NBER WORKING PAPER SERIES

THE EFFECTS OF PRIZE STRUCTURES ON INNOVATIVE PERFORMANCE

Joshua Graff Zivin
Elizabeth Lyons

Working Paper 26737
http://www.nber.org/papers/w26737

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2020

We are grateful to Iwan Barankay, Gordon Hanson, Karim Lakhani, Craig McIntosh, participants at the Duke Workshop on Field Experiments in Strategy, Entrepreneurship, and Innovation and UC San Diego seminar participants for helpful feedback. This research was funded in part by the Ewing Marion Kauffman Foundation. The contents of this publication are solely the responsibility of the authors. This study is included in the AEA RCT Registry (AEARCTR-0004026). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Effects of Prize Structures on Innovative Performance
Joshua Graff Zivin and Elizabeth Lyons
NBER Working Paper No. 26737
February 2020
JEL No. J24,M54,O32

ABSTRACT

Successful innovation is essential for the survival and growth of organizations but how best to incentivize innovation is poorly understood. We compare how two common incentive schemes affect innovative performance in a field experiment run in partnership with a large life sciences company. We find that a winner-takes-all compensation scheme generates significantly more novel innovation relative to a compensation scheme that offers the same total compensation, but shared across the ten best innovations. Moreover, we find that the elasticity of creativity with respect to compensation schemes is much larger for teams than individual innovators.

Joshua Graff Zivin
University of California, San Diego
9500 Gilman Drive, MC 0519
La Jolla, CA 92093-0519
and NBER
jgraffzivin@ucsd.edu

Elizabeth Lyons
School of Global Policy and Strategy
University of California, San Diego
9500 Gilman Drive, MC 0519
La Jolla, CA 92093-0519
lizlyons@ucsd.edu

A randomized controlled trials registry entry is available at
https://www.socialscienceregistry.org/trials/4026

1 Introduction

The modern firm is relegating more of its routine tasks to machines and orienting employees toward increasingly creative undertakings. Given the circuitous path between effort and outcome, what is the best way to encourage innovation? Do winner-takes-all style incentives foster the right mix of effort and risk taking? Or are incentives that reward a greater number of contributors more desirable given firm objectives?

To be clear, nearly everyone seems to have an opinion. CTOs and management gurus extoll the virtues of a wide range of institutional features to get the creative juices flowing.[1] Academic economists, on the other hand, tend to focus on the principal agent problem that governs all innovative endeavors and the remedies to overcome them (e.g. Wright, 1983; Scotchmer, 2004). Nearly all solutions rely upon some combination of performance-based-pay and risk sharing between employers and employees, with the optimal mix hinging on key assumptions regarding the ability and ambitions of employees, their risk preferences, and how costly it is for them to supply effort (Clark and Riis, 1998; Moldovanu and Sela, 2001). Thus, what works best in practice is largely an empirical question and the empirical literature in this space is surprisingly thin.

This paper is designed to fill this void by presenting evidence from an experiment that we ran with Thermo Fisher Scientific, a major life sciences company. In particular, we organized an innovation contest in which participants were asked to design solutions to help share medical equipment across small providers in Mexico and were randomized into two distinct compensation schemes. In the winner-takes-all arm, participants were provided with high-powered incentives to innovate but no insurance for inferior solutions. In the other arm, the ten best proposals received some form of compensation (with the same total payout as the other arm), providing some insurance for participants that their efforts could be rewarded even if their proposals were not quite best-in-class.

Our results reveal a significant impact of incentive structure on the quality of innovative output, with no discernible impact on the quantity of output. In both study arms, approximately one-third of participants submitted a proposal for evaluation. The quality of submission was evaluated by a panel of experts on five distinct dimensions, including the novelty of the proposal relative to other products available in the marketplace. While the two groups did not statistically differ from one another on their overall scores, participants under the winner-takes-all compensation scheme submitted proposals that were significantly more novel than their counterparts in the other compensation scheme. Thus, the risk taking encouraged by the competition with a single prize appears to have driven would be innovators to pursue more creative solutions.

[1] See Henry (2018) and Le Merle and Davis (2017) for two recent examples from the popular management literature.

An important feature of our experiment is that participants could elect to compete as an individual or as a team. The flexibility to choose the composition of one's team was deliberate and intended to provide a reasonable simulacrum of the workplace. As one might expect, teams were generally assembled to diversify skill-sets and deepen professional experience. They did not differ from individuals in terms of their risk appetites, either because individuals did not think they were sufficiently important to inform team composition or because risk preferences were too costly to credibly observe at the time of team formation.[2] While team composition was endogenous by design, the requirement that they be formed before the randomization prize structure enables us to estimate the causal impacts of compensation on team performance.[3]

We find that the output of teams does not statistically vary across compensation schemes, but that the solutions they develop under the winner-takes-all compensation scheme are significantly more novel than those developed under the scheme with more diffuse incentives. In contrast, individual performance does not vary on any dimension under the two compensation schemes. Moreover, consistent with the winner-takes-all prize structure requiring that innovators take on more risk, we find that participants with lower levels of risk aversion produce significantly more novel innovations in the single prize contest than they do in the multiple prize contest.

The 21st century economy is one that prizes novelty. Firms view it as an important source of comparative advantage. It is also an essential ingredient in the development of technological breakthroughs that transform markets and generate large amounts of producer and consumer surplus (Helpman, 1998; Mansfield et al., 1977; Shane, 2001). At the same time, most firms struggle to generate novel innovations and those that manage to succeed often do so at great expense (Azoulay et al., 2019; Krieger et al., 2018; Nanda and Rhodes-Kropf, 2016). The results in this paper have potentially far-reaching implications for the design of institutions and incentives to foster more novel innovation. Providing sizable rewards for only the very top performers appears to inspire the sort of risk-taking required to explore new unproven approaches rather than the exploitation of well-known ones for more incremental progress (Manso, 2011; March, 1991). Insofar as participants interpret the winner takes all contest arm as more competitive, our findings are also consistent with Boudreau et al. (2011) who find that greater competition in software design can lead to more novel solutions when the best way to tackle the problem is more uncertain. Moreover, since compensation schemes do not appear to reduce output levels, it appears that inducing more radical innovation may be less expensive than one might have predicted based on the literature that highlights the discouraging effects of competition on effort (e.g. Cason

[2]Teams were also more likely to include a female participant, consistent with prior literature indicating that women prefer to work in teams (Healy and Pate, 2011).

[3]Since team formation is endogenous, we cannot, however, draw any causal inferences by comparing the performance of teams versus individuals under any particular given compensation scheme.

et al., 2010; Fullerton and McAfee, 1999; Taylor, 1995). Whether these insights generalize to more complex tasks and projects of longer duration remains an open question.

The remainder of our paper is organized as follows. Section 2 describes our study setting and the design of our experiment. Section 3 provides details on our data and econometric strategy. Section 4 present our results and Section 5 offers some brief concluding remarks.

2 Research Setting and Experimental Design

In order to test how prize structure impacts the quantity and quality of innovation, we ran a randomized control trial (RCT) within an innovation contest that we hosted in partnership with Thermo Fisher Scientific, a large biotechnology company with a market cap in excess of $100 billion US. The innovation contest was hosted by their Mexico office in Baja California and was open to all non-management employees of the firm as well as employees at other technology firms in the region.[4] To increase participation and help foster Thermo Fisher's recruitment interests, it was also promoted to STEM students at local universities.

The contest was advertised over a 45-day period. Promotion materials included information about the general topic area of the innovation challenge, the competition dates, and the total prize purse available to participants. The promotion materials also informed potential participants that the contest was being co-hosted by UC San Diego and Thermo Fisher, and that it was part of a research study on motivations for innovation.[5] Participation was open to individuals or teams of up to three people.

At the start of the competition, the innovation challenge details were announced and participants were given 54 hours (from 6 pm on a Friday until midnight the following Sunday) to submit their entries. Submissions were made through DevPost, a popular commercial platform for hosting software innovation contests. The challenge was focused on addressing local health technology needs, with the specifics determined through a consultative process between the study authors and research managers at Thermo Fisher to ensure commercial relevance to the industry. The contest problem was carefully chosen to ensure that reasonable progress could be made during the time allotted for the competition.

In particular, participants were provided with the following text at the opening of the competition window: *Mexico has many small health care providers and research and clinical laboratories that, on their*

[4]Baja California is a Mexican state that borders California, USA. Thermo Fisher has an R&D office in the state and is working with local stakeholders to develop the region's STEM labor force.

[5]We were required to disclose that the contest was part of a research study by UC San Diego's Institutional Review Board. We opted to disclose during recruitment rather than after the competition was complete because ex post disclosure would require that participants are given the option to remove themselves from the study and we were concerned that this could lead to selective attrition based on competition outcomes.

own, cannot afford expensive equipment that would allow them to provide the highest quality care possible. We believe that the proliferation of digital and cloud technologies can help to solve this problem. We are asking you to show us how you think these technologies can be used to support access to high-quality medical equipment even for these small health care providers and labs.

To generate random variation in the prize structure, we randomly assigned participants to one of two prize menus both with a total of 15,000USD available to contest winners, corresponding to approximately 79% percent of 2018 annual incomes for software developers in Mexico (Statista, 2019). The first prize structure was a winner-takes-all design in which a single prize of 15,000USD would be given to the highest ranked submission. The second prize structure, provided awards to the ten highest ranked submissions. Submissions ranked first, second, third, and fourth received $6,000, $3,000, $1,500, and $900 respectively, and submissions ranked fifth to tenth received $600. Given an equal number of competitors in both study arms, the expected return for would be innovators is identical across the two arms, but competitors under the winner-takes-all arm faced a higher risk of failure.

Randomization was performed following the enrollment deadline and stratified by team and individual participants. Participants were given information about the prize structure they would face at the same time they were provided details on the innovation challenge. Judges were told about the different prize structures at the same time the participants were to ensure they did not disclose the prize structures to participants beforehand.[6] To avoid concerns that participants would feel betrayed if they only learned about the alternative prize structures through incidental conversations with other competitors, we disclosed the design upfront. Participants were told that the contest organizers had disagreed over the optimal prize structure and, as a result, had decided to randomly divide participants into two separate and equally sized groups with distinct prize structures. They were also assured that they would only be judged relative to others facing the same prize structure and therefore would only be competing with half of the total participant pool. As Appendix Tables A1 and A2 demonstrates, assignment into contest arms is unrelated to participant characteristics.

Participants were instructed to turn in their complete or incomplete computer scripts, written explanations, and any other non-script output by the end of the competition deadline in order to be eligible for a prize. Contest submissions were judged by six industry experts, including high-level managers at Thermo Fisher, Teradata (a software company headquartered in San Diego, California), and computer science fac-

[6]The exception to this was one of the Thermo Fisher judges who was involved in the planning of the contest and was aware there would be two contest arms. However, she was not told who would be placed in which arm, and we have no evidence that she disclosed any information about the contest prizes to participants.

ulty who actively consult with technology companies in the Baja region. Submissions were judged on a 5-point scale of across five, equally weighted categories: novelty relative to existing products on the market, functionality, user friendliness, the scope of use cases, and the degree to which it addresses the innovation challenge. A detailed description of the scoring categories and criteria is provided in the Data Appendix of this paper.

All submissions were reviewed by 3 of the 6 judges to whom they were randomly assigned. To ensure comparability of judge rankings across prize structures, all submissions were pooled before being randomly assigned to judges. Judges were blinded to all information about the incentive structure under which proposals were submitted. As advertised to participants, awards were determined by rank *within* each study arm.

Our experiment design allows us to control for selection into contest participation based on prize structure. In addition to deciding whether to enroll in the competition prior to prize structure randomization, all participants were required to decide whether they would like to compete as a team or as an individual before prize structures were allocated. They also completed a pre-contest survey under the same conditions. This timing ensures the following three features in our empirical analysis: 1) we are able to observe differences in effort and performance across prize structures among statistically identical populations; 2) our measures of participant characteristics are not biased by the experimental treatment; and 3) selection into teams is not affected by the prize structures.

3 Data and Analysis Plan

A total of 184 individuals signed up for the contest, of whom 91 signed up to participate in a total of 39 teams and 93 signed up to participate on their own.[7] All participants are included in our analysis. Before participants were permitted to register for the contest, they were required to complete a survey that asked for some basic demographic information along with questions about their professional expertise. We also elicited risk preferences from each contestant. Our risk preferences question is based on the Eckel and Grossman measure (Eckel and Grossman, 2002) with the degree of risk aversion taking on a numeric value ranging from 1-5 with higher levels of risk aversion corresponding to lower numbers.[8] The full list of survey questions is

[7]The roughly equal number of team and individual participants is a coincidence and not something we coordinated or anticipated.

[8]We also asked participants to assess their capabilities as a programmer relative to others with similar expertise. This proxy measure for confidence was intended for a distinct study aim designed to randomize information provision about the skills of competitors to examine whether information about relative capabilities would change performance differentially under the two prize structures. Unfortunately, this study aim was abandoned due to insufficient sample size. More details on that proposed aim can be found in the trial history of our RCT registration documents, which can be found at https:

provided in our Data Appendix.

While our measures of participant characteristics are straightforward for solo competitors, assessing them at the team level is more challenging. Most of our core demographic measures – student status, employment status, education, and age – are defined by the average across team members. To capture team gender composition, our female variable is equal to one if any team member is a female.[9] The other team characteristics, which form the basis of our heterogeneity analysis, are measured as follows. Because average expertise is not a useful measure of team skills, we define it as a count of the non-overlapping areas of expertise among team members. Prior contest experience is coded as a binary variable and is set equal to one if any team member had previously participated in an innovation contest. We eschew an average measure to avoid a definition whereby a team could only have the same level of experience as an individual when all of its members had the same level of experience. Finally, risk preferences is defined as the average of individual responses.

As can be seen in Panel A of Table 1, roughly one-quarter of the sample is female, close to half are students with the vast majority of the other half employed either in full-time or part-time jobs. The average participant is in the 25-34 age range, and has between some college education and a Bachelor's degree. As Panel B demonstrates, participants have expertise in 3 of the 8 categories that we had ex-ante identified as relevant for competition success, approximately 32% have some prior innovation contest experience, and the average participant is on the higher end of our risk aversion scale, indicating relatively low risk aversion.

Our contest outcomes of interest are the quantity and quality of innovative output. Our measure of quantity is a simple indicator for whether or not participants submitted a proposal for evaluation by the judges. As Panel C of Table 1 demonstrates, approximately one third of participants submitted a project by the contest deadline. Our primary measures of the quality of innovative output are the overall project rank, and the project novelty rank. Both measures are conditional on a project being submitted for evaluation by the judges. The overall rank measure is appealing because it places equal weight on all five of the categories that we asked the judges to evaluate and is the basis on which prizes were awarded. The novelty measure is of particular interest since that is the primary focus of most R&D units and the one category where we had an a priori clear hypothesis about the role of our compensation schemes. Novel innovations require more risk taking and increase the likelihood of both big successes and big failures. The winner-takes-all prize structure creates strong incentives for those outlier outcomes and therefore should lead to more novel output.[10]

//www.socialscienceregistry.org/trials/4026.

[9] Only one team is made up of all females (a team of two).

[10] For completeness, we also analyze whether the prize structure had an effect on the other evaluation criteria (functionality, user friendliness, wide scope of use cases, and addresses contest problem) though we did not have a priori expectations about

We generate the former measure first by calculating the average rating across all five categories for each judge, then by ranking these averages for each judge, and finally by averaging these rankings across judges who evaluated each submission. We favor a ranking-based measure over an average score measure because it controls for judge-specific differences in how scores are interpreted in a straightforward way. Our results are largely unchanged if we use normalized scores by judge-specific means and standard deviations before averaging across judges (see Appendix Table A3).

We generate our measure of novelty by averaging a submission's novelty rank across judges. As with all judgement criteria, novelty is evaluated on a scale from 1-5. Importantly, novelty is evaluated relative to what is currently and/or soon to be available on the market with the lowest possible score being given for "proposed solutions already available in the target market" and the highest possible score being given for "proposed solutions that are different than anything currently available in the target market and that are so creative judges are almost sure no one else has thought of a similar idea." For simplicity of interpretation, a higher ranking indicates a higher quality submission.

In order to evaluate the impact of prize structure on the quantity and quality of innovativeness, we compare average submission probabilities, and the overall rank and novelty rank by prize structure. Given the success of our randomization (see Appendix Tables A1 and A2), mean comparisons are sufficient to estimate the causal impact of prize structure on innovative performance. We then separately analyze impacts for teams and solo competitors to explore potentially heterogenous responses to incentives along this dimension. Recall that, while the composition of teams is endogenously determined, teams are formed before randomization takes place, and thus a comparison of performance across prize structures within participants competing as a team or within the solo competitor group will continue to provide causal estimates of the impacts of our compensation schemes. Given the important role of risk taking for the development of successful innovations, and the differential incentives to take those risks under the different prize structures, we also examine the degree to which risk preferences shape performance.[11]

4 Results

We begin with our core findings in Table 2, which presents comparisons of mean outcomes by prize structure. Interestingly, despite post-survey responses indicating that people in both prize arms prefer the multiple prize

how our experimental treatment would change these outcomes. Those results are presented in Appendix Table A4.

[11]The separate analysis for teams was part of the inspiration for our experimental design and described in the pre-registration documents for our RCT. We collected data on participant risk preferences with the goal of testing the hypothesis that less risk averse participants would pursue riskier innovation paths and thus be more likely to create novel output. Due to an unfortunate oversight, we failed to include this analysis in our study's pre-registry.

structure (see Table A5), the number of participants who submit an innovation for evaluation by our panel of judges is the same for the single prize and the multiple prize regimes. In both arms, approximately one third of participants felt that their innovation was sufficiently well developed to subject themselves to evaluation by our expert judges.[12] Moreover, the average overall quality of these submissions was statistically similar across prize structures, receiving a score of roughly 2.5 out of a maximum attainable score of 5.

While overall innovation quality is similar between the two prize structures, submissions made under the one prize structure were significantly more novel relative to those made under the multiple prize structure. It is worth reiterating that this measure of novelty is a market-based measure as judges were asked to assess novelty relative to other products available in the market. That innovators under the single prize structure performed better on this metric is consistent with our hypothesis that the strong incentives to generate outlier solutions under this compensation scheme may have led competitors to take more risks (something we will probe further later) and thus generate more novel output. As displayed in Appendix Table A4, submission quality for all other evaluation criteria does not statistically differ across compensation schemes.

Next, we turn our attention to heterogeneity. We begin with a focus on teams, in part, because prior evidence suggests that teams respond differently to competition than individuals (Charness and Sutter, 2012), and that teams are more capable of innovating than individuals (Jones, 2009). Recall that, while the decision to participate as a team or individual is made before randomization and is thus independent of the prize structure, the composition of teams is endogenously determined and reflects individual preferences for teamwork relative to independent work. This endogeneity is desirable as it offers a better reflection of how teams are actually formed within firms, where membership is flexible and teammates generally know one another beforehand (Thompson and Choi, 2006).

As shown in Table 3, teams appear to be constructed with individual team member capabilities in mind, spanning a broader set of skills and encompassing more experience than their individual counterparts. In particular, teams are about 66% more likely than individuals to have prior contest experience (p-value=0.10). They also have a larger number of combined areas of expertise, averaging 3.7 out of 8 relevant domains which is more than 40% higher than the expertise of a typical individual competitor. Given the inherent challenges in observation and verification at the time of team formation, it is perhaps unsurprising that teams do not statistically differ from solo competitors in their risk preferences.[13]

[12] Consistent with the quantity of output being the same under both prize structures, the percentage of participants who registered on the contest DevPost page is statistically the same under both (46% in the multiple prize structure and 52% in the one prize structure), and the likelihood of submitting a project conditional on registering is also the same in the two arms. These findings suggest that, at least at the extensive margin, effort was the same in both prize structures

[13] While teams and individuals differ along the characteristics shown in Table 3, it is important to note that team characteristics are balanced across the two prize structures.

While this endogenous selection into teams implies that we cannot causally identify differences in performance across teams and individuals, the results presented in Table 4 suggest that, consistent with this deliberative process of team formation, teams are more productive and more creative than individual innovators. Perhaps more importantly, the random assignment of teams and individuals to prize structures implies that an analysis of impacts within teams or within individuals allows us to draw causal inferences. Here we see that teams and individuals do, in fact, respond differently to the two prize structures. In particular, team participants submit significantly more novel projects under the single prize structure (approximately 25% more than under the multiple prize) whereas individual participants submit similarly novel projects under both prize structures. Together they suggest that our core finding from Table 3 is being driven by the responsiveness of teams to the incentives for radicality embedded in the single prize structure.

Although the results in Table 3 imply that the performance of teams may simply reflect their superior experience and expertise, the similarity in risk appetites across individuals and teams allows us to examine its role independently. The willingness to take risks has long been associated with success in innovation because innovation is a fundamentally uncertain process (e.g. Fellner, 1966) . Moreover, this uncertainty is larger for more novel innovations because it is harder to compare them with existing pathways and solutions relative to more incremental innovations (Manso, 2011; Sunder et al., 2017).

It is important to note at the outset that our sample size does not permit us to analyze heterogeneous treatment effects by risk preferences separately for teams and individuals.[14] While we cannot rule out that team formation did not take risk preferences into account in some unobservable way not captured by Table 3, given the theoretical uncertainty associated with the optimal risk preference make-up of innovation team members (e.g. Masclet et al., 2009) and the difficulty with which individuals could credibly evaluate it at the time of team formation, we believe it is reasonable to combine teams and individuals for this analysis. Our contention notwithstanding, since risk preferences were not randomly assigned across teams and individuals our pooled analysis should be interpreted with some caution.[15]

To study how risk aversion interacts with our prize structures, we compare the quantity and quality of output across prize structures by splitting the sample into above and below median sample competitor risk aversion levels.[16] These results are presented in Figure 1. Risk preferences do not appear to impact the

[14]Our pre-analysis plan called for a sample size roughly twice that of what we were able to achieve.

[15]We verify that our mean comparison results by risk aversion are robust to controlling for whether or not an observation represents a team or an individual (see Table A6).

[16]In our sample, the median risk aversion is 3 out of 5, so we classify participants with risk aversion above 3 as low risk aversion, and those with 3 or lower as high risk aversion. On our risk aversion scale, 1 represents very high risk aversion, and 5 represents risk neutrality or risk loving. The numbers in between, 2-4, represent declining degrees of risk aversion. Given this, a risk aversion of greater than 3 can be thought of as relatively low risk aversion or risk neutral.

quantity of output nor the overall quality of that output. However, consistent with our expectations, less risk averse participants respond more to the single prize structure. Specifically, those with below sample median risk aversion score about 25% higher on novelty under the single prize structure relative to the multiple prize structure (p-value=0.05). In contrast, those with above median risk aversion do not submit significantly more novel output under the single prize structure.[17]

5 Conclusion

In this paper, we examine the impacts of compensation schemes on innovative output. Our evidence is derived from an innovation experiment that we ran in partnership with Thermo Fisher Scientific, a major life sciences company. In the experiment, scientists were randomized to one of two competition arms with identical aggregate financial resources and then asked to develop a program to facilitate technology sharing applications for small medical providers. Participants in the winner-takes-all tournament faced high-powered incentives to innovate but received no rewards for second-best solutions. In the other arm, participants faced more diffuse incentives, which insured against near misses by spreading out financial rewards across the ten best proposed solutions. Consistent with the embedded incentives for risk-taking, we find that the winner-takes-all prize structure generated significantly more novel output.

Our analysis of heterogeneity reveals further nuance. The elasticity of creativity with respect to compensation schemes is much larger for teams, with the payoff from assembling a diverse team to address the scientific 'burden of knowledge' problem (Jones, 2009) unleashed under the winner-takes-all regime. At the same time, the risk appetites of would-be-innovators also plays an important role. As with a recent study of MBAs, it seems that it is hard to fight nature (Shrader et al., 2019). Those with a greater aversion to risk are less able to pursue the less proven problem-solving strategies that lead to more novel solutions, even when the highly skewed compensation scheme incentivizes them to do so.

Our results have potentially far-reaching implications for the design of institutions and incentives to foster radical innovation. Providing sizable rewards for only the very top performers appears to inspire the sort of risk-taking required to encourage the requisite creativity that delivers scientific and technological novelty. Moreover, since the additional risk under the winner-takes-all compensation scheme did not appear to reduce output levels, it appears that this more radical innovation can be obtained at relatively low cost.

At the same time, it is important to recognize that incentives alone are insufficient to spark creativity.

[17] The astute reader will note that we elicited risk preferences using two distinct questions. This pattern or results is unchanged if we use our less preferred risk preference question which relies upon fewer hypothetical comparisons to determine an individuals risk appetite.

Genius is not created by incentives, but empowered by them. That teams are better able to respond to those incentives is consistent with broader trends in science (Wuchty et al., 2007), but much more work is required to understand the raw ingredients that shape the relationship between creativity and compensation schemes. Whether the insights from our experiment generalize to more complex tasks with less well-defined avenues for intellectual exploration or to projects of longer duration that provide greater opportunities to learn from early failures, are also fruitful areas for additional research.

6 References

Azoulay, Pierre, Erica Fuchs, Anna P Goldstein, and Michael Kearney, "Funding breakthrough research: promises and challenges of the "ARPA Model"," *Innovation Policy and the Economy*, 2019, *19* (1), 69–96.

Boudreau, Kevin J, Nicola Lacetera, and Karim R Lakhani, "Incentives and problem uncertainty in innovation contests: An empirical analysis," *Management science*, 2011, *57* (5), 843–863.

Bruhn, Miriam and David McKenzie, "In pursuit of balance: Randomization in practice in development field experiments," *American economic journal: applied economics*, 2009, *1* (4), 200–232.

Cason, Timothy N, William A Masters, and Roman M Sheremeta, "Entry into winner-take-all and proportional-prize contests: An experimental study," *Journal of Public Economics*, 2010, *94* (9-10), 604–611.

Charness, Gary and Matthias Sutter, "Groups make better self-interested decisions," *Journal of Economic Perspectives*, 2012, *26* (3), 157–76.

Clark, Derek J and Christian Riis, "Competition over more than one prize," *The American Economic Review*, 1998, *88* (1), 276–289.

Eckel, Catherine C and Philip J Grossman, "Sex differences and statistical stereotyping in attitudes toward financial risk," *Evolution and human behavior*, 2002, *23* (4), 281–295.

Fellner, William, "Profit maximization, utility maximization, and the rate and direction of innovation," *The American Economic Review*, 1966, *56* (1/2), 24–32.

Fullerton, Richard L and R Preston McAfee, "Auctionin entry into tournaments," *Journal of Political Economy*, 1999, *107* (3), 573–605.

Healy, Andrew and Jennifer Pate, "Can teams help to close the gender competition gap?," *The Economic Journal*, 2011, *121* (555), 1192–1204.

Helpman, Elhanan, *General purpose technologies and economic growth*, MIT press, 1998.

Henry, Todd, *Herding Tigers: Be the Leader That Creative People Need, By Todd Henry. New York City: Portfolio, 2018*, Penguin Random House, 2018.

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

Krieger, Joshua, Danielle Li, and Dimitris Papanikolaou, "Missing novelty in drug development," *NBER Working Paper*, 2018, (w24595).

Mansfield, Edwin, John Rapoport, Anthony Romeo, Samuel Wagner, and George Beardsley, "Social and private rates of return from industrial innovations," *The Quarterly Journal of Economics*, 1977, *91* (2), 221–240.

Manso, Gustavo, "Motivating innovation," *The Journal of Finance*, 2011, *66* (5), 1823–1860.

March, James G, "Exploration and exploitation in organizational learning," *Organization science*, 1991, *2* (1), 71–87.

Masclet, David, Nathalie Colombier, Laurent Denant-Boemont, and Youenn Loheac, "Group and individual risk preferences: A lottery-choice experiment with self-employed and salaried workers," *Journal of Economic Behavior & Organization*, 2009, *70* (3), 470–484.

Merle, Matthew C Le and Alison Davis, *Corporate Innovation in the Fifth Era: Lessons from Alphabet/Google, Amazon, Apple, Facebook, and Microsoft*, Cartwright Publishing, 2017.

Moldovanu, Benny and Aner Sela, "The optimal allocation of prizes in contests," *American Economic Review*, 2001, *91* (3), 542–558.

Nanda, Ramana and Matthew Rhodes-Kropf, "Financing risk and innovation," *Management Science*, 2016, *63* (4), 901–918.

Shane, Scott, "Technological opportunities and new firm creation," *Management science*, 2001, *47* (2), 205–220.

Shrader, J, R Carson, J Louviere, S Sadoff, and J Graff Zivin, "Fear of Risk Hinders Research Investment," *mimeo UCSD*, 2019.

Statista, "Average salary in the Software Development industry in Mexico in 2018, by position (in 1,000 Mexican pesos)," 2019.

Sunder, Jayanthi, Shyam V Sunder, and Jingjing Zhang, "Pilot CEOs and corporate innovation," *Journal of Financial Economics*, 2017, *123* (1), 209–224.

Taylor, Curtis R, "Digging for golden carrots: An analysis of research tournaments," *The American Economic Review*, 1995, pp. 872–890.

Thompson, Leigh L and Hoon-Seok Choi, *Creativity and innovation in organizational teams*, Psychology Press, 2006.

Wuchty, Stefan, Benjamin F Jones, and Brian Uzzi, "The increasing dominance of teams in production of knowledge," *Science*, 2007, *316* (5827), 1036–1039.

Table 1: Summary statistics

Variable	Mean	(Std. Dev.)	N
Panel A: Participant Demographics			
Student	0.456	(0.486)	132
Employed	0.405	(0.476)	132
Female Participant/Group Member	0.235	(0.426)	132
Age (1-5)	1.809	(0.813)	132
Highest Level of Education (1-6)	3.696	(0.997)	132
Panel B: Participant Characteristics			
Single Prize Contest	0.500	(0.502)	132
Signed Up as Team	0.295	(0.458)	132
Any Prior Contest Experience	0.364	(0.483)	132
Number of Areas of Relevant Expertise (1-8)	2.924	(2.044)	132
Risk Preferences (1-5)	2.868	(91.285)	132
Panel C: Outcomes			
Submitted a Project	0.318	(0.468)	132
Overall Rank (1-5)	2.592	(0.832)	42
Novelty Rank (1-5)	2.923	(0.963)	42

Notes: For team participants, demographics are averaged across teams except for the female variable which is equal to one if any team member is a female. For team characteristics presented in Panel B, risk preferences are equal to the average of individual responses, number of areas of relevant expertise is equal to a count of the non-overlapping areas of expertise among team members, and prior contest experience is equal to one if any team member had previously participated in a contest. Age is categorized into 5 bins where 1 equals 18-24, 2 equals 25-34, 3 equals 35-49, 4 equals 50-59, and 5 equals 60-69. Highest Level of Education is categorized into 6 bins where 1 represents less than high school, 2 is high school, 3 is some college or vocational training, 4 is a Bachelor's degree, 5 is a Master's degree, and 6 is a PhD or equivalent. A submission's Overall Rank is equal to the within-judge rank of the average rating assigned to the five evaluation criteria, averaged across judges who evaluated the submission. A submission's Novelty Rank is the average novelty rating rank across judges who evaluated the submission. Overall Rank and Novelty Rank are conditional on a project being submitted for evaluation by a judge. For both rank measures, a higher rank is associated with a higher quality submission.

Table 2: Outcomes by Prize Structure

	Multiple Prizes	One Prize	p-value of difference
Submitted a Project	0.303	0.333	0.711
	(0.057)	(0.058)	
Overall Rank	2.428	2.7842	0.227
	(0.211)	(0.150)	
Novelty Rank	2.608	3.208	0.042**
	(0.230)	(0.175)	

Notes: A submission's Overall Rank is equal to the within-judge rank of the average rating assigned to the five evaluation criteria, averaged across judges who evaluated the submission. A submission's Novelty Rank is the average novelty rating rank across judges who evaluated the submission. Overall Rank and Novelty Rank are conditional on a project being submitted for evaluation by a judge. For both rank measures, a higher rank is associated with a higher quality submission. The statistics reported in the p-value of difference column are the p-values from tests of equality between the single prize and multiple prize contest arms. Standard errors are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 3: Characteristics by Team and Individual Participants

	(1) Participate as Individual	(2) Participate as Team	(3) p-value of difference
Any Prior Contest Experience	0.312	0.463	0.103
	(0.048)	(0.081)	
Unique Areas of Relevant Expertise	2.581	3.744	0.003***
	(0.182)	(0.361)	
Risk Preferences	2.892	2.811	0.744
	(0.146)	(0.150)	
N	93	39	

Notes: Any Prior Contest Experience is equal to one if the individual has prior innovation contest experience and zero otherwise in column 1; and equal to 1 if any team member has prior contest experience and zero otherwise in column 2. Unique Areas of Relevant Expertise corresponds to the number of unique areas of expertise (from a pre-specified list of relevant domains) held by an individual or the combined members of the team.. Risk Preferences ranges from 1-5 with 1 representing the highest level of risk aversion and 5 representing risk neutrality or risk loving based on our risk preference elicitation tool. The statistics reported in the p-value of difference columns are the p-values from tests of equality between the individual and team characteristic means. Standard errors are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

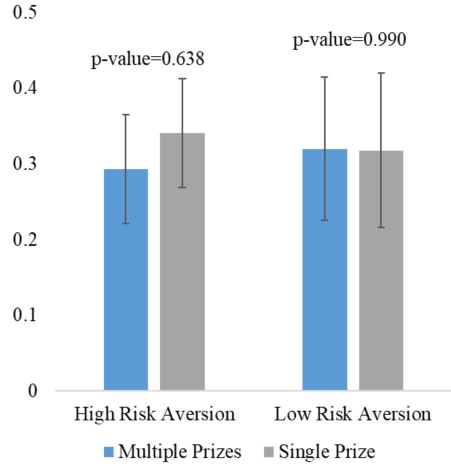
Table 4: Heterogeneous Impacts of Prize Structure by Teams

	No Team			**Team**		
	Multiple Prizes	One Prize	p-value of difference	Multiple Prizes	One Prize	p-value of difference
Submitted a Project	0.239	0.174	0.445	0.474	0.700	0.159
	(0.064)	(0.057)		(0.114)	(0.105)	
N	46	46		20	20	
Overall Rank	2.339	2.419	0.860	2.537	2.926	0.207
	(0.320)	(0.277)		(0.275)	(0.163)	
Novelty Rank	2.515	2.907	0.405	2.722	3.381	0.095*
	(0.356)	(0.217)		(0.289)	(0.239)	
N	11	8		9	14	

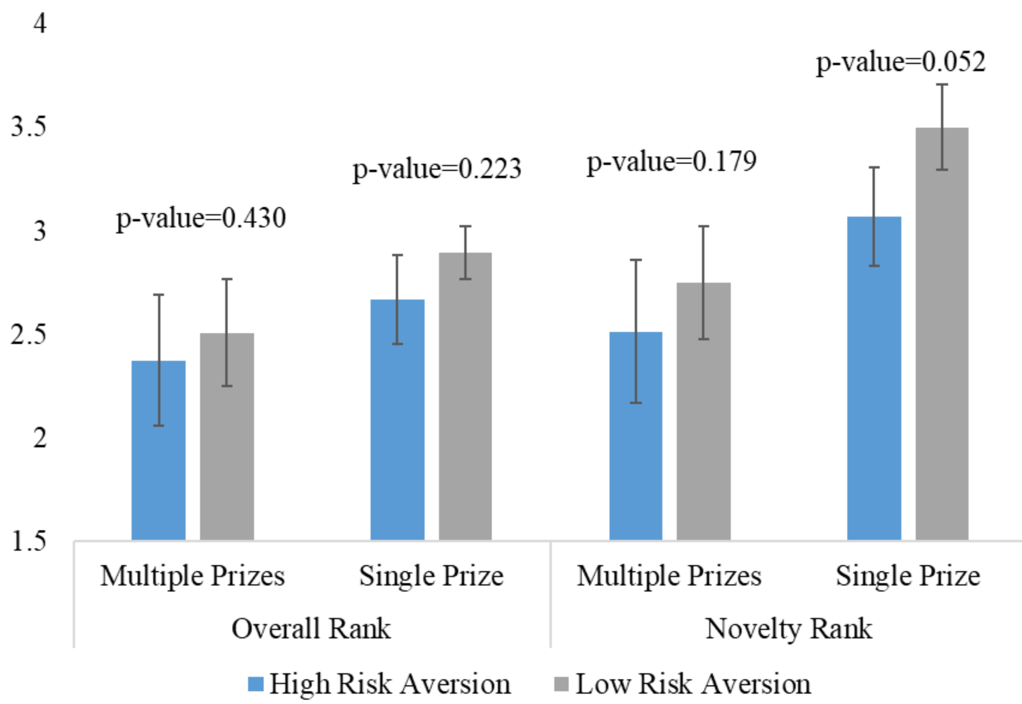
Notes: A submission's Overall Rank is equal to the within-judge rank of the average rating assigned to the five evaluation criteria, averaged across judges who evaluated the submission. A submission's Novelty Rank is the average novelty rating rank across judges who evaluated the submission. Overall Rank and Novelty Rank are conditional on a project being submitted for evaluation by a judge. For both rank measures, a higher rank is associated with a higher quality submission. The statistics reported in the p-value of difference columns are the p-values from tests of equality between the single prize and multiple prize contest arms within each subsample. Standard errors are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Figure 1: The Effect of Prize Structure on Innovative Performance by Risk Preferences

Panel A: The Effects of Prize Structure on Submissions by Risk Preferences



Panel B: The Effects of Prize Structure on Innovation Scores by Risk Preferences



Notes: A submission's Overall Rank is equal to the within-judge rank of the average rating assigned to the five evaluation criteria, averaged across judges who evaluated the submission. A submission's Novelty Rank is the average novelty rating rank across judges who evaluated the submission. Overall Rank and Novelty Rank are conditional on a project being submitted for evaluation by a judge. For both rank measures, a higher rank is associated with a higher quality submission. Our risk elicitation tool produces values that range from 1 (highly risk averse) to 5 (risk neutral or possibly loving). All individuals with scores above 3 (the sample median)are classified as having low levels of risk aversion. The remainder are classified as having high-levels of risk aversion. p-values report the p-value from a t-test of the difference between the reported means. Standard errors are reported as black bars.

Appendix A Additional Tables

Balance Checks for Experimental Internal Validity. To verify that our randomization was successful at assigning statistically identical populations into the single and multiple prize structures, we compare participant mean demographics and characteristics in Table A1. These mean comparisons confirm that there are no statistically significant differences in mean participant observables by treatment group. As an alternative test of randomization success, Table A2 analyzes whether the joint relationship between participant demographics and characteristics and treatment assignment is zero, as suggested by Bruhn and McKenzie (2009). In particular, we regress the variables presented in Table A1 on treatment status and run a test for joint orthogonality. Table A2 demonstrates both that no single participant observable is correlated with treatment status, and that the variables are not jointly related to treatment status (p-value=0.995). Combined, Tables A1 and A2 provide strong evidence that participants were randomly assigned into innovation contest prize structures.

Table A1: Mean Demographics and Characteristics by Treatment Group

	Multiple Prize	One Prize	p-value of difference
Student	0.412	0.500	0.298
	(0.059)	(0.060)	
Employed	0.444	0.366	0.347
	(0.059)	(0.058)	
Female Participant/Group Member	0.212	0.242	0.681
	(0.051)	(0.053)	
Age Range	1.846	1.773	0.607
	(0.109)	(0.091)	
Highest Level of Education	3.697	3.694	0.989
	(0.107)	(0.138)	
Signed Up as Team	0.288	0.303	0.850
	(0.056)	(0.057)	
Any Prior Contest Experience	0.378	0.333	0.589
	(0.060)	(0.058)	
Number of Unique Areas of Relevant Expertise	3.000	2.848	0.672
	(0.259)	(0.245)	
Risk Preferences (Average within Teams)	2.886	2.851	0.875
	(0.153)	(0.165)	
Observations	66	66	

Notes: For team participants, demographics are averaged across teams except for the female variable which is equal to one if any team member is a female. For team characteristics, risk preferences is equal to the average of individual responses, number of areas of relevant expertise is equal to a count of the non-overlapping areas of expertise among team members, and prior contest experience as equal to one if any team member had previously participated in a contest. The statistics reported in the p-value of difference column are the p-values from tests of equality between the single prize and multiple prize contest arms. Standard errors are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table A2: Omnibus Test of Random Assignment Success

	Treatment Assignment
Student	0.063
	(0.167)
Employed	-0.044
	(0.145)
Female Participant/Group Member	0.012
	(0.123)
Age Range	-0.009
	(0.077)
Highest Level of Education	0.019
	(0.051)
Signed Up as Team	0.022
	(0.114)
Any Prior Contest Experience	-0.046
	(0.100)
Number of Unique Areas of Relevant Expertise	-0.007
	(0.025)
Risk Preferences (Average within Teams)	0.002
	(0.037)
Omnibus p-value	0.995
Observations	132
R-squared	0.013
Mean dep var	0.500

Notes: The Table presents regression coefficients from a regression of participant characteristics on prize arm assignments. For team participants, demographics are averaged across teams except for the female variable which is equal to one if any team member is a female. For team characteristics, risk preferences is equal to the average of individual responses, number of areas of relevant expertise is equal to a count of the non-overlapping areas of expertise among team members, and prior contest experience as equal to one if any team member had previously participated in a contest. The Omnibus p-value reports the p-value from testing whether the sum of coefficients reported in the table is equal to zero. Standard errors are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Alternative Measures of Innovation Quality. Our primary innovation quality outcomes are rank-based measures of the overall quality of project submissions across evaluation criteria and of the novelty of project submissions. We favor a rank-based measure over an average score measure because it controls for judge-specific differences in how scores are interpreted in a straightforward way. However, to verify that our estimated effects of prize structure on our two quality measures of innovative output are robust to an average score measure, we report mean differences in average scores normalized by judge-specific means and standard deviations in Table A3. In Table A3, overall score is measured using a standard z-score normalization where we take the judge-specific average score across the 5 evaluation criteria for each project, normalizing this average by the judge's overall score average and standard deviation across all projects she evaluated, and taking the average of the normalized judge-specific scores across the judges who evaluate the project. Novelty score is measured by normalizing the novelty score each judge assigned a project by the judge's novelty score average and standard deviation across all projects she evaluated, and taking the average of the normalized novelty scores across the judges who evaluated the project. The normalization accounts for judge-specific differences in how scores are interpreted. We cannot employ judge fixed effects because each project was randomly assigned to be evaluated by three of the six contest judges, and there are very few instances of the same set of three judges evaluating multiple projects.

Table A3 demonstrates that projects submitted under the single prize structure are scored about 2.5% higher than those in submitted to the multiple prize contest arm (p-value=0.14). These mean comparisons are smaller and less significant than those presented in Table 2, but we believe that is because the judges exhibit an aversion to assigning large score differences across projects.

While we contend that innovation novelty is the most important dimension of innovative quality to test in our setting because of the differences in the rewards to risk-taking under the two prize structures, we also test whether the two prize structures drove quality differences along the other dimensions of quality on which projects were assessed. Table A4 presents the average project rank across the four non-novelty quality dimensions; functionality, user experience, wide scope of use cases, and solves contest problem. These mean comparisons demonstrate that the quality of innovations under the single prize structure is higher across all four dimensions in the single prize structure, but that none of these differences are statistically significant. These patterns further demonstrate that the single prize structure is primarily effective at driving innovators to produce more novel output.

Table A3: Alternative Measures of Overall Quality and Novelty by Prize Structure

	Multiple Prizes	One Prize	p-value of difference
Overall Score	3.572	3.662	0.187
(Judge-specific normalization)	(0.067)	(0.018)	
Novelty Score	0.972	0.996	0.137
(Judge-specific normalization)	(0.012)	(0.010)	

Notes: A submission's Overall Score (Judge-specific normalization) is equal to the within-judge average rating assigned to the five evaluation criteria normalized by the judge's overall rating average and standard deviation, averaged across judges who evaluated the submission. A submission's Novelty Score (Judge-specific normalization) is the novelty rating normalized by each judges' mean novelty rating and standard deviation, averaged across judges who evaluated the submission. Both outcome variables are conditional on a project being submitted for evaluation by a judge. The statistics reported in the p-value of difference column are the p-values from tests of equality between the single prize and multiple prize contest arms. Standard errors are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table A4: Alternative Measures of Innovation Quality by Prize Structure

	Multiple Prizes	One Prize	p-value of difference
Functionality Rank	3.300	3.428	0.646
	(0.231)	(0.160)	
User Experience Rank	2.808	3.015	0.436
	(0.205)	(0.168)	
Wide Scope of Use Cases Rank	2.975	3.212	0.375
	(0.215)	(0.159)	
Solves Contest Problem Rank	2.617	3.011	0.214
	0.247	0.196	

Notes: A submission's Rank for each category is the average category rating rank across judges who evaluated the submission. Outcomes are conditional on a project being submitted for evaluation by a judge. For both rank measures, a higher rank is associated with a higher quality submission. The statistics reported in the p-value of difference column are the p-values from tests of equality between the single prize and multiple prize contest arms. Standard errors are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Post-Contest Survey. Following the innovation contest deadline and before the contest winners were announced, we asked contest participants to complete a survey about their experience in the contest. The complete list of survey questions is provided in Appendix Appendix B. Survey completion was not required to be eligible for contest prizes, and, as a result, only 57 individuals (out of 184) completed the survey. These individuals were significantly more likely to have submitted a project for judgement than those who did not complete the survey. They were also younger, more likely to be students, and less likely to be employed than those who did not complete the survey. However, those who completed the survey were equally likely to be drawn from the single prize contest as from the multiple prize contest, and, conditional on submitting a project for judgment, had equally ranked innovative quality. While we worry that our post-survey sample is too un-representative for us to analyze how prize structure impacts responses to most questions, we think it is informative to examine prize structure preferences across the two contest arms given that our survey sample captures most of those who did submit a project under both project arms (including the contest winners). We compare whether survey respondents prefer a multiple prize structure to a winner-takes-all structure across treatment groups in Table A5, and find that participants in both the single prize and the multiple prize arms are much more likely to state they prefer multiple prizes (over 80% in both cases), and that there is no significant difference in this likelihood by the prize structure to which participants were assigned. This information is particularly interesting given the inconsistency it suggests between participant preferences and innovative performance across prize structures.

Table A5: Post-Contest Survey Responses by Prize Structure

	Multiple Prizes	One Prize	p-value of difference
Prefer Multiple Prizes to One Prize	0.885	0.800	0.339
	(0.058)	(0.067)	
Observations	25	32	

Notes: The sample used in this analysis is restricted to participants who agreed to fill in the post-contest survey. The statistics reported in the p-value of difference column are the p-values from tests of equality between the single prize and multiple prize contest arms. Standard errors are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Robustness of Prize Structure Effects on Performance by Risk Aversion. Despite the lack of difference in mean risk aversion between teams and individuals (see Table 3 in the main text), it is still possible that teams endogenously form around risk preferences in ways that might bias the interpretation of some of our results. To provide further reassurance that differences in how teams and individuals respond to innovation prize structures are not driving are finding that less risk averse participants produce more novel output under the single prize contest arm than they do under the multiple prize arm, we test whether controlling for whether or not a participant is a team or individual changes the relationship between prize structure, risk aversion, and innovative output. In particular, Table A6 presents regression coefficients from an analysis of the effect of prize structure for more and less risk averse participants on outcomes, controlling for an indicator for team participants. These estimates demonstrate that, consistent with Figure 1 in the main text, prize structure does not impact innovation quantity or overall quality for any participants, but that less risk averse participants generate more novel innovations under the single prize structure than the multiple prize structure. We lose some power in this analysis relative to our mean comparisons, and, as such, this relationship is not quite significant at traditional levels of significance (p=0.12). However, the effect size is very similar to our mean comparisons and the results demonstrate the differential effect of prize structure on novelty by risk preferences is robust to controlling for teams.

Table A6: Effects of Prize Structure on Innovative Performance by Risk Aversion, Controlling for Teams

	(1) Submitted	(2) Overall Rank	(3) Novelty Rank
Low Risk Aversion	-0.017	0.028	0.136
	(0.112)	(0.389)	(0.433)
Single Prize	0.020	0.174	0.443
	(0.096)	(0.335)	(0.374)
Low Risk Aversion*	0.012	0.232	0.325
Single Prize	(0.161)	(0.549)	(0.612)
Signed Up as Team	0.386***	0.363	0.345
	(0.084)	(0.269)	(0.300)
Test: Single Prize + Low Risk Aversion*Single Prize=0	0.805	0.353	0.119
Observations	132	42	42
R-squared	0.143	0.092	0.159
Mean dep var	0.318	2.592	2.923

Notes: Overall Rank is equal to the within-judge rank of the average rating assigned to the five evaluation criteria, averaged across judges who evaluated the submission. A submission's Novelty Rank is the average novelty rating rank across judges who evaluated the submission. Overall Rank and Novelty Rank are conditional on a project being submitted for evaluation by a judge. For both rank measures, a higher rank is associated with a higher quality submission. Low risk aversion is equal to one for participants with below median risk aversion. The statistics presented in the Test: Single Prize + Low Risk Aversion*Single Prize=0 row are p-values from a test of equality between the performance of low risk averse participants across prize structures. Standard errors are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Appendix B Data Appendix

Pre-Contest Survey Questions. Each individual who signed up to participate in the contest as a team member or independently was required to complete all pre-contest survey questions. In order to require this completion in compliance with IRB, participants were informed that their survey responses would be linked to their contest performance through their anonymous contest ID and that this data would be used for research undertaken at UC San Diego but not shared with Thermo Fisher or any other organization or individual. Moreover, all participants were informed from the outset that they would not be able to remove their data once they had completed their contest sign-up. This ensured that we had baseline information on all participants even if they never submitted a proposal for evaluation by our panel of judges.

The survey was offered in English and Spanish, and was completed through Google Forms. While we have reproduced the skip codes in our survey in the transcript below, this process was automated in the electronic version of our survey.

Questions 5 and 9 were intended to measure participant confidence levels. We included these questions in order to test heterogeneity in the effects of an information treatment that we had planned to run but were unable to due to sample size constraints. The details of our information treatment can be found in our RCT pre-registry (https://www.socialscienceregistry.org/trials/4026).

1. Contest ID (do not use your legal name, this is to allow us to link your survey responses to your contest performance)
2. Employment Status
   (a) Student
   (b) Full time employee
   (c) Part time employee
   (d) Self-employed
   (e) Unemployed
   (f) Retired
3. Highest level of education
   (a) Less than high school
   (b) High school
   (c) Some college or vocational training
   (d) Bachelor's
   (e) Master's
   (f) PhD
4. Areas of expertise (either through experience or formal education). Check all that apply.
   (a) Desktop software development
   (b) Ecommerce development
   (c) Game development
   (d) Mobile development
   (e) QA & Testing
   (f) Scripts & Utilities

(g) Web Development

(h) Web & Mobile Design

(i) Other – Software Development (Specify)

5. Relative to people with similar expertise as yourself, how would you rank your skill sets on a scale of 0-10 where zero is lower skills than everyone, 10 is better skills than everyone, and 5 is average?

6. Number of contests/hackathons previously participated in (if 0, proceed to question 11; if more than 0, proceed to question 7)

7. Have you ever placed first in an innovation contest/hackathon? If yes, how many times?

8. What is the highest rank you've achieved in prior innovation contests/hackathons you've participated in?

9. Relative to people you have competed against in these contests, how would you rank your skill sets on a scale of 0-10 where zero is lower skills than everyone, 10 is better skills than everyone, and 5 is average?

10. Why have you chosen to sign up to participate in this contest? (check all that apply)

    (a) Prize money

    (b) Develop my skills

    (c) Have fun working on the problem

    (d) Try something new

    (e) Exposure to Thermo Fisher

    (f) Exposure to UC San Diego

    (g) Exposure to judges

    (h) Build my CV

    (i) Other:

11. Choose which of the following gamble you prefer. In all instances, you have a 50% chance of receiving the low payoff, and a 50% chance of receiving the high pay-off. Answer carefully, a random 30% of respondents will receive the pay-off from their selected gamble.

Choice (50/50 Gamble)	Low Payoff (in USD)	High Payoff (in USD)
Gamble 1	16	16
Gamble 2	12	24
Gamble 3	8	32
Gamble 4	4	40
Gamble 5	0	48

12. As in the previous question, choose which of the following gamble you prefer. In all instances, you have a 50% chance of receiving the low payoff, and a 50% chance of receiving the high pay-off.

Choice (50/50 Gamble)	Low Payoff (in USD)	High Payoff (in USD)
Gamble 1	10	10
Gamble 2	6	18
Gamble 3	2	26
Gamble 4	-2	34
Gamble 5	-6	42

13. Gender
    (a) Female
    (b) Male
    (c) Other
    (d) Prefer not to answer
14. Age
    (a) 18-24
    (b) 25-34
    (c) 35-49
    (d) 50-59
    (e) 60-69
    (f) 70+

Post-Contest Survey Questions. Following the contest deadline and before the contest winner(s) were announced, all participants were asked to complete a post-contest survey. Completion of this survey was voluntary, participants were told their responses would be linked to their pre-contest survey responses and contest performance; and that their responses would also be used to better understand how to run effective innovation contests. The survey was offered in English and Spanish, and was completed through Google Forms. While we have reproduced the skip codes in our survey in the transcript below, this process was automated in the electronic version of our survey. Those redirected to section 2 were also asked to complete section 3 upon completing section 2.

In total, 58 individuals completed the survey, 67% of whom submitted a project for evaluation by the judges (compared to 4% among those who did not complete the survey). Thus, survey completion was higher among participants more engaged in the contest.

1. Contest ID:

2. As you know, we decided to split the contest into two separate competitions each with different prize structures.

   As a reminder, you were assigned a prize structure with multiple winners in which first prize will be $6,000, second prize will be $3,000, third prize will be $1,500, 4th prize will be $900, and those who place in the 5th-10th place will receive $600 OR a prize structure with a single prize of $15,000 for the first place submission.

   Participants in the other competition were assigned a prize structure with a single prize of $15,000 for the first place submission OR with multiple winners in which first prize will be $6,000, second prize will be $3,000, third prize will be $1,500, 4th prize will be $900, and those who place in the 5th-10th place will receive $600. Would you have put more, less, or the same amount of effort into the contest if you were assigned to the competition with the other prize structure?

   (a) More effort

   (b) Less effort

   (c) The same amount of effort

3. Please tell us what you think the best prize structure for an innovation contest is from the following list of options.

(a) Single prize for the best submission

(b) Multiple prizes for many of the top submissions

(c) A prize for everyone who submits

(d) Other (please explain)

4. What was the most important factor in determining your participation in this contest?

   (a) The contest prizes

   (b) The networking and exposure opportunities

   (c) Personal challenge

   (d) Other: (Please explain)

5. When did you start working on your hackathon project?

   (a) Friday evening

   (b) Friday night

   (c) Saturday morning

   (d) Saturday afternoon

   (e) Saturday evening

   (f) Saturday night

   (g) Sunday morning

   (h) Sunday afternoon

   (i) Sunday evening

   (j) Sunday night

   (k) I never began working on the project

6. Did you submit a project for consideration by hackathon judges?
   (a) Yes (If yes, go to section 3)
   (b) No (If no, go to section 2)

Section 2

7. Why did you decide not to submit a project for consideration by the judges? Check ALL options that apply:
   (a) I did not have enough time to dedicate to the project due to (check all that apply):
      i. Competing work obligations
      ii. Competing personal obligations
      iii. The project proved more difficult than I had expected.
   (b) My project was incomplete, and as a result, I did not think it was good enough to submit
   (c) I completed my project but did not think it was good enough to submit
   (d) I was worried about the judges thinking poorly of my submission
   (e) I did not think I could win a prize in the contest so decided not to spend time on it
   (f) I lost interest in the contest
   (g) Other: (Please explain)
8. If you had submitted a project, how likely do you think you would have been to win a prize?
   (a) 0%
   (b) 1-15%
   (c) 16-25%
   (d) 26-35%
   (e) 36-50%
   (f) 51-60%
   (g) 61-70%
   (h) 71-80%
   (i) 81-90%

(j) 91-99%

(k) 100%

9. Do you have suggestions for how the hackathon could have been organized differently to convince you to submit a project for consideration by the judges? (open-ended)

Section 3

10. How much time did you spend working on the hackathon problem

    (a) 0 hours

    (b) 1-3 hours

    (c) 4-7 hours

    (d) 8-10 hours

    (e) 11-15 hours

    (f) 16-20 hours

    (g) 21-26 hours

    (h) 27-32 hours

    (i) 32-40 hours

    (j) More than 40 hours

11. Do you think you made the right decision in signing up for the hackathon?

    (a) Yes

    (b) No

12. What do you think your chance of winning a prize in the contest is (if you did not submit a project to the hackathon, choose 0

    (a) 0

    (b) 1-15

    (c) 16-25

    (d) 26-35

(e) 36-50

(f) 51-60

(g) 61-70

(h) 71-80

(i) 81-90

(j) 91-99

(k) 100

13. How would you rate your experience with the hackathon? (Scale from 1-7)

14. Would you consider participating in another hackathon?

    (a) Yes

    (b) No

15. Please provide us with any suggestions for how we could improve the hackathon. (open-ended)