AEA Papers and Proceedings 2021, 111: 577–581
https://doi.org/10.1257/pandp.20211119

ECONOMICS OF INNOVATION[‡]

The Effects of Prize Structures on Innovative Performance[†]

By Joshua Graff Zivin and Elizabeth Lyons*

The modern firm is relegating more of its routine tasks to machines and orienting employees toward increasingly creative undertakings (Autor, Levy, and Murnane 2003). The associated knowledge creation is a fundamental driver of firm prosperity and economic growth (Romer 1990). Given the circuitous path between effort and outcome, what is the best way to encourage innovation? Do winner-takes-all 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? While existing evidence demonstrates that financial prizes can act as important incentives for innovation (Brunt, Lerner, and Nicholas 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, 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 run within a software innovation contest hosted by the University of California, San Diego, and a major life sciences company. Contest participants were randomized into two compensation schemes: a winner-takes-all arm and a multiple-prize arm. By implementing our experiment in the context of a high-profile firm-sponsored innovation contest that required participants to develop a complete and novel product, our findings have direct implications for firm-directed innovation systems.

The findings from our study have potentially far-reaching implications for the design of institutions and incentives to foster more novel innovation. First, we demonstrate that providing sizeable rewards for only the very top performers appears to inspire the sort of risk taking required to explore new, unproven approaches rather than exploiting well-known ones for more incremental progress (March 1991). Second, 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 riskier activity provides empirical support for theoretical evidence on rank-order tournaments and employee pay (Lazear and Rosen 1981). Moreover, we find that the winner-takes-all compensation scheme does not reduce output levels on average and increases them when innovators are working in teams,[1] demonstrating that inducing more

[‡]*Discussants:* John C. Haltiwanger, University of Maryland; Daniel P. Gross, Duke University; Kevin Stange, University of Michigan; Walker Hanlon, New York University-Stern.

*Graff Zivin: School of Global Policy & Strategy, University of California, San Diego, and NBER (email: jgraffzivin@ucsd.edu); Lyons: School of Global Policy and Strategy, University of California, San Diego (email: lizlyons@ucsd.edu). We are grateful to Iwan Barankay; Gordon Hanson; Karim Lakhani; Craig McIntosh; participants at the Duke Workshop on Field Experiments in Strategy, Entrepreneurship, & Innovation; and participants at the NBER Summer Institute Productivity/Innovation 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).

[†]Go to https://doi.org/10.1257/pandp.20211119 to visit the article page for additional materials and author disclosure statement(s).

[1]To be consistent with the research and development (R&D) workplace, participants could elect to compete as an individual or in a team of their choosing.

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., Fang, Noe, and Strack 2020).

I. Experimental Design and Data

We use data from a randomized control trial within an innovation contest that we hosted in partnership with Thermo Fisher Scientific's Mexico office in Baja California.[2] The contest was open to all nonmanagement employees of the firm, employees at other technology firms in the region, and STEM students at local universities.[3] A total of 184 individuals signed up for the contest, 91 of which signed up to participate in a total of 39 teams and 93 of which signed up to participate on their own.

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. 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.[4]

Consistent with Thermo Fisher's technological focus, participants were asked to develop a software solution that would allow small healthcare providers and health-science labs to access high-quality medical equipment in Mexico. Contest submissions were judged by six industry experts on a five-point scale 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 addressed the innovation challenge.

We randomly assigned participants to one of two prize menus, both with a total of $15,000 available to contest winners. The first prize structure was a winner-takes-all design in which the full $15,000 was given to the highest ranked submission. The second prize structure provided awards to the ten highest ranked submissions.[5] 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 the 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.[6] Participants were given information about the prize structure that they would face at the same time that they were provided details on the innovation challenge.[7]

Our outcomes of interest are the quantity and quality of innovative output. Our measure of quantity is an indicator for whether or not participants submitted a proposal for evaluation by the judges.[8] Our primary measures of the quality of innovative output are (i) overall project

Importantly, prize-structure randomization was performed following selection into teams.

[2] Thermo Fisher is a large biotechnology company. Thermo Fisher has an R&D office in Baja California, and is working with local stakeholders to develop the region's STEM labor force.

[3] For complete details on the contest organization and promotion, please see our online data Appendix.

[4] For instance, many Fortune 500 companies run short contests for both 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 one to three days. 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 (e.g., Boudreau, Lacetera, and Lakhani 2011; Gross 2020).

[5] 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.

[6] Balance tests demonstrate that assignment into contest arms is unrelated to participant demographics and characteristics (see Tables A1 and A2 of the online Appendix).

[7] To avoid concerns about unobservable information spillover between participants in different prize arms, we disclosed the design up front. Participants were also assured that they would be judged only relative to others facing the same prize structure. This could bias our estimates if it caused participants to underinvest in the contest. We do not think this occurred in our setting. 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, and both contest arms offer substantial prizes. Second, we had zero participants complain about the prize structure in which they were placed. Third, we find the prize structure the majority of participants report preferring in the post-contest survey led to worse performance than the alternative.

[8] The types of submissions made to the contest ranged from a word document describing what a solution could look like at the low end to platforms ready for beta testing on the high end. The majority of submissions included either a basic website or website demonstration with relatively detailed product descriptions and explanations.

rank, which places equal weight on all five evaluation categories, and (ii) the project novelty rank, which captures the primary focus of most R&D units and is the one category for which we had an a priori clear hypothesis about the role of our compensation schemes. Novelty is evaluated relative to what is currently on the market, with the lowest possible score being given for "proposed solutions already available in the target market" and the highest possible score being given for "proposed solutions that are different than anything currently available in the target market and that are so creative judges are almost sure no one else has thought of a similar idea."[9]

II. Results

Table 1 presents comparisons of mean outcomes by prize structure.[10] Despite post-contest survey responses indicating that people in both prize arms prefer the multiple-prize structure,[11] the number of participants who make a submission is the same for the single-prize and multiple-prize regimes.[12]

While we find no differences in the quantity of innovative output by prize structure, we find that submissions made under the winner-takes-all prize structure received overall ranks that are about 15 percent higher than those submitted under the multiple-prize structure, though 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 ranked almost 25 percent higher than multiple-prize arm submissions and that this difference is statistically significant.[13]

We analyze whether these treatment effects differ for teams relative to individuals 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).[14] Consistent with teams being selected to improve upon individual capabilities, teams span a broader set of skills and encompass more experience than their individual counterparts.[15] While this endogenous selection into

Table 1—Innovation Outcomes by Prize Structure

	Multiple prizes	One prize	*p*-value of difference
Submitted a project	0.303	0.333	0.711
	(0.057)	(0.058)	
Overall rank	2.428	2.742	0.227
	(0.211)	(0.150)	
Novelty rank	2.608	3.208	0.042
	(0.230)	(0.175)	

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

[9] 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 and if we include judge group fixed effects in a regression framework.

[10] The treatment effect estimates from mean comparisons are equivalent to regression coefficients from regressions that control for measures of student, employed, female, age range, education, team participants, risk preferences, prior contest experience, and areas of expertise.

[11] To ensure that whether or not participants won a prize did 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.

[12] We also compare the number of participants that registered on the contest submission page as an alternative measure of effort and find that the percentage of participants who registered was statistically the same under both study arms. These findings suggest that, at least at the extensive margin, effort was the same in both prize structures.

[13] We do not find that participants in the winner-takes-all arm outperform participants in the multiple-prize arm on any of the other four evaluation categories.

[14] Given the significant reduction in sample size for these heterogeneity analyses, our findings should be viewed as exploratory. For instance, our analysis of prize structure on outcomes that are conditional on submissions (e.g., novelty) is only powered to detect statistical significance for effect sizes larger than 22 percent. Since effect sizes of 20 percent are quite large, our null results should be interpreted with caution.

[15] While it is not surprising that individuals formed teams in an effort to improve upon their own experience and skills, this need not be the case. Social preferences, lower

Table 2—Treatment Effects of Prize Structure on Teams versus Individuals

	Individual mean difference *one prize − multiple prizes* (1)	Team mean difference *one prize − multiple prizes* (2)	Difference-in-difference *p*-value (3)
Submitted a project	−0.060	0.226	0.103
	(0.084)	(0.157)	
Observations	93	39	
Overall rank	0.079	0.389	0.542
	(0.444)	(0.299)	
Novelty rank	0.391	0.659	0.619
	(0.458)	(0.377)	
Observations	19	23	

Notes: Outcome measures are as described in Table 1. The statistics reported in the "difference-in-difference *p*-value" 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.

teams implies that we cannot analyze differences in performance across teams and individuals, because of the random assignment of teams and individuals to prize structures, we can analyze treatment effects within teams and within individuals and compare these treatment effects to each other.[16]

Table 2 presents these within-participant-type comparisons and demonstrates that while individuals have an almost-identical submission rate under each prize structure, teams are about 50 percent more likely to submit a project under the winner-takes-all prize structure than under the multiple-prize structure.[17] Additionally, we find that teams in the winner-takes-all scheme ranked 24 percent higher than their multiple-prize-structure-team counterparts on the novelty of their submissions. Individuals in the winner-takes-all structure also ranked higher 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. Combined, our findings on the effects of prize structure among teams suggest 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.

III. Conclusion

Our results have potentially far-reaching implications for the design of institutions and incentives to foster radical innovation. Providing sizeable 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.[18] 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 genius is not created by incentives but empowered by them. That teams

communication costs, and endogenous networks could all motivate more homogeneous team formation.

[16] Our sample size prevents us from comparing treatment effects among very similar teams and individuals. However, the effect of prize structure on submissions by teams and by individuals looks similar across the skill and experience distribution.

[17] We find that the 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 percent level.

[18] Interestingly, given the relationship between risk and novelty, we do not find evidence that more risk-loving participants perform better on the winner-takes-all scheme than less risk-loving participants.

are better able to respond to those incentives is consistent with broader trends in science (Wuchty, Jones, and Uzzi 2007), but much more work is required to understand the raw ingredients that shape the relationship between creativity and compensation schemes. 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 underexplored in the empirical literature. Together they comprise a future research agenda.

REFERENCES

Autor, David H., Frank Levy, and Richard J. Murnane. 2003. "The Skill Content of Recent Technological Change: An Empirical Exploration." *Quarterly Journal of Economics* 118 (4): 1279–1333.

Boudreau, Kevin J., Nicola Lacetera, and Karim R. Lakhani. 2011. "Incentives and Problem Uncertainty in Innovation Contests: An Empirical Analysis." *Management Science* 57 (5): 843–63.

Brunt, Liam, Josh Lerner, and Tom Nicholas. 2012. "Inducement Prizes and Innovation." *Journal of Industrial Economics* 60 (4): 657–96.

Charness, Gary, and Matthias Sutter. 2012. "Groups Make Better Self-Interested Decisions." *Journal of Economic Perspectives* 26 (3): 157–76.

Erat, Sanjiv, and Uri Gneezy. 2016. "Incentives for Creativity." *Experimental Economics* 19 (2): 269–80.

Fang, Dawei, Thomas Noe, and Philipp Strack. 2020. "Turning Up the Heat: The Discouraging Effect of Competition in Contests." *Journal of Political Economy* 128 (5): 1940–75.

Gross, Daniel P. 2020. "Creativity under Fire: The Effects of Competition on Creative Production." *Review of Economics and Statistics* 102 (3): 583–99.

Jones, Benjamin F. 2009. "The Burden of Knowledge and the "Death of the Renaissance Man": Is Innovation Getting Harder?" *Review of Economic Studies* 76 (1): 283–317.

Kremer, Michael, and Heidi Williams. 2010. "Incentivizing Innovation: Adding to the Tool Kit." *Innovation Policy and the Economy* 10 (1): 1–17.

Lazear, Edward P., and Sherwin Rosen. 1981. "Rank-Order Tournaments as Optimum Labor Contracts." *Journal of Political Economy* 89 (5): 841–64.

March, James G. 1991. "Exploration and Exploitation in Organizational Learning." *Organization Science* 2 (1): 71–87.

Moser, Petra, and Tom Nicholas. 2013. "Prizes, Publicity and Patents: Non-monetary Awards as a Mechanism to Encourage Innovation." *Journal of Industrial Economics* 61 (3): 763–88.

Rathi, Anil. 2014. "To Encourage Innovation, Make It a Competition." *Harvard Business Review*, November 19. https://hbr.org/2014/11/to-encourage-innovation-make-it-a-competition.

Romer, Paul M. 1990. "Endogenous Technological Change." *Journal of Political Economy* 98 (5, Part 2): S71–S102.

Wuchty, Stefan, Benjamin F. Jones, and Brian Uzzi. 2007. "The Increasing Dominance of Teams in Production of Knowledge." *Science* 316 (5827): 1036–39.