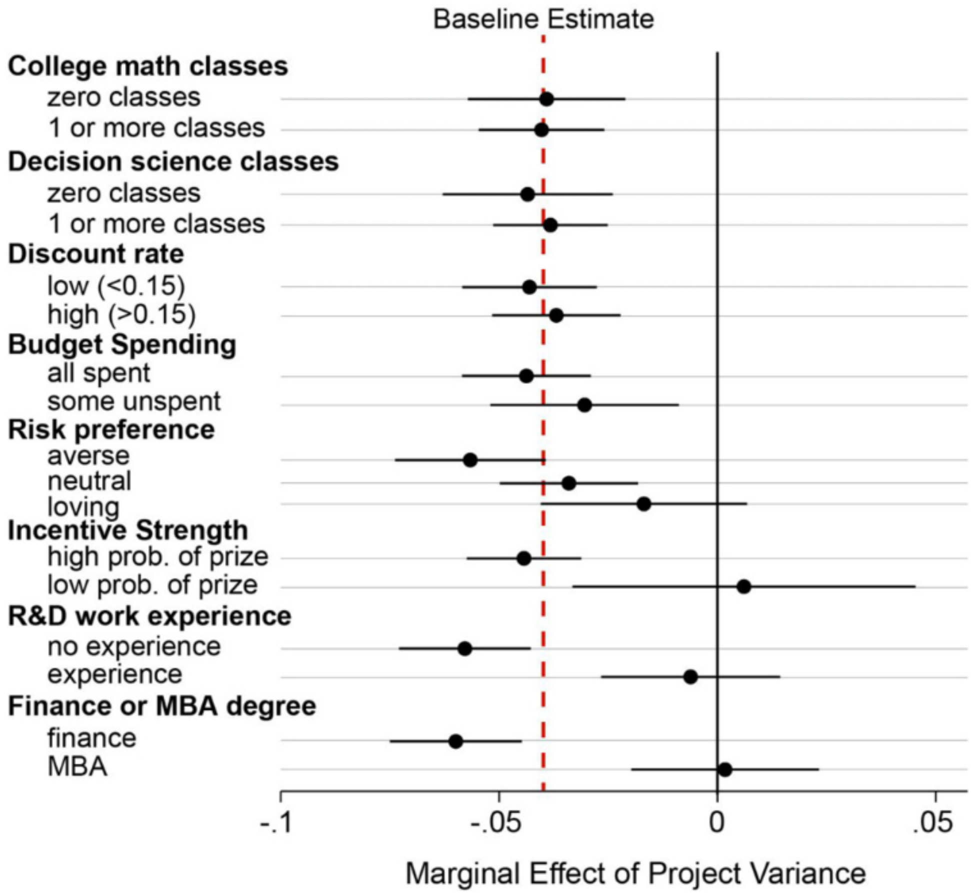
Table 4. Project Choice Heterogeneity, Choice Scenario Set 1, Phase 1

Dependent variable Dimension of heterogeneity	(1) Project choice Coefficient RRA	(2) Project choice Budget unspent	(3) Project choice R&D work experience	(4) Project choice Discount rate	(5) Project choice Math classes	(6) Project choice Decision science	(7) Project choice Finance or MBA	(8) Project choice Prize probability
Average Utility Weight								
Average Score	5.46***	5.73***	5.10***	4.90***	6.26***	4.71***	4.55***	3.73***
	(1.11)	(0.63)	(0.41)	(0.93)	(1.35)	(0.43)	(0.47)	(0.50)
Score Variance	0.80**	−0.52***	−0.74***	−0.74***	−0.46**	−0.47***	−0.23***	−0.25*
	(0.37)	(0.11)	(0.076)	(0.13)	(0.19)	(0.14)	(0.083)	(0.14)
Average × Het.	−0.24	−0.051	0.33	0.49	−1.71	1.00**	1.11***	2.04***
	(0.59)	(0.69)	(1.07)	(4.68)	(1.44)	(0.47)	(0.42)	(0.28)
Variance × Het.	−1.11***	−0.13	0.38	0.49**	−0.21	−0.29	−0.67***	−0.50***
	(0.30)	(0.15)	(0.32)	(0.23)	(0.23)	(0.18)	(0.13)	(0.18)
Utility Weight Heterogeneity								
Average Score	0.57***	1.44***	2.38***	1.79***	0.018	2.31***	2.24***	1.92***
	(0.11)	(0.23)	(0.23)	(0.22)	(0.15)	(0.24)	(0.25)	(0.24)
Score Variance	0.72***	0.79***	0.64***	0.74***	0.68***	0.69***	0.58***	0.56***
	(0.068)	(0.076)	(0.078)	(0.068)	(0.085)	(0.085)	(0.084)	(0.11)
Average × Het.	2.16***	1.37**	1.91***	4.99	1.69***	0.99***	1.42***	2.02***
	(0.23)	(0.56)	(0.27)	(4.05)	(0.28)	(0.13)	(0.28)	(0.31)
Variance × Het.	0.014	0.16	0.70***	0.23	0.44***	0.44**	0.56***	0.57***
	(0.049)	(0.11)	(0.27)	(0.38)	(0.095)	(0.20)	(0.12)	(0.13)
τ	0.37***	0.58***	0.075	0.024	0.39**	0.14***	0.11***	0.058*
	(0.058)	(0.080)	(0.11)	(0.098)	(0.16)	(0.015)	(0.026)	(0.031)
Subjects	150	150	150	150	150	150	150	150
Observations	13,500	13,500	13,500	13,500	13,500	13,500	13,500	13,500

Notes. The table shows results from estimating Equation (1) using choice scenario data from choice scenario set 1 in the phase 1 experiment. The outcome variable is an indicator for whether the project was chosen. "*Average score*" is the average of the five scores for the project. "*Score variance*" is the variance of the scores. Both explanatory variables are standardized. Each column also shows the effect of the interaction between those explanatory variables and a dimension of heterogeneity. The dimension of heterogeneity is given at the top of the column. Coefficient RRA in Column 1 refers to the coefficient of relative risk aversion. All models contain subject and choice scenario random effects in addition to the variables shown in the table. Standard errors, clustered at the subject level, are in parentheses.

*$p < 0.10$; **$p < 0.05$; ***$p < 0.01$.

Figure 2. (Color online) Heterogeneity in Marginal Effect of Variance on Project Choice



Notes. The figure shows marginal effects of variance on the probability of choosing a project in the first set of choice scenarios from the phase 1 experiment broken down by subject demographics and characteristics. The points are based on coefficient estimates from the eight G-MNL regressions reported in Table 4. The black solid horizontal lines show the 95% confidence intervals based on standard errors clustered at the subject level. The vertical dashed line shows the marginal effect based on the G-MNL estimates in Table 3. For continuous heterogeneity measures (coursework, discount rate, and incentive strength), the measure is discretized by dividing into above and below mean values.

In contrast, the next four measures do correlate with variance preferences. Participants who were risk averse were substantially and significantly more variance averse than participants who were risk loving. A risk-averse subject responded to a one-unit increase in project score variance by reducing the probability of choosing that project by six percentage points. A risk-loving participant, on the other hand, reduced the probability by only two percentage points.[11]

Participants with prior work experience in the R&D sector (30% of this sample) were also substantially less variance averse. The difference in marginal effects of variance between participants with R&D experience versus those without was 4.8 percentage points (95% confidence interval of 2.3 to 7.3). This result also holds when simultaneously estimating the effect of all of these dimensions of heterogeneity (see Online Appendix C, Figure C2), suggesting that it is not simply selection into prior work experience because of coursework, risk or time preferences, or other factors.

Similarly, participants who were pursuing an MBA were, on average, variance loving, whereas those pursing an MFin were substantially more variance averse, potentially because the MFin students were more likely to treat the choice scenarios as typical financial portfolio problems. The difference in marginal effects of variance between the two groups was 6.3 percentage points (95% confidence interval of 3.7 to 8.9).[12]

The strength of incentives faced by participants also affected preferences. Based on the number of other participants in a session, the probability that a subject would receive an extra $25, for example, varied between 20% and 25%. For a session with 36 people, for instance, exactly 9 people would receive a $25 or $100 award. In a session with 35 people, however, only 8 (or 22.9%) would receive a larger reward. Participants in the more competitive (lower-probability) sessions were substantially less variance averse.[13] This result, however, is not robust to including all dimensions of heterogeneity simultaneously, again suggesting that the incentive structure faced by the participants played a weaker role in determining behavior.

Online Appendix B, Table B9 and Online Appendix C, Figure C2 show the correlation of these measures when included simultaneously in the same regression. As discussed, elicited risk aversion and prior R&D experience continue to be two of the strongest predictors of variance aversion, suggesting that these two measures (or their correlates) independently predict the preferences of participants. The discount rate also becomes a strong predictor, with higher discount rates being associated with lower variance aversion. Online Appendix B, Table B9 also shows that these dimensions of observable heterogeneity explain a substantial fraction (74%) of the subject-level heterogeneity estimated in the baseline G-MNL regression reported in Table 3.

5.2.2. Assessing Cognitive Limitations in Effort or Attention. Second, we assess whether cognitive limitations in effort or attention help explain the response to uncertainty. The second set of choice scenarios tested a simple informational intervention in which in addition to the individual project scores and mean of the scores, participants were shown the variance of project scores. The variance is straightforward to infer from the individual scores in the first set of choice scenarios, so the second set of choice scenarios measures the effect of lowering the cognitive calculation costs and increasing the salience of score variance.

Table 5 compares behavior between the first and second choice scenarios. The results show that participants were substantially and significantly more variance averse when shown the variance. The G-MNL model in column (2) that nests responses from both sets of choice scenarios indicates that participants were twice as variance averse when the variance was displayed. The marginal effect of project variance in choice scenario set 2 is −0.061 versus −0.032 in set 1. The difference is significant at the 1% level. That the act of reporting variance, which should have made it easier for subjects to respond to the incentives of the

Table 5. Comparison of Choice Scenarios 1 and 2, Phase 1 Experiment

Dependent variable	(1) Choice scenario 2 data only Project choice	(2) Both choice scenarios 1 and 2 Project choice	(3) Choice scenario 2 data only Project choice	(4) Both choice scenarios 1 and 2 Project choice
Average Utility Weight				
Average project score	5.01***	4.82***	2.07***	2.48***
	(0.46)	(0.29)	(0.11)	(0.18)
Project score variance	−1.59***	−0.62***	−0.85***	−0.45***
	(0.16)	(0.085)	(0.073)	(0.051)
Average × Choice scenario 2		0.12		−0.41***
		(0.28)		(0.15)
Variance × Choice scenario 2		−0.92***		−0.40***
		(0.11)		(0.064)
Utility Weight Heterogeneity				
Average project score	2.68***	1.76***		
	(0.31)	(0.18)		
Project score variance	1.57***	0.81***		
	(0.13)	(0.086)		
Average × Choice scenario 2		1.14***		
		(0.23)		
Variance × Choice scenario 2		1.19***		
		(0.086)		
τ	0.046	0.27***		
	(0.042)	(0.045)		
Model	G-MNL	G-MNL	C-Logit	C-Logit
Subjects	150	150	150	150
Observations	13,500	27,000	13,500	27,000

Notes. The table shows results using choice scenario-level data from choice scenario sets 1 and 2 in the phase 1 experiment. The sample restrictions are indicated at the top of each column. Columns (1) and (2) are estimated using G-MNL models (Equation (1)). Columns (3) and (4) are estimated using conditional logit models. The outcome variable is an indicator for whether the project was chosen. "*Average project score*" is the average of the five scores for the project. "*Project score variance*" is the variance of the scores. Both explanatory variables are standardized. "*Choice scenario 2*" is an indicator equal to one if the data come from choice scenario 2 and zero otherwise. All models contain subject and choice scenario random effects in addition to the variables shown in the table. Standard errors, clustered at the subject level, are in parentheses.

***$p < 0.01$.

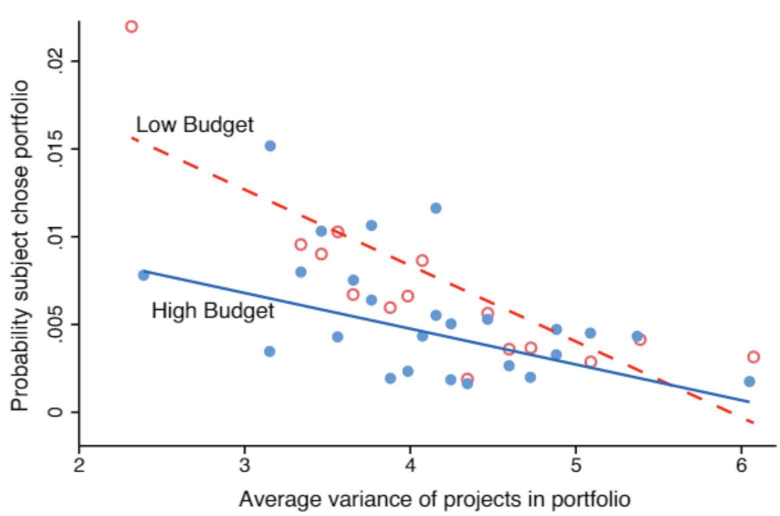
contest, was associated with more risk-averse choices is quite surprising. Two important caveats pertain to the results, however. First, participants were not randomized into being treated with choice scenario set 1 or 2 (all participants took part in both), so although the estimates come from a within-subject comparison, they still fall short of the ideal experiment and should be treated with some circumspection. Second, participants could have interpreted our display of variance as a form of experimenter demand, leading them to act in an apparently more variance-averse way.

5.2.3. Effect of Risk Diversification and Budgetary Pressure. Next, we examine behavior in a portfolio choice setting—the third set of choice scenarios—to determine whether the ability to diversify the set of projects affects risk taking. Consistent with the first two sets of choice scenarios, individuals continued to make variance-averse choices. We can see this result from a simple analysis of choices over similar portfolios. For a given budget, subjects could often construct two portfolios with identical expected values but different variances. For instance, with a budget of $12 million, there were two different portfolios with the highest-possible mean score (53.48), one with higher variance (15.14) and one with lower variance (8.44). Most subjects (71%) chose the lower-variance project when faced with this choice. Online Appendix B, Table B7 reports regression results that further corroborate this analysis. The point estimates suggest that individuals were slightly more variance averse when choosing portfolios than in the first set of choice scenarios, although the difference is not statistically significant at conventional levels. As in the other choice tasks, we do not know what assumptions the participants made about the correlation of outcomes, and this will likely impact decision making in a portfolio context. In practice, each choice entered the participant's score independently.

By randomizing the budget given to participants in each choice scenario, we are also able to assess the effect of budgetary pressure. Figure 3 shows that participants were more variance averse when they faced a smaller budget. For ease of presentation, we show the relationship for budgets that were less than the average (between $12 and $15 million) and for budgets that were greater than average (between $16 and $19 million). The difference in slope between the two fitted lines indicates that the effect of variance was smaller for choices made with larger budgets. This difference is statistically significant at the 1% level and shows that for two otherwise similar portfolios (same mean score, same cost), subjects were roughly twice as reluctant to choose a portfolio with a higher variance if they had a smaller budget than if they had a larger budget.[14]

5.2.4. What Participants Said About Their Own Behavior. After the third set of choice scenarios, participants were asked an open-ended question about their decision-making process.[15] The answers can help shed light on the results and heterogeneity reported. The most common answers indicated that the participant was treating the choices as a typical mean-variance

Figure 3. (Color online) Effect of Budget on Preference for Portfolio Variance



Notes. The figure shows the average effect of variance on portfolio choice (choice scenario set 3) in the phase 1 experiment broken down by budget. The hollow circles show the effect of variance on portfolio choice for low budgets (less than $15 million). The dashed line is a linear fit through the low-budget points. The solid circles show the same relationship for higher budgets (between $16 and $19 million). The solid line shows a linear fit through the high-budget points. All values are conditional on average project mean, average project cost, the interaction between average project mean and cost, the interaction between average project variance and cost, and choice scenario indicator variables.

trade-off, with typical respondents saying that they "[f]irst sought lowest variance with the highest average value." More than 70% of participants gave an answer along these lines (see Online Appendix B, Table B11).

One-fifth of participants indicated that they were variance loving. For example, one respondent wrote, "I went with the highest number and then often the highest number with the highest variance. Some people can have a difference of opinion." Some were even more explicit about how the R&D process is improved by judicious risk taking: "Consensus was not applied to my choices. When the respondents had a unanimous consensus, I took that to mean that the work was not groundbreaking. Therefore, I chose proposals that would challenge the experts and thus may drive at better outcomes, positive or negative." Most participants who indicated that they were variance loving also said that they looked at individual project scores and not just the mean or variance. Finally, participants who stated in their answers that they sought out variance were significantly more likely to exhibit variance-loving preferences in the first and second sets of choice scenarios, corroborating the self-reported statements.

Reassuringly, only about 1% of participants gave answers that were directly contradictory to the experimental instructions. A larger fraction—12% of participants—expressed some form of loss aversion, indicating that they left budget unspent out of concern that the remaining projects would result in losses or saying that projects with a high proportion of low scores might generate losses.[16] Loss aversion is a potential alternative explanation for the behavior we observe, and we designed the phase 2 experiment to test for the effect of loss aversion directly.

Table 6. Choice as a Function of Project Attributes and Treatment Arm: Phase 2 Experiment

Dependent variable	(1) Project choice	(2) Project choice
Average Utility Weight		
Average project score	6.43***	2.80***
	(0.65)	(0.28)
Project score variance	−0.36***	−0.31***
	(0.100)	(0.085)
Average × Arm 2	−1.25***	−0.69
	(0.20)	(0.43)
Average × Arm 3	−0.29	−0.52
	(0.18)	(0.35)
Variance × Arm 2	−0.23	−0.025
	(0.14)	(0.12)
Variance × Arm 3	−0.34***	−0.21
	(0.13)	(0.13)
Utility Weight Heterogeneity		
Average project score	3.75***	
	(0.53)	
Project score variance	0.70***	
	(0.064)	
Average × Arm 2	0.16*	
	(0.086)	
Average × Arm 3	0.52**	
	(0.20)	
Variance × Arm 2	0.11	
	(0.071)	
Variance × Arm 3	0.11***	
	(0.038)	
τ	0.16***	
	(0.014)	
Model	G-MNL	C-Logit
Subjects	140	140
Observations	12,600	12,600

Notes. The table shows results from estimating Equation (1) using choice scenario-level data from choice scenario set 1 in the phase 2 experiment. The outcome variable is an indicator for the chosen project. "*Average project score*" is the average of the five scores for the project. "*Project score variance*" is the variance of the scores. Both explanatory variables are standardized. The interactions are indicators for the experimental arm. The base category is the replication arm. "*Arm 2*" is the no loss framing treatment, and "*Arm 3*" is the objective costs treatment. All models contain subject and choice scenario random effects in addition to the variables shown in the table. Standard errors, clustered at the subject level, are in parentheses.
*$p < 0.10$; **$p < 0.05$; ***$p < 0.01$.

5.2.5. Phase 2 Experiment: Assessing Loss and Ambiguity Aversion. The phase 2 experiment consisted of a replication of the first set of choice scenarios from phase 1 (treatment arm 1), a replication of choice scenario set 1 from the phase 1 experiment but without any mention of potential losses (treatment arm 2), and a version of treatment arm 2 that further presented the project attributes as objective financial returns rather than review scores (treatment arm 3). The similar results between the phase 1 experiment and the arm 1 replication are reported in Section 5.1.

The second and third arms allow us to assess the effects of loss aversion and ambiguity aversion. Table 6 compares behavior across the arms and shows that preferences were largely the same in all arms. Participants had a statistically significantly lower preference for average project score in arm 2 compared with arm 1, but there was not a significant difference in variance preferences.

In addition, Online Appendix B, Table B10 adds elicited loss aversion and risk aversion parameters to the estimation model and shows that a larger loss aversion preference parameter is associated, if anything, with greater taste for variance. Variance aversion is also relatively unaffected by the inclusion of controls for elicited loss aversion. Together, these results indicate that loss aversion is not driving the variance aversion we find.

Arm 3 tested for the effect of ambiguity aversion. Table 6 shows that presenting the project attributes using explicit financial returns led participants to be more variance averse, if anything.[17] This result is consistent with the debriefing results and heterogeneity analysis showing that participants who treated their choices as financial portfolio problems sought to maximize mean while minimizing variance.

6. Discussion

Anemic research pipelines and the apparent slowdown of paradigm-shifting discoveries over the past quarter century have drawn considerable ire from both the research and investor communities. This has led to episodic concerns of policy makers regarding national scientific competitiveness and its role in shaping economic growth. If a small number of breakthrough research projects are responsible for a disproportionate amount of scientific progress, then research funders should target projects with greater uncertainty in order to have any chance of hitting upon rare but important results (Lotka 1926, Helpman 1998).

Although no experiment is a perfect simulacrum of complex decision making in the real world, in our setting, participants did not behave this way. They consistently chose lower-variance projects, despite incentives that expressly rewarded risk taking and that mirrored the risk-reward trade-off laid out. The results suggest that one possible reason for the lack of scientific breakthroughs could be the risk appetite of R&D managers. We found that subjects routinely made dominated decisions—choosing lower-variance projects even when projects with higher variance and the same mean score were available. These decisions caused excessively risk-averse subjects to leave money on the table. Comparing participants by the variance of the projects they actually chose, those subjects in the top quartile of variance were three times more likely to earn a reward than subjects in the bottom quartile.[18]

To highlight the effect of variance aversion and personal preferences, consider a typical project in the experiment that had an average rating of 3.5 of 5. The variance of ratings ranged between 0.28 and 2.28. A participant with an elicited coefficient of relative risk aversion of 1.185 (the 25th percentile in the sample) would be just over two percentage points less likely to choose the project with higher variance. In contrast, a participant with a preference parameter of 1.45 (the 75th percentile) would be more than five percentage points less likely to choose the project. On the other extreme, a participant with an elicited parameter of 0.93 (the 10th percentile) would be indifferent to project variance in this case.

To put this into context, we can construct a stylized example of how these differences could affect broader R&D. Patent citations and other measures of R&D performance have highly skewed distributions, with the top 10%–20% of inventions typically capturing more than 90% of returns (Scherer and Harhoff 2000, Silverberg and Verspagen 2007). If we imagine that scores of five in the experiment represent the 80th percentile of quality, then a project with mean of 3.5 and variance of 0.28 has only a 3% chance of yielding a score of five. The higher-variance project has a much higher 25% chance. Given the convexity in returns found in citation data, the expected value of a project that could yield \$1 million if successful is just \$65,000 for the low-variance project and \$250,000 for the high-variance project. Individuals with 75th percentile risk aversion, by preferentially choosing the lower-variance project, would lose out on \$10,000 in expectation (or 5.5% of the difference in expected value between the two projects). Individuals with 25th percentile risk aversion would lose 2.2% of the difference in value, whereas 10th percentile individuals would lose nothing in expectation.

These findings suggest that the personal preferences of those individuals in charge of research investments may be important for investment decisions. That is, who is placed in charge of research investment decisions may be as important as the incentives that firms provide them to make those decisions. As such, efforts to integrate risk preferences into the criteria driving hiring and promotion decisions could yield increases in the productivity of the R&D divisions in which they are employed. Whether this is best achieved through the importation of one of the many assessment approaches developed in economics and psychology, through new artificial intelligence tools that are beginning to be integrated into human resources departments (Li et al. 2020, McKinsey 2020), or some combination of the two will likely depend on firm characteristics and goals as well as the continued evolution of these tools over time (Cowgill 2018). The use of machine learning may be particularly well suited to firm management decisions to jointly determine contract incentives and the individuals who will face them.

Although the results suggest that incentives alone might not ensure appropriate risk taking, the effect on ultimate R&D investment depends on multiple factors. In practice, R&D managers can likely engage in dialogue with expert reviewers in an iterative process that could lead to better decision making. In our setting, we find that participants who looked more at individual scores outperformed those who only focused on simple summary statistics, so efforts to provide more context for those evaluations, such as encouraging more dialogue between reviewers and managers, could be beneficial.

At the same time, our finding that individuals with more R&D experience perform better suggests that the relevant decision-making skills can be learned and also suggests that real-world R&D firms might already be taking action to promote more appropriate risk taking. A critical question for managers is determining which employees are best positioned to acquire this knowledge and the best approaches to accelerate this learning. The recent evidence that suggests that entrepreneurship training, as distinct from standard business school training, can greatly improve decision making in highly uncertain domains (Lyons and Zhang 2018, Camuffo et al. 2020) offers some reason for optimism. How to optimize those for R&D choices has important implications for the advancement of science and the fate of research-intensive firms.

We conclude by noting that, as with all experiments, it is unclear how the experience in the "laboratory" generalizes to more realistic work settings. Two important considerations merit particular attention in our context. First and perhaps most importantly, we examined individual decision making, but many R&D decisions are made by teams. Although team risk preferences do appear to influence how teams respond to incentives to innovate (Graff Zivin and Lyons 2020), precisely how risk preferences are aggregated within a team and what that implies about group decision making remain open questions. Second, decisions in our experiment were one shot, whereas many firm-level decisions are sequential with opportunities for midcourse corrections and early project determination. Whether the trade-off of risk and reward in a more options-oriented framework differs significantly from those in our setting is equally unknown. Finally, it is important to recognize that not all additional risk taking is equally valuable to the firm. Efforts to encourage more radical forms of exploration must balance concerns regarding moral hazard and the relative values of more incremental forms of innovation, with the optimal mix likely to vary across firms and sectors. Together, these questions represent an area ripe for future research.

Acknowledgments
The authors thank three anonymous referees and seminar participants at Columbia University for helpful feedback and comments.

Endnotes

[1] See the personal communication with Hanneke Schuitemaker, PhD, VP, Head Viral Vaccine Discovery and Translational Medicine, Janssen Vaccines and Prevention B.V., Johnson and Johnson (Schuitemaker 2020).

[2] Limiting the study to upside risk means that our results do not speak to the possibly important role of losses (and loss aversion) in driving R&D decisions. Although all experimental arms involved only upside risk, we also explored the role of potential losses by varying whether losses were mentioned in the instructions presented to subjects and by eliciting loss-averse preferences.

[3] Theoretical work in this area continues to develop. For instance, Hopenhayn and Squintani (2021) study the role that R&D fads can play in diverting researchers away from potentially more valuable but riskier R&D areas.

[4] The participants are shown reviewer scores in the phase 1 experiment to mimic the ratings procedure used by some private sector firms, government agencies, and other organizations when deciding how to allocate R&D funds. This increased relevance to real-world R&D decisions comes at the cost of potential ambiguity in how participants viewed the scores. The phase 2 experiment is designed to address this potential ambiguity.

[5] More specifically, in both choice scenario sets 1 and 2, subjects engaged in 10 choice scenarios. For each choice scenario, they first selected their top and bottom choice from a set of four options. They then selected their second favorite choice from the remaining two options. We model this as three choice occasions per scenario, so there were four observations for the first choice occasion, three options in the second occasion, and two in the final occasion. For choice scenario set 3, the set of feasible portfolios determined the choice set faced by the subject in each of the eight choice scenarios. Feasible portfolios were those that had total cost less than or equal to the budget. Because budget was randomized, the size of the choice set varied by subject and choice scenario.

[6] Random utility theory was developed by Thurstone (1927) and underlies applications of the Method of Paired Comparisons (e.g., David 1988). Models for multiple choices were proposed by Luce (1959), and random utility theory was extended to statistical models for multiple discrete choices by McFadden (1974). Louviere and Woodworth (1983) proposed discrete choice experimental designs consistent with random utility theory.

[7] Basically, a DCE is a sparse, incomplete contingency (crosstab) table, one side of which represents the observed discrete choice options presented in the DCE. Thus, DCEs use experimental designs from the factorial family of combinatorics designs to create sets of choice options called choice sets. The experimental design provides the basis for creating the choice options and the choice sets to which they are assigned.

[8] This is a G-MNL type I model in the terminology of Fiebig et al. (2010) because the standard deviation of $\boldsymbol{\eta}_i$ is assumed to be independent of the scaling of $\boldsymbol{\beta}$. We make this assumption to speed convergence of the model and based on analyses that showed that this constraint led to superior model fit relative to the other choices of constraints commonly used in the literature (including not constraining the relationship between the standard deviation of η_i and the scaling of $\boldsymbol{\beta}$). These alternative results are available upon request.

[9] For the project choice questions, subjects ranked all projects by first choosing their first and fourth favorite projects and then choosing their second favorite project from the remaining two choices. In the analysis in Table 3, we treat these decisions as three separate choice scenarios. In the first scenario, the choice set is all four projects, and the subject's choice is their top ranked project. In the second scenario, the choice set is the three remaining projects after excluding the top ranked project, and the choice is their second ranked project. The third scenario's choice set is the remaining two projects, and the choice is the third ranked project. Results using just the first choice (of the most preferred project) are similar and available from the authors. Rank-based multinomial logit results are shown in Online Appendix B, Table B6.

[10] These are a score of 3 of 5 (corresponding to the 25th percentile of scores shown to subjects), a score of 3.5 of 5 (the median), and a score of 4 of 5 (the 75th percentile), respectively.

[11] However, as noted, these results could be driven by true risk preferences or correlates, including cognitive ability (Frederick 2005, Benjamin et al. 2013).

[12] Online Appendix C, Figure C2 shows that this effect is attenuated (although still significant at the 5% level) when all dimensions of

heterogeneity are included simultaneously. See the discussion at the end of this section.

[13] The marginal effect of score variance for participants with a 20% chance of winning was five percentage points more negative than the effect for other participants (95% confidence interval of 0.6–8.8).

[14] Regression estimates corresponding to the differences shown in the figure are in Online Appendix B, Table B7.

[15] The exact wording was as follows: "Briefly describe how you went about deciding which projects you put in the portfolios you wanted to fund."

[16] For example, one subject wrote, "Firstly, I will rank by the average score from high to low and prefer those with a significantly small variance. After that, I will check whether there is a possibility of extreme loss in this project. If so, I would like to not fund the project."

[17] The difference is statistically significant for the G-MNL model but not for the conditional logit in column (2). The point estimates from both models indicate substantially greater variance aversion on average.

[18] Because incentives were competitive, raising the variance of project choices would only have led to a larger expected payment, conditional on unchanged choices by other participants.

References

Abdellaoui M, Bleichrodt H, L'Haridon O (2008) A tractable method to measure utility and loss aversion under prospect theory. *J. Risk Uncertainity* 36(3):245–266.

Aghion P, Howitt P (1992) A model of growth through creative destruction. *Econometrica* 60(2):323–351.

Amabile TM (2017). In pursuit of everyday creativity. *J. Creative Behav.* 51(4):335–337.

Amit R, Zott C (2001) Value creation in e-business. *Strategic Management J.* 22(6–7):493–520.

Astebro T, Herz H, Nanda R, Weber RA (2014) Seeking the roots of entrepreneurship: Insights from behavioral economics. *J. Econom. Perspect.* 28(3):49–70

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

Benjamin DJ, Brown SA, Shapiro JM (2013) Who is 'behavioral'? Cognitive ability and anomalous preferences. *J. Eur. Econom. Assoc.* 11(6):1231–1255.

Bloom N, Jones CI, Van Reenen J, Webb M (2017) Are ideas getting harder to find? NBER Working Paper No. 23782, National Bureau of Economic Research, Cambridge, MA.

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

Camuffo A, Cordova A, Gambardella A, Spina C (2020) A scientific approach to entrepreneurial decision making: Evidence from a randomized control trial. *Management Sci.* 66(2):564–586.

Charness G, Gneezy U, Imas A (2013) Experimental methods: Eliciting risk preferences. *J. Econom. Behav. Organ.* 87:43–51.

Chu JS, Evans JA (2021) Slowed canonical progress in large fields of science. *Proc. Natl. Acad. Sci. USA* 118(41):e2021636118.

Cowgill B (2018) Bias and productivity in humans and algorithms: Theory and evidence from resume screening. Working paper, Columbia Business School, New York.

Dasgupta P, David PA (1994) Toward a new economics of science. *Res. Policy* 23(5):487–521.

Dasgupta P, Maskin E (1987) The simple economics of research portfolios. *Econom. J.* 97(387):581–595.

David HA (1988) *The Method of Paired Comparisons* (Griffin, London).

Fiebig DG, Keane MP, Louviere JJ, Wasi N (2010) The generalized multinomial logit model: Accounting for scale and coefficient heterogeneity. *Marketing Sci.* 29(3):393–421.

Frederick S (2005) Cognitive reflection and decision making. *J. Econom. Perspect.* 19(4):25–42.

Goel AM, Thankor AV (2008) Overconfidence, CEO selection, and corporate governance. *J. Finance* 63(6):2737–2784.

Graff Zivin J, Lyons E (2020) The effects of prize structures on innovative performance. NBER Working Paper No. 26737, National Bureau of Economic Research, Cambridge, MA.

Helpman E (1998) *General Purpose Technologies and Economic Growth* (MIT Press, Cambridge, MA).

Hirshleifer D, Low A, Teoh SH (2012) Are overconfident CEOs Better innovators? *J. Finance* 67(4):1457–1498.

Hopenhayn H, Squintani F (2021) On the direction of innovation. *J. Political Econom.* 129(7):1991–2022.

Imas A, Sadoff S, Samek A (2017) Do people anticipate loss aversion? *Management Sci.* 63(5):1271–1284.

Jones BF (2009) The burden of knowledge and the "death of the renaissance man": Is innovation getting harder? *Rev. Econom. Stud.* 76(1):283–317.

Kagan E, Leider S, Lovejoy WS (2020) Equity contracts and incentive design in start-up teams. *Management Rev.* 66(10):4879–4898.

Keupp MM, Palmié M, Gassmann O (2012) The strategic management of innovation: A systematic review and paths for future research. *Internat. J. Management Rev.* 14(4):367–390.

Krieger J, Li D, Papanikolaou D (2019) Missing novelty in drug development. NBER Working Paper No. 24595, National Bureau of Economic Research, Cambridge, MA.

Li D, Raymond LR, Bergman P (2020) Hiring as exploration. NBER Working Paper No. 27736, National Bureau of Economic Research, Cambridge, MA.

Lotka AJ (1926) The frequency distribution of scientific productivity. *J. Washington Acad. Sci.* 16(12):317–323.

Louviere JJ, Woodworth G (1983) Design and analysis of simulated consumer choice or allocation experiments: An approach based on aggregate data. *J. Marketing Res.* 20(4):350–367.

Louviere JJ, Flynn TN, Marley AAJ (2015) *Best-Worst Scaling: Theory, Methods and Applications* (Cambridge University Press, Cambridge, United Kingdom).

Luce RD (1959) On the possible psychophysical laws. *Psychol. Rev.* 66(2):81–95.

Lyons EE, Zhang L (2018) Who does (not) benefit from entrepreneurship programs? *Strategic Management J.* 39(1):85–112.

Madjar N, Greenberg E, Chen Z (2011) Factors for radical creativity, incremental creativity, and routine, noncreative performance. *J. Applied Psych.* 96(4):730.

March JG (1991) Exploration and exploitation in organizational learning. *Organ. Sci.* 2(1):71–87.

Marks AR (2011) Repaving the road to biomedical innovation through academia. *Sci. Translational Medicine* 3(89):89cm15.

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

McFadden D (1974) The measurement of urban travel demand. *J. Public Econom.* 3(4):303–328.

McKinsey (2020) The state of AI in 2020. Accessed January 14, 2022, https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/global-survey-the-state-of-ai-in-2020.

Merton RK (1973) *The Sociology of Science: Theoretical and Empirical Investigations* (University of Chicago Press, Chicago).

Munos BH, Chin WW (2011) How to revive breakthrough innovation in the pharmaceutical industry. *Sci. Translational Medicine* 3(89): 89cm16.

Porter ME (1985) *Competitive Advantage: Creating and Sustaining Superior Performance* (Free Press, New York).

Sauermann H, Cohen WM (2010) What makes them tick? Employee motives and firm innovation. *Management Sci.* 56(12):2134–2153.

Scherer FM, Harhoff D (2000) Technology policy for a world of skew-distributed outcomes. *Res. Policy* 29(4-5):559–566.

Schuitemaker H (2020) Personal communication, February 3.

Silverberg G, Verspagen B (2007) The size distribution of innovations revisited: An application of extreme value statistics to citation and value measures of patent significance. *J. Econometrics* 139(2):318–339.

Stephan PE (2010) The economics of science. Hall BH, Rosenberg N, eds. *Handbook of the Economics of Innovation*, vol. 1 (Elsevier, Amsterdam), 217–273.

Stern S (2004) Do scientists pay to be scientists? *Management Sci.* 50 (6):835–853.

Teece DJ (2010) Business models, business strategy, and innovation. *Long Range Planning* 43(2–3):172–194.

Thurstone LL (1927) A law of comparative judgment. *Psychol. Rev.* 34(4):273–286.

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

Tishler A (2008) How risky should an R&D program be? *Econom. Lett.* 99(2):268–271.