The Effects of Prize Structures on Innovative Performance*

Josh Graff Zivin[†] & Elizabeth Lyons[‡]

June, 2020

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, the winner-takes-all scheme does not reduce innovative output on average, and among teams of innovators, generates more output than the less risky prize structure.

JEL Classification: J24, M54, O32

*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). IRB approval for this study was obtained from the UC San Diego Human Research Protections Program (approval number 180938).

[†]School of Global Policy & Strategy, University of California, San Diego, 9500 Gilman Drive, MC 0519 La Jolla, CA 92093-0519 and NBER. jgraffzivin@ucsd.edu.

[‡]School of Global Policy & Strategy, University of California, San Diego, 9500 Gilman Drive, MC 0519 La Jolla, CA 92093-0519. lizlyons@ucsd.edu.

1 Introduction

The modern firm is relegating more of its routine tasks to machines and orienting employees toward increasingly creative undertakings (Autor et al., 2003). The associated knowledge creation is a fundamental driver of firm prosperity and economic growth (Romer, 1990; Aghion and Howitt, 1992). 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?

In practice, firms deploy a wide range of institutional features to get the creative juices flowing. While some are faddish and others persist, the nature of firms in the real world makes it exceedingly difficult to evaluate their effectiveness. 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. Scotchmer, 2004; Wright, 1983). 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, while existing evidence demonstrates that financial prizes can act as important incentives for innovation (Brunt et al., 2012; Moser and Nicholas, 2013), and argues that innovation prizes may be an effective substitute for patents in some circumstances (Kremer and Williams, 2010), how best to structure these incentives in practice is largely an empirical question (Williams, 2012) and the empirical literature in this space is surprisingly thin.

This paper is designed to help 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 a weekend-long innovation contest in which participants were asked to propose and build a software-based solution to help share medical equipment across small providers in Mexico. The innovation contest was hosted by Thermo Fisher's Mexico office in Baja California and was open to all non-management employees of the firm, as well as to employees at other technology firms, freelancers, and STEM students in the Baja California region.

Contest participants were randomized into two distinct compensation schemes following the contest registration deadline. 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. In both contest arms, the available prizes were large relative to average monthly salaries for software developers in the region and the total prize money was identical across arms. By implementing our experiment in the context of a high profile, firm-sponsored innovation contest that requires participants to develop a complete and novel product, our findings have direct implications for firm-directed innovation systems.

In total, 132 participants[1] signed up to participate in the contest with 66 randomly assigned to each prize structure. In contrast to what existing theory suggests about the impact of prize inequality across competitors on effort (Fang et al., 2020), our results reveal that the incentive structure had no impact on the quantity of innovative output. In both study arms, approximately one-third of participants submitted a solution for evaluation by the judges.

The quality of submissions 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 participants in the winner-takes-all compensation scheme generated

[1] As we discuss below, individual and team participation was permitted. The total number of contest participants combines team and individual participants. The total number of individuals who participated in the contest either on their own or as part of a team is 184.

output with higher overall average scores than their counterparts in the other compensation scheme, this difference was not statistically significant. Importantly, however, participants under the winner-takes-all compensation scheme submitted proposals that were significantly more novel than those in the other compensation scheme. Thus, the risk taking encouraged by the competition with a single prize appears to have driven innovators to pursue more creative solutions.

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 R&D 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 the performance of teams separately from its impact on the performance of individuals and then compare them in a difference-in-difference framework.[3]

We find that teams assigned the winner-takes-all prize structure make approximately 50% more submissions to the contest than those assigned the multiple prize structure. In contrast, the quantity of submissions made by individual participants is unaffected by prize structure. We also find that teams in the winner-takes-all scheme develop significantly more novel solutions than teams in the multiple prize structure. Contrary to expectations that the winner-takes-all prize structure should have larger effects on less risk averse participants

[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.

(Shrader et al., 2019), we find no such evidence across any of our study outcomes.[4]

Our study is closely tied to two distinct empirical literatures on innovation contests. One is focused on historical contests offered by governments and monarchs to examine how these prizes created incentives to innovate (Brunt et al., 2012; Moser and Nicholas, 2013). These studies examine large, economically important contests but, by their nature, are unable to generate the counterfactuals required for studying the impact of differential prize structures on innovation.

The second provides evidence from modern innovation contests with a primary focus on how the size and composition of competitors influences effort and outcomes rather than on the role played by prize structures. Boudreau et al. (2011) study a large number of programming contests on TopCoder, each of which lasts approximately 75 minutes. Gross (2016) takes a similar approach to studying logo competitions. In both cases, financial incentives are modest, and the contests are focused on narrowly defined tasks that are much closer to the input than output end of innovation. In contrast, we study the impact of prize structure on performance in the context of a contest that requires participants to conceive of a complete product in pursuit of substantial financial remuneration. As such, our paper should be viewed as bridge between the historical and modern studies on innovation contests that analyzes how a central dimension of contest design can impact the development of novel commercializable innovations at a scale commensurate with activities that generally take place within the modern firm in the tech sector.

More generally, our study contributes to literature on why and how innovation occurs. Technological breakthroughs transform markets and generate large amounts of producer and consumer surplus (Helpman, 1998; Mansfield et al., 1977; Shane, 2001). Yet, most firms struggle to generate novel innovations and those that manage to succeed often do so at great

[4] As we discuss in section (empirical approach), our subgroup analysis relies on relatively small sample sizes. As a result, statistical significance is indicative of large treatment effect magnitudes, while null results, which might still reflect economically meaningful effect sizes, should be treated with greater skepticism.

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). In contrast to existing empirical findings on the relationship between pay and creativity (Erat and Gneezy, 2016), our finding that a more competitive pay structure is preferable when workers are performing a more risky activity provides empirical support for theoretical evidence on rank-order tournaments and employee pay (Lazear and Rosen, 1981). Moreover, our finding that the winner-takes-all compensation scheme does not reduce output levels on average, and increases them when innovators are working in teams, demonstrates that inducing more radical innovation is less expensive than one would have predicted based on the literature that highlights the discouraging effects of competition on effort (e.g. Cason et al., 2010; Fang et al., 2020; Fullerton and McAfee, 1999; Taylor, 1995).

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

2.1 Innovation Contest

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 USD. 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.[5] 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 through email blasts, posters, and announcements in industry newsletters. Promotion materials included information about the general topic area of the innovation challenge, the competition dates, and the total prize purse available to participants (15,000USD). 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.[6] Participation was open to individuals or teams of up to three people. In order to register for the contest, participants were required to complete a brief survey and agree to the terms and regulations of the contest.[7]

At the start of the competition, the details of the innovation challenge 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. Our contest design was informed by discussions with management at Thermo Fisher and other large companies that use innovation contests as part of their R&D strategy. This effort was undertaken to ensure that it is representative of the types of contests being used throughout industry. For instance, many Fortune 500 companies run short contests both for customers and employees and, as in our setting, these contests lay out a specific problem to be solved (Rathi, 2014). Moreover, many of these contests are run over a short period of time. For instance, the typical hackathons hosted by DevPost, the world's largest hackathon platform, last 1-3 days. These

[5]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.

[6]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.

[7]The contest rules are included in the Data Appendix.

include contests hosted by CapitalOne, IBM, Y-Combinator and many other Fortune 500 companies. Importantly, companies report that these contests generate important innovations. For instance, a 24-hour contest hosted by an insurance company yielded a completely new model for processing healthcare claims (Grijpink et al., 2015). Moreover, our contest is significantly longer and more closely resembles the types of contests companies are hosting to generate new products and services than those studied by much of the existing literature. Crowdsourced innovation, like innovation contests, is forecasted to be worth up to 2 billion USD by 2024 (Markets & Markets, 2019).

Our contest 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 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.*

2.2 Experiment Implementation

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 the standard deviation of expected returns and corresponding participant risk of failure, is higher in the winner-takes-all arm.

Randomization was performed following the enrollment deadline and stratified by team and individual participants to ensure equal numbers of teams and individuals in both contest arms. As Appendix Tables A1 and A2 demonstrate, assignment into contest arms is unrelated to participant demographics and characteristics.

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.[8] 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.

While we felt it was necessary to disclose the two prize regimes to all participants before

[8]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.

the contest, by doing so, we risked upsetting participants who would have preferred the alternative prize structure to the one they were assigned. If this disclosure caused participants to under-invest in the contest, this could bias our estimates. We do not think this occurred in our setting for several reasons. First, when participants signed up to participate, they were explicitly told the prize structure had not been announced to ensure that they did not sign up with a particular expectation in mind. Moreover, both contest arms offer substantial prizes.[9] Second, we had zero participants complain about the prize structure they were placed in. Third, as we discuss in section 4, we find the prize structure the majority of participants report preferring in the post-contest survey led to worse performance than the alternative. Thus, if under-investment due to dissatisfaction with the prize structure is occurring, it is working against our findings rather than biasing our findings upwards.[10]

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 faculty 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 for 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

[9] First place in the multiple prize arm is equal to about 31% of the annual salary of full-time employed computer scientists in the region.

[10] Moreover, as we report in section 4, we do not find differences in the quantity of output or in the likelihood that any effort is put into producing output between the two arms. This further minimizes our concern that the disclosure is biasing our findings.

before being randomly assigned to judges. Judges were blinded to all information about the incentive structure under which proposals were submitted, which allows us to analyze the quality of innovations across contest arms. 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

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.[11] 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.[12]

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

[12]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

The full list of survey questions is 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.[13] 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 both 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, and an aggregate experience measure which would allow teams to have higher maximum experience than individuals. Finally, risk preferences is defined as the average of individual responses.

Panel A of Table 1 presents summary statistics of participant demographics, which were collected from the pre-contest survey. These statistics demonstrate that roughly one-quarter of our 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.

Panel B of Table 1 presents summary statistics of other relevant participant characteristics. These statistics demonstrate that participants have expertise in an average of 3 of

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

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

the 8 programming categories that we had ex-ante identified as relevant for competition success,[14] 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. The types of submissions made to the contest range from very undeveloped at the low end (specifically, a word document describing what a solution could look like), to platforms ready for beta testing on the high end. For example, in both contest arms, the winning submissions included fully developed and functional platforms that addressed the contest problem that included user data tracking capabilities for the firm, and were accompanied by detailed product descriptions and explanations. The majority of submissions include either a basic website or website demo with relatively detailed product descriptions and explanations.

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 because that is the primary focus of most R&D units[15] and the one category where we had an a priori clear hypothesis about the role of our compensation schemes. Novel innovations require

[14]These categories are desktop software development, ecommerce development, game development, mobile development, Q&A Testing, scripts and utilities, web development, web and mobile design. We also included an "other" category, but we do not include skill sets filled in "other" in our measure of areas of expertise because of a lack of consistency in responses, and because most entries were only tangentially related to the contest problem.

[15]For instance, a firm cannot patent an idea unless it is sufficiently novel (Williams, 2017). Moreover, more novel innovations are associated with higher and more persistent returns (Hirshleifer et al., 2018).

more risk taking and increase the likelihood of both big successes and big failures. Given its relatively high standard deviation of expected returns, the winner-takes-all prize structure creates strong incentives for those outlier outcomes and therefore should lead to more novel output.[16]

We generate our overall rank 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. For simplicity of interpretation, a higher ranking indicates a higher quality submission. Although judges evaluated 21 submissions each, our ranking measure ranges from 1-5 because many proposals were assigned the same scores in a given category and disagreement across judges compressed average rankings for each proposal. That our rank measures have the same range as the scoring scales is an unintended coincidence.

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, and allows us to analyze mean comparisons without worrying about whether judge-specific scoring differences are affecting our findings. 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), and if we include judge group fixed effects in a regression framework (see Table A5).

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, with 5 representing the best possible score. 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

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

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."

4 Results

In order to evaluate the impact of prize structure on the quantity and quality of innovativeness, we compare average submission probabilities, and average overall and novelty ranks 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. However, we also run regressions with participant characteristic controls and judge fixed effects to verify that our results are unchanged when we account for these observables.

Once we have established the average treatment effect of prize structure on our main outcome variables, we explore whether treatment effects differ for participant types who may be more or less responsive to risk-reward tradeoffs. The objective of this subgroup analysis is to try to develop a better understanding of the drivers of our main effects. In particular, we separately analyze impacts for teams and solo competitors and participants with different risk preferences.[17]

[17]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. We present an analysis of treatment effects by risk preferences despite its omission from our pre-analysis plan because of the importance of understanding the link between risk taking and prize structure effects in our setting (Duflo et al., 2020).

4.1 Main Results

We begin by discussing our core findings on how prize structure affects the quantity and quality of innovative output. Table 2 presents comparisons of mean outcomes by prize structure in Panel A, and estimated treatment effects of prize structure conditional on participant demographics and characteristics in Panel B.[18] Consistent with randomization being balanced across participant observables, including controls for these observables does not change our treatment effect estimates. We, therefore, focus our discussion on the mean comparisons contained in Panel A.

Interestingly, despite post-survey responses indicating that people in both prize arms prefer the multiple prize structure (see Table A6),[19] the number of participants who make a submission is the same for the single prize and the multiple prize regimes. In both arms, approximately one third of participants submitted their innovation for evaluation by our expert judges. Importantly, participants who did not submit a project may have invested no effort in the contest, or may have invested some effort but believed their project was not sufficiently developed to subject themselves to evaluation. To further probe effort across study arms, we look at the number of participants that registered on the contest DevPost page, not all of whom made a submission for evaluation. We find that the percentage of participants who registered was statistically the same under both study arms and higher than the percentage of participants who made submissions (46% in the multiple prize structure and 52% in the one prize structure). Unsurprisingly, we also find that 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.

[18] The regression coefficients presented in Panel B are conditional on measures of student, employed, female, age range, education, team participants, risk preferences, prior contest experience, and areas of expertise.

[19] To ensure that whether or not participants win a prize does not influence their response to their preferred prize structure, we conducted the post-contest survey before winners were announced. Interestingly, even the winner of the winner-takes-all prize reported preferring a prize structure with multiple prizes.

Despite finding no differences in the quantity of innovative output by prize structure, Table 2 demonstrates there are important differences in the quality of innovations submitted under the two prize schemes. Submissions made under the winner-takes-all prize structure receive an overall rank that is about 15% higher than those submitted under the multiple prize structure. However, as is emphasized by the regression coefficient presented in Panel B, this difference is noisy and insignificant. When we turn our attention to our primary outcome of interest, novelty, we find that submissions made under the winner-takes-all prize structure were significantly better than those made under the multiple prize structure. Specifically, winner-takes-all prize arm submissions are ranked almost 25% higher than multiple prize arm submissions. 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 and thus generate more novel output.

We verify that our treatment effect estimates are robust to including judge-group fixed effects (Table A5), and to using average scores normalized by judge rather than rankings to measure innovation quality (Table A3). In addition, we demonstrate that prize structure did not have a significant impact on any quality dimension other than novelty (Table A4).

4.2 Heterogeneity Analysis

Our findings demonstrate that the winner-takes-all prize structure generates significantly more novel output without altering the total quantity of innovative output generated. We now turn to whether these treatment effects differ across participant characteristics hypothesized to influence innovative activity in order to deepen our understanding of the mechanisms that may be driving our results. Given the significant reduction in sample size for these heterogeneity analyses, our findings should be viewed as exploratory. This is especially important for any analysis of prize structure on outcomes that are conditional on submissions (e.g. novelty), where our analysis is only powered to detect statistical significance for effect sizes larger than 22%. Since effect sizes of 20% are clearly economically meaningful, our null results should be interpreted with caution. On the other hand, any statistically significant relationship is indicative of large effects and is thus particularly informative.

We begin by analyzing whether the relationship between prize structure and innovative performance differs for teams relative to individuals, 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.[20] Given the inherent challenges in observation and verification at the time of team formation, it is perhaps unsurprising that

[20]While our prior was that people would form teams to increase skills and experience in an effort to improve their chance of success, this need not be the case. Social preferences, lower communication costs, and endogenous networks (Gompers et al., 2017; Lyons, 2017; Ruef et al., 2003) could all motivate more homogenous team formation.

teams do not statistically differ from solo competitors in their risk preferences.[21] While this endogenous selection into teams implies that we cannot analyze differences in performance across teams and individuals, the random assignment of teams and individuals to prize structures implies that we can analyze treatment effects within teams and within individuals. We can then then compare these treatment effects to each other, in a framework conceptually akin to a difference-in-difference strategy, to analyze whether prize structure differentially affects teams and individuals.[22]

Table 4 presents these within participant type comparisons. The difference in means between the winner-takes-all and the multiple prize arm for individuals and teams is reported in columns 1 and 2 respectively. Row 3 reports the p-value from a seemingly unrelated estimation test to assess whether these differences in means in individuals and teams are significantly different from one another. This analysis demonstrates that, while individuals have an almost identical submission rate under each prize structure, teams are about 50% more likely to submit a project under the winner-takes-all prize structure than under the multiple prize structure, and that the differences in these treatment effects is statistically significant at the 10% level.[23]

Turning our attention to the quality of output, we find that teams in the winner-takes-all scheme ranked 24% higher on the novelty of their submissions than their counterpart teams in the multiple prize structure. Individuals in the winner-takes-all structure also ranked higher

[21]Consistent with teams being formed to improve upon individual productivity and creativity, the number of submissions is more than twice as high among team participants, and teams also have higher average overall and novelty rankings (p-values=0.12 and 0.14 respectively). These comparisons are presented in Table A7.

[22]Although our sample size is too small for us to compare treatment effects among very similar teams and individuals, we do find that the effect of prize structure on submissions by teams and by individuals is similar for more and less experienced participants and for more and less skilled participants. This suggests that the differences in treatment effects for teams relative to individuals is not be driven exclusively by either differences in experience or differences in skill sets between the two participant types.

[23]As we demonstrate in Table A8, the estimated regression coefficients on the relationship between prize structure and the quantity of innovative output between individuals and teams conditional on participant characteristics are statistically different at the 10% level.

on novelty than those in the multiple prize structure, though this difference is statistically insignificant. Given our small sample sizes and the noisiness of our estimates, we cannot reject that the difference-in-difference treatment effects on either of our quality measures are the same across teams and individuals (see also Table A8)).

Combined, these patterns suggest that winner-takes-all prize schemes generate higher output among teams but not among individuals, but that conditional on making submissions, winner-takes-all leads both participant types to generate more novel innovations.

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. 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 not be interpreted causally.[24]

To study how risk aversion interacts with our prize structures, we compare the quantity

[24] 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, as well as other participant observables (see Table A9).

and quality of output across prize structures by splitting the sample into above and below median sample competitor risk aversion levels.[25] These results are presented in Table 5. Risk preferences do not appear to impact the quantity of output nor the overall quality of that output. We find that those with below sample median risk aversion rank about 25% higher on novelty under the single prize structure relative to the multiple prize structure (p-value=0.05), and that those with above sample median risk aversion in the single prize structure similarly rank about 22% higher on novelty than those in the multiple prize structure (p=0.17).[26] These effect sizes are not statistically different from each other, suggesting that risk preferences do not mediate the relationship between prize structure and novelty in our setting (see also Table A9).

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, innovators 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

[25]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 levels of 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 someone exhibiting relatively low risk aversion or risk neutrality.

[26]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.

for risk-taking, we find that the winner-takes-all prize structure generated significantly more novel output. Surprisingly, the riskiness associated with the winner-takes-all prize structure did not reduce the total quantity of innovation produced.

Our analysis of heterogeneity reveals further nuance. We find that not only does the winner-takes-all compensation scheme not reduce innovative output, it increases and is more novel among team participants suggesting that the payoff from assembling a diverse team to address the scientific 'burden of knowledge' problem (Jones, 2009) is more effectively unleashed under the winner-takes-all regime. Our results on risk aversion were more surprising.

Given that participants face more risk under the winner-takes-all prize structure, we hypothesized that those with lower aversion to risk would perform better under the winner-takes-all prize structure. Consistent with recent evidence from the classroom (Shrader et al., 2019), we expected this to be particularly true for project novelty, as those with less aversion to risk should be better positioned to pursue the less proven problem-solving strategies that lead to more novel solutions. However, our analysis reveals that risk aversion exerts little influence on any of our outcome measures. Whether this null result is a reflection of statistical power, risk measurement, or something more profound about the importance of risk preferences for the types of individuals that self-select into innovation contests is unknown and an important topic for future exploration.

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 and, in fact, increased output among teams, 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. 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 open questions. The implications for contract design beyond the innovation context is also a fruitful area for additional research. The principal agent problem that characterizes many labor contracts requires a careful balancing of the risks borne by employer and employees (Lazear and Rosen, 1981) and how that risk sharing might depend on the tasks performed by workers is under-explored in the empirical literature. Together they comprise a future research agenda.

6 References

Aghion, Philippe and Peter Howitt, "The Schumpeterian approach to technical change and growth," in "Economic Growth in the World Economy: Symposium," Vol. 1993 1992.

Autor, David H, Frank Levy, and Richard J Murnane, "The skill content of recent technological change: An empirical exploration," *The Quarterly journal of economics*, 2003, *118* (4), 1279–1333.

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.

Brunt, Liam, Josh Lerner, and Tom Nicholas, "Inducement prizes and innovation," *The Journal of Industrial Economics*, 2012, *60* (4), 657–696.

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.

Duflo, Esther, Abhijit Banerjee, Amy Finkelstein, Lawrence F Katz, Benjamin A Olken, and Anja Sautmann, "In praise of moderation: Suggestions for the scope and use of pre-analysis plans for rcts in economics," Technical Report, National Bureau of Economic Research 2020.

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.

Erat, Sanjiv and Uri Gneezy, "Incentives for creativity," *Experimental Economics*, 2016, *19* (2), 269–280.

Fang, Dawei, Thomas Noe, and Philipp Strack, "Turning up the heat: The discouraging effect of competition in contests," *Journal of Political Economy*, 2020, *128* (5), 000–000.

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.

Gompers, Paul A, Kevin Huang, and Sophie Q Wang, "Homophily in entrepreneurial team formation," Technical Report, National Bureau of Economic Research 2017.

Grijpink, F, A Lau, and J Vara, "Demystifying the Hackathon.," 2015.

Gross, Daniel P, "Creativity under fire: The effects of competition on creative production," *Review of Economics and Statistics*, 2016, pp. 1–49.

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.

Hirshleifer, David, Po-Hsuan Hsu, and Dongmei Li, "Innovative originality, profitability, and stock returns," *The Review of Financial Studies*, 2018, *31* (7), 2553–2605.

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.

Kremer, Michael and Heidi Williams, "Incentivizing innovation: Adding to the tool kit," *Innovation policy and the economy*, 2010, *10* (1), 1–17.

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

Lazear, Edward P and Sherwin Rosen, "Rank-order tournaments as optimum labor contracts," *Journal of political Economy*, 1981, *89* (5), 841–864.

Lyons, Elizabeth, "Team production in international labor markets: Experimental evidence from the field," *American Economic Journal: Applied Economics*, 2017, *9* (3), 70–104.

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.

Markets & Markets, "Crowdsourced Testing Market by Testing Type (Performance Testing, Functionality Testing, Usability Testing, Localization Testing, and Security Testing), Platform, Organization Size, Vertical, and Region - Global Forecast to 2024," 2019.

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.

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

Moser, Petra and Tom Nicholas, "Prizes, publicity and patents: Non-monetary awards as a mechanism to encourage innovation," *The Journal of Industrial Economics*, 2013, *61* (3), 763–788.

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

Rathi, Anil, "To encourage innovation, make it a competition," *Harvard Business Review*, 2014, *11.*

Romer, Paul M, "Endogenous technological change," *Journal of political Economy*, 1990, *98* (5, Part 2), S71–S102.

Ruef, Martin, Howard E Aldrich, and Nancy M Carter, "The structure of founding teams: Homophily, strong ties, and isolation among US entrepreneurs," *American sociological review*, 2003, pp. 195–222.

Scotchmer, Suzanne, *Innovation and incentives*, MIT press, 2004.

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.

Williams, Heidi, "Innovation inducement prizes: Connecting research to policy," *Journal of Policy Analysis and Management*, 2012, *31* (3), 752–776.

Williams, Heidi L, "How do patents affect research investments?," *Annual review of economics*, 2017, *9*, 441–469.

Wright, Brian D, "The economics of invention incentives: Patents, prizes, and research contracts," *The American Economic Review*, 1983, *73* (4), 691–707.

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

Panel A: Mean Comparisons

	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)	

Panel B: Regression Analyses

	Submitted a Project	Overall Rank	Novelty Rank
Single Prize Contest	0.031	0.274	0.689**
	(0.076)	(0.315)	(0.327)
Observations	132	42	42
R-squared	0.212	0.212	0.294
Mean dep var	0.318	2.592	2.923

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 in Panel A are the p-values from tests of equality between the single prize and multiple prize contest arms. Panel B present regression coefficients from regressions that include controls for participant characteristics. The controls are measures of student, employed, female, age range, education, team participant, risk preferences, prior contest experience, and areas of expertise. 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.

Table 4: Heterogeneous Impacts of Prize Structure by Teams

	(1) **No Team Mean Difference** One Prize - Multiple Prizes	(2) **Team Mean Difference** One Prize - Multiple Prizes	(3) Difference-in-Difference p-value
Submitted a Project	-0.060	0.226	0.103
	(0.084)	(0.157)	
N	93	39	
Overall Rank	0.079	0.389	0.542
	(0.444)	(0.298)	
Novelty Rank	0.391	0.659*	0.619
	(0.458)	(0.377)	
N	19	23	

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 difference in means in the sample of individuals and the sample of teams using seemingly unrelated estimation in Stata. Standard errors are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 5: Heterogeneous Impacts of Prize Structure by Risk Preferences

	(1) **High Risk Aversion** **Mean Difference** One Prize - Multiple Prizes	(2) **Low Risk Aversion** **Mean Difference** One Prize - Multiple Prizes	(3) Difference-in-Difference p-value
Submitted a Project	0.048	0.002	0.769
	(0.102)	(0.139)	
N	85	47	
Overall Rank	0.295	0.387	0.841
	(0.369)	(0.273)	
Novelty Rank	0.558	0.750**	0.711
	(0.405)	(0.322)	
N	27	15	

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 difference in means in the sample of participants with at least the sample median risk aversion and the sample participants with below sample median risk aversion using seemingly unrelated estimation in Stata. Standard errors are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

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%

Robustness to Judge Fixed Effects. Our main analysis does not include judge fixed effects because our ranking measure accounts for judge specific differences which allows us to analyze mean comparisons. In addition, judge fixed effects are challenging in our setting because of how submissions were assigned to judges. We had too many submissions for each judge to review all of them,[27] and we wanted each submission to be reviewed by at least three judges. While we could have assigned the same set of submissions to three-judge groupings, which would have allowed us to include judge-group fixed effects, we were concerned that this would still prevent us from disentangling judge-specific evaluation norms from innovation quality given that we would have had only two judge groupings. We decided that, in order to ensure the scores were most informative, it would be best to randomize submissions within each judge such that no two judges reviewed the same set of submissions. As a result, few submissions have the same judge grouping as each other. In total, we have 20 judge-groupings with as few as one participant per group in 7 cases (and an average of 2.7 participants per group). We are confident that by combining both the average ranking across three judges (such that no single judge determines the ranking of a submission), and by using rankings

[27] Because our judge pool was made up of busy company executives and computer science faculty, we were limited in how much time we could demand from them.

which eliminate any judge-specific grading preferences (e.g. some judges prefer assigning higher scores than others), our analysis does sufficiently control for judge effects.

However, to verify that our ranking measures do sufficiently account for potentially important differences across the scoring behavior of judges, we re-ran our main quality analyses with judge fixed effects in a regression framework to test whether including them overturns our findings. We include separate fixed effects for each of the three-judge groupings in our data. These results are presented in Table A5, and demonstrate that the inclusion of these fixed effects does not affect our findings. Table A5 also includes controls for participant observables to be consistent with the results presented in Panel B of Table 2 (our main results table).

Table A5: Effects of Prize Structure on Innovation Quality, Controlling for Judge Fixed Effects & Participant Characteristics

	(1) Overall Rank	(2) Novelty Rank
Single Prize Contest	0.335	0.890*
	(0.371)	(0.349)
Observations	42	42
R-squared	0.796	0.830
Mean dep var	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. All regressions include fixed effects for each judge, and controls for participant characteristics (student, employed, female, age, education, team membership,. prior contest experience, skills, and risk preferences). 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 A6, 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 A6: 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.

Mean Outcomes by Team and Individual Participants. As we discuss in section 4.2, 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. Consistent with this, Table 3 in the main text demonstrates that teams in our setting were constructed to improve on individual capabilities. In Table A7 below, we show that teams also perform better than individual participants on average. In particular, they are more than twice as likely to submit a project for consideration by the judges (p=0.00), and rank about 15% higher overall (p=0.12) and about 17% higher on novelty (p=0.14).

Table A7: Outcomes by Team and Individual Participants

	(1) Participate as Individual	(2) Participate as Team	(3) p-value of difference
Submitted a Project	0.204	0.590	0.000***
	(0.042)	(0.080)	
Overall Rank (1-5)	2.373	2.774	0.121
	(0.213)	(0.148)	
Novelty Rank (1-5)	2.680	3.123	0.140
	(0.225)	(0.192)	
N	93	39	

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 individual and team characteristic means. Standard errors are in parentheses.

Robustness of Prize Structure Effects in Heterogeneity Analysis. The analysis of heterogeneous effects of prize structure by teams and individuals and by participant risk preferences presented in Tables 4 and 5 of the main paper relies on mean comparisons across prize structures by these subgroups. We verify that this analysis is unchanged by conditioning on participant observable characteristics in Tables A8 and A9 below. Regression coefficients on prize structure in these tables are conditional on measures of whether participants are students, employed, or female; participant age range, education, prior contest experience, and areas of expertise. Table A8, which analyzes the effects of prize structure on outcomes separately for teams and individuals, also includes a control for participant risk preferences. Table A9, which analyzes the effects of prize structure on outcomes separately for more and less risk averse participants, controls for whether participants are in teams or not.

Verifying that controlling for team participants in our risk preference analysis does not change our findings is particularly important because, 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. The estimates presented in Table A9 demonstrate that, consistent with Table 5 in the main text, prize structure does not impact innovation quantity or overall quality for any participants, and that more and less risk averse participants produce similarly more novel output under the single prize structure. We lose some power in this analysis relative to our mean comparisons, and, as such, the relationship between prize structure and novelty 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.

Table A8: Effects of Prize Structure on Innovative Performance by Teams and Individuals, Controlling for Participant Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)
	Submitted		Overall Rank		Novelty Rank	
	No Team	Team	No Team	Team	No Team	Team
Single Prize	-0.074	0.214	0.010	-0.051	0.289	0.703
	(0.088)	(0.175)	(0.685)	(0.417)	(0.630)	(0.527)
Test: No Team=Team	0.095*		0.916		0.486	
Observations	93	39	19	23	19	23
R-squared	0.082	0.331	0.162	0.461	0.212	0.472
Mean dep var	0.204	0.590	2.373	2.774	2.680	3.123

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. The statistics presented in the Test: No Team=Team row are p-values from a test of equality between the coefficient estimates on prize structure among individual and team participants. All regressions include controls for measures of whether participants are students, employed, or female; participant age range, education, prior contest experience, risk aversion and areas of expertise. Standard errors are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table A9: Effects of Prize Structure on Innovative Performance by Risk Aversion, Controlling for Participant Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)
	Submitted		Overall Rank		Novelty Rank	
	No Team	Team	No Team	Team	No Team	Team
Single Prize Single Prize Contest	0.026	0.075	0.122	0.376	0.541	1.000
	(0.102)	(0.132)	(0.516)	(0.529)	(0.473)	(0.501)
Test: High Risk Aversion=Low Risk Aversion	0.748		0.623		0.339	
Observations	85	47	27	15	27	15
R-squared	0.205	0.311	0.277	0.345	0.305	0.642
Mean dep var	0.318	0.319	2.539	2.689	2.824	3.100

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: High Risk Aversion=Low Risk Aversion row are p-values from a test of equality between the coefficient estimates on prize structure among low and high risk averse participants. All regressions include controls for measures of whether participants are students, employed, or female; participant age range, education, team status, prior contest experience, and areas of expertise. Standard errors are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Appendix B Data Appendix

Contest Promotion Text. The following text was used in flyers and emails used to promote the contest. Flyers and emails contained the text in Spanish and in English. The flyer text was also included on the contest webpage through which all participants were required to sign-up.

Promotion Email Text

Dear XYZ,

We are excited to announce a research study being conducted by Professors at UCSD that involves a weekend long Digital Baja Online Hackathon open to all residents of Baja California over the age of 18.

With 15,000 USD in prizes to be won, the contest offers participants the opportunity to propose and work towards the development of a technology that will improve healthcare in Mexico!

Thermo Fisher is assisting in the organization and judging of the contest. While we think this is a great opportunity to use your skills for a fun and important project, participation is completely voluntary and unrelated to your work if you are a Thermo Fisher employee.

Contest details, including sign-up instructions are below.

Please share this announcement with anyone you think might be interested!

Sign-Up Deadline: April 24, 2019

Contest Start Time: April 26 at 6 pm

Submission due by: April 28 at midnight

Location: Wherever you are most creative!

Sign-Up and Additional Contest Information:

Promotion Flyer Text

Digital Baja Online Hackathon!

Weekend Long Hackathon as part of a research study to better understand what motivates participation in these types of activities.

With $15,000 in prizes to be won, this contest offers participants the opportunity to propose and work towards the development of a technology that will improve healthcare in Mexico!

For more information and to sign-up to participate, go to: Contest URL

Contest Rules. All individuals who signed up to participate in the contest as a team member or independently were emailed the following contest rules in Spanish and English at the start time of the innovation contest (6 pm on a Friday). The contest rules were also posted on the DevPost host pages, and on the contest webpage through which all participants were required to sign-up.

Who is eligible to participate?

Any resident of Baja California over the age of 18 is eligible to participate in the contest.

Who is eligible to win the contest?

You must sign up for the contest by filling in the participant survey and confirming your eligibility and willingness to participate by April 24 at midnight. To complete the sign-up process, visit: (*contest webpage*). Everyone who signs-up for the contest by the sign-up deadline must submit a project in compliance with the rules outlined herein to be in consideration for prize money.

Where will participants work on their projects?

This is a fully digital contest, so all work will be done remotely and submissions will be made online through a link provided to participants after the sign-up deadline. Participants can work on their solutions wherever they are most productive. Submissions require access to the internet.

Project requirements:

- Project ideas must be the original work of the submitters
- If participants' solutions uses someone else's technology, participants must make reference to this and, in the case of proprietary technology must receive permission from the owner to use it.
- Projects that violate the copyright, trademark rights, patent rights, or any other intellectual property rights of a third party will be disqualified.

- Projects that are deemed to include offensive, inappropriate, or illegal content will be disqualified
- Projects can be submitted by an individual or teams of 2 to 3

How will contest winners be selected?

Prizes will be awarded to projects with the highest scores from our panel of expert judges based on the following equally weighted factors: functionality, user friendliness, wide scope of use cases, novelty, addresses contest problem. Winners will be informed of their prize by email.

Contest dates and times:

April 24, 2019	Midnight	Deadline to sign-up for contest
April 26, 2019	6:00 PM	Contest problem is announced, contest begins
April 28, 2019	Midnight	Deadline to submit a project for consideration by judges
May 22, 2019		Contest winners will be informed of their winning

What can participants do with my project after the contest is over?

It's totally up to them! Participants retain all intellectual property rights over their submissions. The contest judges have signed an agreement to respect these rights.

What if something unexpected happens?

If something unforeseen occurs that affects the integrity of the contest, either during the contest sign-up period, the contest, or while judges are evaluating projects, we reserve the right to make changes to the contest timeline and submission requirements.

Contest Instructions. All individuals who signed up to participate in the contest as a team member or independently were emailed the following instructions in Spanish and English at the start time of the innovation contest (6 pm on a Friday). Instructions were also posted on the contest DevPost host pages. Participants in the winner-takes-all contest arm received the italicized prize information, and those in the multiple prize contest arm received the prize information in bold text. Otherwise, the information provided to participants in each arm was identical.

Dear Digital Baja Hackathon Participant,

Thank you for your interest in participating in the first ever Digital Baja Hackathon! We are thrilled to have participants like you working on a real world and important problem. Below we have provided you with details on the specific contest problem we would like you to work on, the contest prize structure, and instructions on how to submit your project.

Contest Problem: Mexico has many small health care providers and research and clinical laboratories that, on their 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.

Contest Prize Structure: As we could not agree on what the best prize structure for hackathons is, we have divided participants into two separate groups with two separate prize structures, each totaling $15,000! Importantly, submissions within each prize structure will only be judged relative to others in the same prize structure and this means you will only be competing with half of the contest participant pool! *You have been randomly assigned into a winner takes all prize structure with a single first place prize of $15,000!* **You have been randomly assigned into 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**

place will receive $900, and those who place in the 5th-10th place will receive $600!

Submission Instructions: Please submit any computer script through the following DevPost link: . Please submit any written explanations or any other non-script output to . Please submit any and all output, even if it is incomplete.

Remember that the deadline for contest submissions is Sunday, April 28 at 11:59 pm! We look forward to seeing what you come up with!

Sincerely,

Your Digital Baja Hackathon Organizers!

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)