Incentives for Public Goods Inside Organizations: Field Experimental Evidence[1]

Andrea Blasco   Olivia S. Jung   Karim R. Lakhani   Michael Menietti

Last updated: 10 April, 2017

Abstract

We report results of a natural field experiment conducted at a medical organization that sought contribution of public goods (i.e., projects for organizational improvement) from its 1200 employees. Offering a prize for winning submissions boosted participation by 85 percent without affecting the quality of the submissions. The effect was consistent across gender and job type. We posit that the allure of a prize, in combination with mission-oriented preferences, drove participation. Using a simple model, we estimate that these preferences explain about a third of the magnitude of the effect. We also find that these results were sensitive to the solicited person's gender.

JEL Classification: D23; H41; M52.

Keywords: innovation contest; free rider problem; social preferences; altruism; idea generation; organization of work.

1 Introduction

Employees divide their time between production tasks tied to pay and promotion and organizational improvements without explicit compensation. This dilemma leads to two key issues in the analysis of incentives inside organizations. First, compensation schemes need to address a multitasking problem (Holmstrom and Milgrom, 1991; Hellmann and Thiele, 2011; Manso, 2011) as time and effort are more likely to be directed towards the rewarded tasks. Second, the non-rival nature of improvement work that benefits the organization as a whole may tempt workers to free-ride on the efforts of others.

An organizational contest that seeks contribution of public goods (e.g., suggestion of organizational improvement opportunities) and offers a winning prize may be a proper way to remedy the incentives. Contests stimulate risk taking and participation in less-easily contractible tasks, as discussed by Lazear and Rosen (1981); Green and Stokey (1983); Mary et al. (1984) among others. However, traditional contest literature does not take into account the non-rival nature of contributing which constitutes the public good problem. Standard contest theory models presumes that only the contest sponsors (e.g., managers) enjoy the contest outcome, with small gains for the participants beyond the prizes given to the winners.[1] This assumption may become questionable when contest participants can expect gains from the work of others, as in the case of a competition aiming at organizational improvements. These circumstances raise an important question: Are contests an effective incentive mechanism in fostering contributions to public goods inside organizations?

In the present study, we report results of a natural field experiment that considers two dominant perspectives on the issue. The first comes from the seminal work of Morgan (2000), who stresses that awarding prizes in lottery-like contests will mitigate, or even eliminate, the incentive to free ride. At the margin, employees choose their level of effort so that expected gains and costs are equal. Because these gains include direct utility from receiving a prize and from improving their organization, employees' participation is likely to be boosted beyond the pecuniary value of the prize. In the second perspective, employees will volunteer their effort when the contest aim matches the preferences of workers, as discussed by Besley and Ghatak (2005), Prendergast (2007), Delfgaauw and Dur (2008) among others.[2] According to this perspective, announcing a competition for prizes can cause unintentional adverse consequences, such as evoking mixed behaviors in workers with varying inclinations towards competition or risk (e.g., men and women

---

[1]One notable exception is the model of tournaments with positive externalities introduced by Drago and Turnbull (1988).

[2]Everyday evidence for this kind of "altruistic" behavior comes from blood donations, charitable giving, social workers. See also the results of laboratory experiments based on economic games (see Levitt and List, 2007) and field studies showing evidence on pro-social preferences at work (Bandiera et al., 2005; Della Vigna et al., 2016).

Niederle and Vesterlund, 2007; Croson and Gneezy, 2009); promoting unethical behavior (Lazear, 1989; Charness et al., 2013); weakening intrinsic motivations (Reeve and Deci, 1996; Frey, 1997); and even hurting the creativity of innovative work (Erat and Gneezy, 2015).

We conducted this study in collaboration with the Massachusetts General Hospital's (MGH) Corrigan Minehan Heart Center ("Heart Center"), a prominent medical organization in the United States and a teaching hospital of the Harvard Medical School. The health care delivery context is particularly relevant as the need for organizational improvement and innovation is vastly noted (e.g., Cutler et al., 2012). In addition, health care professionals are commonly seen as willing to step beyond the boundaries of their contractual duties to offer better care (Delfgaauw, 2005), which makes the comparison of different incentives towards a public good especially relevant and interesting.

The Heart Center launched a contest aimed to improve the operations of the organization, in the spirit of "open innovation" discussed in Terwiesch and Xu (2008), Lakhani et al. (2013), and Glaeser et al. (2016). The contest solicited employees to submit project proposals describing an existing problem and providing a solution to address the problem. After the submission phase, the contest invited all employees to read and rate each proposal on a five-point scale. The winning proposal would receive funding for implementation, implying additional costs and responsibilities from making a winning proposal, such as providing further guidance or a direct involvement in implementation, that were not compensated by winning the competition. Winners of the contest received an iPad mini, the value of which was relatively small in comparison to the foreseeable costs of unpaid proposal implementation time and effort.

The subject pool consisted of over 1,200 staff members of the Heart Center including physicians, nurses, and administrative staff (the entire population). Our main intervention involved altering the content of solicitation: in the PRIZE treatment, the solicitation nudged employees to participate to win an iPad mini; in the FUND treatment, the solicitation nudged employees to participate to win $20,000 towards implementation of the proposal (the submitters would not receive this money for personal use); in the remaining two treatments, we emphasized the opportunity to improve the health care of the patients (PCARE) or the workplace (WPLACE). We randomly assigned each Heart Center staff to receive one of the four solicitations.

By doing so, we obtained causal estimates of the effect of different incentives on two main outcomes: (a) the decision to submit a proposal and engage in an organizational improvement task and (b) the quality of the submissions measured by over 12,000 peer ratings and about 100 evaluations made by the contest organizers. Using data on profession and gender, we also characterized the heterogeneity of responses to the treatment, testing the implications of our intervention for the organization.

We find a high participation rate despite its cost: 196 employees (16 percent) across the entire

organization participated in the initiative, with 5 percent of our sample submitting a proposal.[3] About half of the participants were invited to submit detailed implementation plans. Two of these submissions received funding for implementation.

We observe that small prizes boosted participation without lowering the quality of the submissions. Mentioning the prize in the solicitation email produced a 2.5 percentage points (85 percent) increase in participation. This effect appears too large for a contest without public good incentives. The relatively high incomes of our subject pool, the low chance of winning, and the anticipated costs of being selected for implementation suggest that very few would find it advantageous to participate. We discuss several possible explanations for this observation and conclude that offering a prize increased participation because it reduced the incentive to free ride, as in Morgan (2000). We then use a simple model to estimate employees' underlying preferences towards improving the organization (i.e., the public good), showing that these preferences can account for 25 percent of individual costs of participation.

Analysis of the peer ratings indicates that there was no crowding-out effect. The higher propensity of participation for employees in the PRIZE treatment does not seem to be driven by low-quality submissions. This result is consistent with ratings by the Heart Center managers. We also find small differences in content in terms of the number of proposals submitted by a participant, areas of focus, and proposal length across treatment groups. Overall, these findings suggest no trade-off between quantity and quality; treatments that attracted more participation resulted in proposals of comparable quality and content. This also implies that monetary incentives were not counter-productive to creativity compared to voluntary contributions.

We find that framing around the mission of the organization resulted in responses that were sensitive to the gender of the solicited person. Women's participation was greater when emphasizing the patient care whereas men's participation was significantly lower, controlling for the profession. At the same time, we do not find gender-based differences with respect to participation in the PRIZE treatment: women's participation was slightly higher but not significant than men's. The first finding suggests that gender may influence sensitivity to framing concerning the organizational mission (i.e., patient care). The second evidence indicates that gender differences in preferences, such as competitive inclinations or risk aversion, may not exert great influence on responses of workers to contests inside organizations.

Finally, employees were less likely to participate when solicited with the funding opportunity alone – even with a non-trivial amount of $20,000 in implementation funding – compared to all

---

[3]By comparison, in a purely public good setting such as List and Lucking-Reiley (2002)'s field experiment on individual monetary contributions to charity, the authors find very similar participation rates between 3 and 8 percent. Similarly, in a setting that involves employees of a consulting company making proposals to clients with no clear public good incentive, Gibbs et al. (2015) finds participation rates that are slightly higher (about 10 percent) but over a two-year period versus our four-week competition.

other treatments. This finding suggests that the opportunity to lead implementation of one's own submitted idea, a non-production task, was not perceived as a reward.

2   Literature

The present study contributes to the literature on the use of prizes (relative incentives) in the workplace (Lazear and Rosen, 1981; Green and Stokey, 1983; Mary et al., 1984, among others). Our main contribution consists in studying the role of prizes in fostering workers' participation in the field and in activities that produce organizational improvements. In particular, the observed increase in participation in the PRIZE treatment is consistent with the results of existing empirical studies (Bull et al., 1987; Knoeber and Thurman, 1994; Eriksson, 1999; Ehrenberg and Bognanno, 1990; Terwiesch and Xu, 2008; Terwiesch and Ulrich, 2009; Boudreau et al., 2011, 2016). However, while most studies focus on tournaments that result in benefits enjoyed exclusively by the sponsor of the competition (increasing sales, production, revenues), we show that this positive result generalizes to situations that generate positive externalities for the contestants (innovation projects to improve the organization). Despite being a common situation,[4] this setting has received relatively less attention in past studies.

Our work also contributes to the empirical literature on the use of contest-type lotteries to finance public goods that was first studied by Morgan (2000). A number of works have shown a positive effect of prizes on the extent of individual contributions to a public good in the laboratory (Morgan and Sefton, 2000; Dale, 2004; Lange et al., 2007) and in the field (Landry et al., 2006). However, as noted by Vesterlund (2012), the existing evidence on the profitability of contest-type mechanisms for raising money for public goods (e.g., charity donations) is only mixed. Our work provides further evidence to this theory as we extend the results of past studies to an organizational setting. Within this setting, individual contributions are non-monetary but consist of time and effort in putting forward (and implementing) a proposal, and the public good consists of the potential improvements for the organization. Under such circumstances, we find evidence indicating that offering prizes can effectively raise the level of participation (compared to voluntary mechanisms) and overall appear to be a profitable solution for organizations.

Our results also relate to the literature on social preferences at work. A number of studies have shown that people tend to contribute to public goods despite strong incentives to free ride. According to the World Health Organization, about 60 percent of blood donations collected globally each year is from voluntary unpaid blood donors. According to List (2011), charitable gifts of money are worth two percent of gross domestic product for the United States. Lacetera et al.

---

[4]Members of the same organization often end up competing on the basis of their ability to solve common issues at the organizational level, such as addressing specific problems, innovating, providing business ideas.

(2014) stresses that: "27% of Americans volunteer with formal organizations, for a total of about 8 billion hours per year." A sizable scientific evidence on this topic comes from a series of studies based on economic games in the laboratory (see Levitt and List, 2007, for a review) and in the field (Bandiera et al., 2005; Della Vigna et al., 2016). Bandiera et al. (2005), for instance, shows that workers internalize preferences of co-workers and may reduce effort under relative incentives. Likewise, Della Vigna et al. (2016) shows a positive effect of mission-oriented preferences, also called "vertical social preferences,"[5] on the level of effort of free-lance workers folding envelopes for a charity. Our work is consistent with a positive effect of vertical social preferences; adding evidence that such social preferences not only increase effort in mandatory pre-specified tasks but also affect voluntary participation to non-mandatory ones.[6]

Another important aspect of the present study is that we focus on incentives to carry out a complex task (writing a project proposal) as opposed to standardized production tasks (Knoeber and Thurman, 1994) or sport (Ehrenberg and Bognanno, 1990). We also focus on a competition among individual workers instead of teams (e.g., Erev et al., 1993, Hamilton et al. (2003) and more recently Gibbs et al. (2015)). This allows us to remove from consideration important team dynamics such as peer pressure, monitoring, reciprocity among team members that may also affect the participation and effort quality of employees.

Finally, our work provides support to the incentive effect of a personal satisfaction derived from helping the organization achieving its goals (Akerlof and Kranton, 2005; Besley and Ghatak, 2005; Delfgaauw, 2005; Delfgaauw and Dur, 2008; Prendergast, 2007). This type of altruism is believed to be an important driver of effort for workers in organizations for social public goods, such as hospitals, universities, schools, administrations, and the military. Theoretical models suggest different ways in which managers can exploit these pro-social motivations to raise individual productivity; in the current study, we use framing to make particular motivations salient. We find that emphasizing pro-social motivations has countervailing effects on participation; negative for men and positive for women.[7] While this finding is consistent with altruism being one important driver of effort inside organizations, it also suggests that people are sensitive to framing and in ways that may be difficult to predict ex-ante.

---

[5]Social preferences towards peers are instead called "horizontal."

[6]In Della Vigna et al. (2016), workers can choose how much effort to exert but cannot choose which task to work on (in this sense the task is "mandatory").

[7]Concerning framing, many studies have explored the effects of positive or negative framing on the private provision of public goods in the laboratory (Andreoni, 1995). Inside organizations, Hossain and List (2012) and Hong et al. (2015) are among the first studies to measure the impact of framing interventions on productivity. The current study adds to this literature by showing significant effects associated with a particular type of framing such as appealing to internal motivations towards the mission of the organization.

3 Analytical framework and predictions

In this section, we conceptualize an internal solicitation for innovation project proposals to improve the operations of the organization as a voluntary contribution mechanism for a public good. Successful proposals are viewed as non-excludable because innovation leads to improvements for everyone in the workplace (including customers by increasing the quality and efficiency of the services provided). Submitting a proposal requires costly effort by employees, such as the time to identify a problem, form a proposal, write up a concise description, and the potential for further involvement during proposal implementation.

Consider a linear model of the utility of a typical employee who contributes x and benefits from total contributions of $Y = \sum x$:

$$u(R,\ Y) = \gamma Y + \delta x + \frac{x}{Y} R - cx. \tag{1}$$

The benefits of contributing derive from three sources. First, there is an altruistic benefit from the improved workplace, γY. The altruistic benefits are the crux of public goods. Only the existence of an improved workplace is desired and the source of contributions is irrelevant. Thus, everyone would prefer to free ride on others' efforts. Second, participants have some chance of winning the contest and can expect to derive benefits from the prizes, $\frac{x}{Y} R$, where, for simplicity, all efforts have an equal chance of being selected as the winner, as in Morgan (2000). The personal reward R can be thought of as a pecuniary prize, but it could also be an increase in prestige or recognition or any combination of the above. Finally, employees may have an egoistic motivation for contributing "per se," regardless of winning and the effect on others, which is captured by δx. This includes the case in which workers may derive a personal satisfaction from contributing personally to the organization, often called warm glow preferences for giving (Andreoni, 1995). Since we cannot observe the distinction between altruistic and warm-glow motives in our empirical setup, we are going to impose later that these preferences are such that $\delta = 0$.

Contributors incur some cost from developing and submitting a proposal, cx. If there are n employees the public goods dilemma arises when $\gamma + \delta < c < n\gamma + \delta$. Then no individual would contribute without a reward as costs exceed individual benefits, but everyone would be better off if everyone contributes.

Suppose contributing a proposal is a discrete choice by employees. An employee can either contribute a single proposal $x = 1$ and receive utility of

$$u_1 = \gamma \hat{Y} + \delta + \sum_{k=1}^{n} \Pr(Y = k) \frac{R}{k} - c, \tag{2}$$

where $\hat{Y}$ denotes the expected level of contributions and $\Pr(Y = k)$ is the probability of having

k total contributions. Or they can contribute nothing $x = 0$ and receive utility of

$$u_0 = \gamma(\hat{Y} - 1). \tag{3}$$

If there are n employees, then the unique symmetric mixed-strategy equilibrium is for each employee to contribute a proposal with probability $p > 0$. After using the binomial probability for $\Pr(Y = k)$, the payoff-equating condition to find a mixed-strategy equilibrium is:

$$\frac{1 - (1 - p)^n}{np} = (c - \gamma - \delta)/R. \tag{4}$$

This equation admits one single solution p^* which cannot be expressed explicitly. Using a first order Taylor expansion around p, the equilibrium probability can be approximated as follows:

$$p^* \approx \frac{2(R - c + \gamma + \delta)}{(n - 1)R}. \tag{5}$$

The analysis of the above model is used to derive the following predictions.

1) The probability of contributing a proposal to improving the organization is zero when the prize for winning is sufficiently small relative to the individual cost of effort minus the preference for the public good (i.e., $R < c - \gamma + \delta$).

2) The probability of contributing a proposal to improve the organization increases with the value of the prize for winning.

3) The probability of contributing a proposal to improve the organization increases with the extent of individual preference for the public good ($\gamma + \delta$).

Now suppose that the public good Y constitutes the sum of innovation projects to improve the organization. Imagine that the quality of each project is randomly drawn from a discrete distribution, the same for every contributor (every employee who contributes is assumed to be equally likely to come up with a useful idea). Each proposal can be of high quality with probability ν and of low quality with probability $1 - \nu$. If a proposal is of low quality, then the value for the organization is normalized to zero. The quality of proposals is learned only after the agent paid the cost of effort. Now the equilibrium public good Y is not deterministic but follows a binomial distribution with average $E[Y] = p^{**}\nu n$, where the equilibrium probability p^{**} can be derived as before with the only difference being that it is also an increasing function of the probability ν. This leads to the following prediction.

4) If the public good depends on the quality of each contribution and every agent is equally likely to make a proposal of high quality, then the higher the probability of contributing, the higher is the average public good.

This framework can be extended to the case of individuals with heterogeneous costs. In the appendix, we explicitly consider the case of two types of individuals with different marginal costs of effort that form two groups of equal size. The symmetric mixed-strategy equilibrium is then characterized by the vector of probabilities of contributing with a proposal $(p_1^\star, p_2^\star)$. Here, the analysis of the payoff-equating conditions for the mixed-strategy equilibrium shows that the higher the marginal cost of effort minus preference for contributing, the lower the equilibrium probability of individuals (i.e., $p_1^\star > p_2^\star$ when $c_1 < c_2$, and vice versa). This leads the final prediction.

5) If individuals have heterogeneous costs, then the probability of contributing a proposal to improve the organization is higher for agents with lower costs (positive sorting).

4   Experimental Design

4.1   The context

The Heart Center is a leading academic medical center specializing in clinical cardiac care and research in the United States. Founded more than a hundred years ago, the Heart Center serves thousands of patients every year, occupies more than 35,000 square feet of office space, and employs more than 1,200 people (nurses, physicians, researchers, technicians, and administrative staff) scattered across several buildings on the Massachusetts General Hospital's main campus in downtown Boston and a few other satellite locations.

The study was in cooperation with the Heart Center's launch of the Health-care Transformation Lab (HTL), an initiative aimed at developing innovative health care process improvements to enhance the health care safety and delivery of the hospital.[8] The launch of the HTL was accompanied by the announcement of an internal "innovation contest," called the Ether Dome Challenge (the name is taken from a historical place on MGH's main campus where the first public surgery using anesthesia was demonstrated in 1846) that sought to engage all staff members to participate.

The communication around the innovation contest highlighted the opportunity for staff to help in the selection process of the ideas and a commitment by the Heart Center Management that the leading ideas would be provided appropriate resources so that they could be implemented. The announcement on the contest's website read:

> "If you've noticed something about patient experience, employee satisfaction, workplace efficiency, or anything that could be improved; if you've had an inspiration about a new way to safeguard health; or if you simply have a cost-saving idea, then now is the time to share your idea."
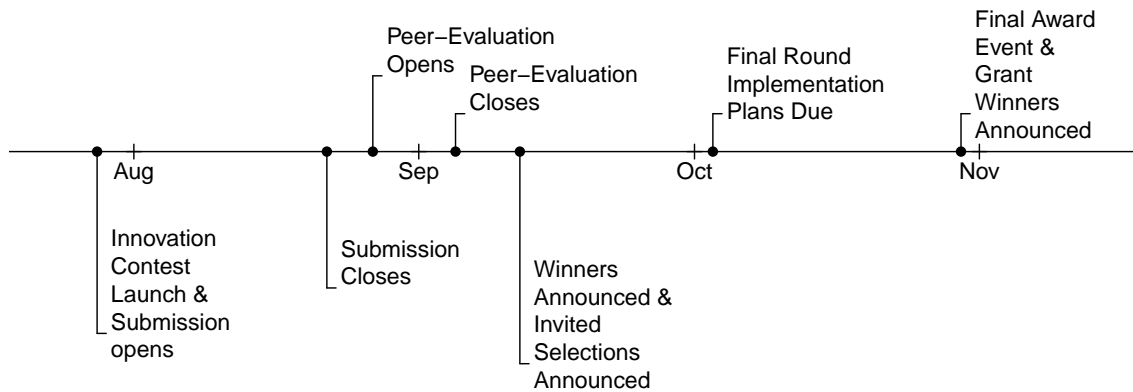
Figure 1: Timeline of the innovation contest

The innovation contest was divided into three main phases: the submission phase, the peer evaluation phase, and the implementation phase. The timing is shown in Figure 1.

In the four-week submission phase, all staff members were encouraged to identify one or more organizational problems and submit proposals addressing them. Employee participation was voluntary. All project submissions were done online via the website of the contest. There was no limit to the project proposals to submit (proposals could cover any issue within the organization, as described above), but each proposal was limited to approximately 300 words to lower the costs of entry and encourage broader participation. To ensure that treatment effects could be isolated, identified, and matched to participants, team submissions were not permitted. Limiting submissions to individual participation allowed us to match each submitter's characteristics to the randomly assigned treatment. It also lowered incentives to communicate or exchange information with other employees. Also, the website was designed to not provide any information about the status of the contest during the submission period. In this way, decisions could not be easily influenced by the perceived popularity of the contest or previous submissions.

In the two-week peer evaluation phase, all staff members were invited to rate the merit and potential of submitted proposals on a five-point rating scale. All evaluations were done online on the website of the contest. Each signed-up employee was shown a list of anonymized proposals to read and rate. Proposals were presented at random in batches of 10 each. Each proposal was described by a title, a main description of the problem to solve, and the proposal. Voting was then introduced by the following text: "Rate this idea" followed by the rating scale: 1-low; 2; 3; 4; 5-high. Ratings were kept confidential and the website did not provide any feedback or

---

[8]See http://www.healthcaretransformation.org for more information about the HTL initiative.

any other kind of additional information that might have influenced individual judgment until the voting phase was over. Evaluators were free to decide how many (and which) proposals to rate. Since these were presented in a random order, every proposal had on average the same exposure to people asked to rate its quality. Evaluators were offered a limited edition T-shirt as a compensation for the effort in voting.

In the final implementation phase, employees having submitted proposals highly rated by peers and judged as particularly promising by the HTL staff were invited to submit a full proposal detailing plans for implementation. Following evaluation by MGH senior leadership, top proposals were selected to receive support and funding for implementation. This final phase took a few months to complete, essentially the time necessary to select and implement the best projects.

4.2   The design

The basic idea of the experiment was to randomize the content of the communication announcing the innovation contest to all staff members. The start of the submission phase was indeed announced to everyone in a series of personalized emails. A direct message was sent to each contact in the list of employees' emails from our subject pool.

The content of this communication with a placeholder for our treatment is reported below (a copy of the exact email is in the Appendix).

> Dear Heart Center team member,
>
> **Submit your ideas to [TREATMENT HERE]**
>
> The Ether Dome Challenge is your chance to submit ideas on how to improve the MGH Corrigan Minehan Heart Center, patient care and satisfaction, workplace efficiency and cost. All Heart Center Staff are eligible to submit ideas online. We encourage you to submit as many ideas as you have: no ideas are too big or too small!
>
> Submissions will be reviewed and judged in two rounds, first by the Heart Center staff via crowd-voting, and then by an expert panel. Winning ideas will be eligible for project implementation funding in the Fall of 2014!

The first paragraph of the above message was randomized into *four* different solicitation treatments creating as many treatment groups of equal size (Table 1).

In the first treatment group (PRIZE), the solicitation nudged employees to participate by announcing a personal prize (iPad mini's) for top submissions. In the second treatment group (FUND), the solicitation nudged employees to participate by announcing a $20,000 budget for developing their project proposals. In the remaining two groups, the solicitation "framed" the

contest as an opportunity to improve the health care of their patients (PCARE) or the workplace (WPLACE). The exact words used in each group are reported in Table 1. In all groups, employees were not told that they were part of an experiment.

Table 1: Experimental design

	Employees:		*Randomized solicitation:*
	Obs.	%	Submit your ideas to
PRIZE	312	25	win an Apple iPad mini
FUND	308	25	win project funding up to $20,000 to turn your ideas into actions
PCARE	310	25	improve patient care at the Heart Center
WPLACE	307	25	improve the workplace at the Heart Center
Total	1237	100	

A sample size of more than 300 units for each treatment ensured a sufficiently high statistical power based upon standard power calculations on the difference of proportions (Cohen, 1992). In testing the difference of proportions between any two treatments, the probability of type-I errors was slightly below 0.80 for *small* differences at 5 percent significance level but higher than 0.80 for *medium* and *large* differences at the more stringent 1 percent significance level.[9]

Also, note the lack of a traditional "control" treatment in this study. Since the experiment was run in a workplace, we were constrained to carry out treatments having equal chances of being successful. This prevented us from having a 'null' treatment with no personalized incentives messaging as a control group. Indeed, the analysis focused on multiple comparisons of several unordered discrete treatments (e.g., prizes vs funding vs framing).[10]

The website of the innovation contest had supporting information about the available prizes, funding, and timing of the initiative. The website also required an institutional email address to login. Using this feature, we designed the website graphics and layout to reinforce the effect of the announcement: the headings, background images, a short video, and the space just below a "submit your ideas" button were designed to show the exact same first paragraph of the solicitation that the employee received by email (i.e., text in Table 1).

---

[9]The definition of small, medium and large differences is given by Cohen (1992); e.g., a difference of 5 percentage points of the pair $(0.05, 0.10)$ is considered a small effect: see Cohen (1992) p. 158.

[10]Nevertheless, if we were to think of one treatment as the benchmark against which to compare the others, the FUND treatment would be our best candidate because giving information about the size of available funding is the default option for announcing grant programs and was part of the HTL's initial design before our cooperation in the experiment.

The MGH management and the HTL staff members were blind to group assignment, which prevented potential bias in the communication of the innovation contest that was not under our direct control. We also made an effort to create a "safe" environment for employees submitting proposals by making clear (in the application form) that the identity of the proponents was going to be kept private unless the employee self-identified, so that management could not identify workers without their consent.

Finally, we relied only on official channels for communication to strengthen the effect of the announcement and signal legitimacy of the contest. Each employee received the same exact solicitation email three times: at the launch, eight days from the launch and two days before the end of the submission phase of the challenge. Starting from the second week of the submission phase, information booths, flyers, and posters were used to encourage everyone to take part in the event and respond to the email solicitation. These flyers and posters were based on a generic, undifferentiated version of the solicitation email without the text of the treatments.

5 Data

Our subject pool was the entire population working at the Heart Center as of the end of 2014, a total of 1,237 individuals. For each individual, we collected administrative data on the gender, the type of profession, and whether they had a fixed office location or not. Additional, complementary data were obtained for a limited group of 378 employees (31 percent). These extra data had self-reported information about employees' demographics, such as age and years of tenure at the Heart Center, that were obtained from an online survey that was run about two months before the launch of the innovation contest.

Table 2 presents summary statistics showing that the variables in the four treatment groups were statistically balanced.

Notice that the large majority (72 percent) of employees in our sample were women. This is due to the high fraction of workers being nurses (52 percent) and the presence of a gender separation by profession with nurses being predominantly women (92 percent). It is also important to remark that, although we do not have data on income, there were large differences in earnings by profession. According to the United States Bureau of Labor Statistics, the median annual wage of a physician was $187,200 in 2015, which is about 60 percent higher than the that of a registered nurse ($67,490) and about 70 percent higher than that of a laboratory technician ($38,970).

Table 2: Summary statistics by treatment

	Assigned treatments:				All:		
	FUND	PCARE	WPLACE	PRIZE	%	Obs.	P-value
Other	30	30	26	32	29	362	0.844
MD/Fellow	19	18	18	18	18	226	
Nursing	51	52	56	51	52	649	
Female	69	70	75	75	72	890	0.159
Male	31	30	24	26	28	347	
No office	50	46	47	45	47	577	0.556
Office	50	54	52	56	53	660	
18-25 years old*	6	8	8	6	6	24	1.000
26-35 years old*	29	29	31	26	29	107	
36-45 years old*	18	19	24	16	22	81	
>45 years old*	44	46	51	45	42	157	
< 10 years tenure*	40	31	36	37	36	132	0.891
10-20 years tenure*	26	29	38	28	30	111	
20-30 years tenure*	12	19	15	10	14	50	
30-40 years tenure*	10	16	15	12	13	48	
>40 years tenure*	10	4	8	8	8	28	

Notes: This table reports the percentage of employees in our sample cross tabulated by the assigned treatment across the gender, profession, whether the employee had a fixed office location, age, and years of tenure at the Heart Center. For each categorical variable, the last column reports the p-value from a Pearson's Chi-squared test with the assigned treatment and the variable. The asterisk * indicates non-representative self-reported information obtained from an online survey polling 378 employees that was run about two months before the launch of the innovation contest.

6 Results

6.1 Submitting project proposals

At the end of the four-week submission phase, we collected a total of 118 project proposals made by 60 employees (excluding an additional 20 proposals from 11 employees who were not part of the Heart Center when the experiment was designed). As shown in Table 3 (left panel), the percentage of employees submitting project proposals was highest in the PRIZE treatment, followed by the WPLACE treatment, the PCARE treatment, and the FUND treatment. Table 3 (right panel) also presents statistics for the count of project proposals per person. Based on these data, we find a statistically significant (a Fisher's Exact Test for Count Data gives a p-value of 0.026) association between submission rates and treatments, but no significant difference in the count of proposals (a Kruskal-Wallis rank sum test gives a p-value of 0.787). Therefore, while we detect treatment effects on participation rates (the "extensive margin"), there is no evidence indicating effects on the intensity of participation as measured by the count of submitted project proposals (the "intensive margin").

Table 3: Outcomes of the submission phase

|	Submitting proposals:			Submitted proposals:		
	No	Yes	% yes	Total	Mean	Median
PRIZE	289	23	7.4	40	1.7	1
FUND	301	7	2.3	11	1.6	1
PCARE	296	14	4.5	36	2.6	1
WPLACE	291	16	5.2	31	1.9	1
Total	1177	60	4.9	118	2.0	1

A pairwise comparison of the probability of submitting project proposals (Figure 2) reveals that employees in the PRIZE treatment were significantly more likely (5 percentage points) to submit than those in the FUND treatment. We also find a significant positive difference (about 3 percentage points) between the WPLACE and FUND treatments, although slightly below the 95 confidence level. These results are robust to bootstrap resampling that yields smaller confidence levels (see the Appendix). Also, using the more conservative Holm-Bonferroni correction for multiple comparisons gives essentially the same results (see the Appendix).[11]

Results are also robust to restricting attention to staff members that were then selected and

---

[11]The Holm-Bonferroni precedure is perhaps too conservative in this case, also considered the experimental intervention was fairly small (small effect sizes).

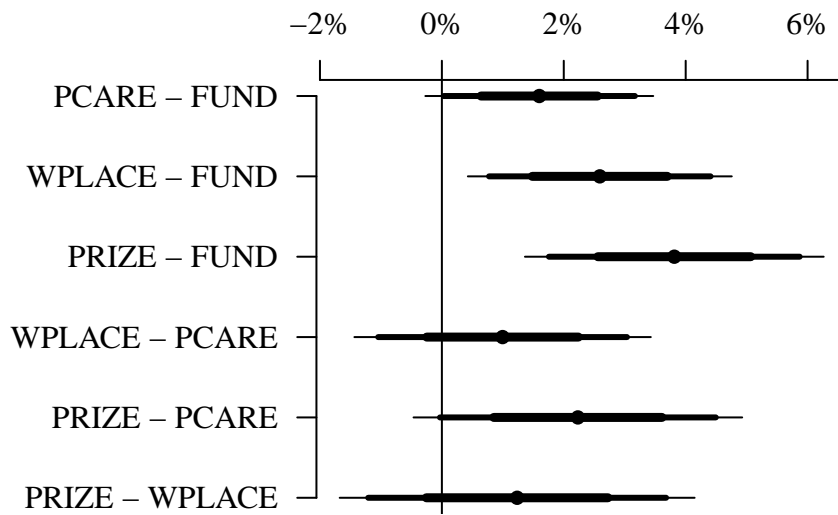Figure 2: Difference in the probability of submitting project proposals
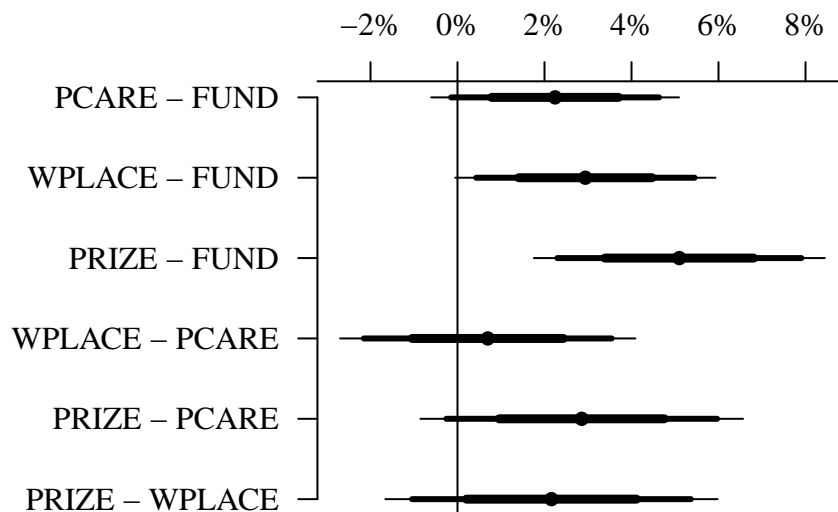
Notes: This figure plots the point estimates of the difference plus ± 1, ± 1.6, and ± 2 standard errors. Bootstrap resampling and confidence intervals based on the more conservative Holm-Bonferroni method yield very similar results (see the Appendix).

invited by the HTL staff to submit implementation plans for their proposals. Of the 29 workers invited to participate in the implementation phase, most were in the PRIZE treatment (13 employees), followed by the WPLACE treatment (9 employees), the PCARE treatment (6 employees), and the FUND treatment (only 1 employee). Also in this case, a pairwise comparison of the probability of submitting proposals and being selected (Figure 3) returns a significant and positive difference in participation between the PRIZE and FUND treatments, as well as between the WPLACE and FUND treatments.

A potential concern with a causal interpretation of the above differences lies in the possibility of *contamination* among experimental units, a topic we will discuss in greater detail in Section 7. For the moment, let us point out that a "contaminated" sample will yield estimates of the difference in participation biased towards zero. Intuitively, if everyone was exposed to the content of each solicitation, participation would be the same in each condition. Therefore, if solicitations were shared through face-to-face communication, one should expect participation rates to quickly converge over time. Contrary to these expectations, an analysis of the submissions over time (Figure 4) does not show signs of a strong convergence. The growth of the number of staff submitting proposals in the PRIZE was higher in almost each week. Only in the last week, participation in the WPLACE and PCARE had a little boost exceeding that in the PRIZE treatment. Thus, if

Figure 3: Differences in the probability of submitting finalist project proposals

Notes: This figure plots the point estimates of the difference plus ± 1, ± 1.6, and ± 2 standard errors. Estimates have been adjusted for the small counts of finalists (Agresti and Caffo, 2000) resulting in more conservative confidence intervals. Bootstrap resampling and confidence intervals based on the more conservative Holm-Bonferroni method yield very similar results (see the Appendix).

anything, the effects of contamination occurred at the very end of the competition. And even so, these might have biased downwards (instead of inflating) the estimated positive effect of prizes on participation. In this sense, our interpretation of a large effect of prizes on participation is robust to contamination.

Figure 4: The dynamic of submissions
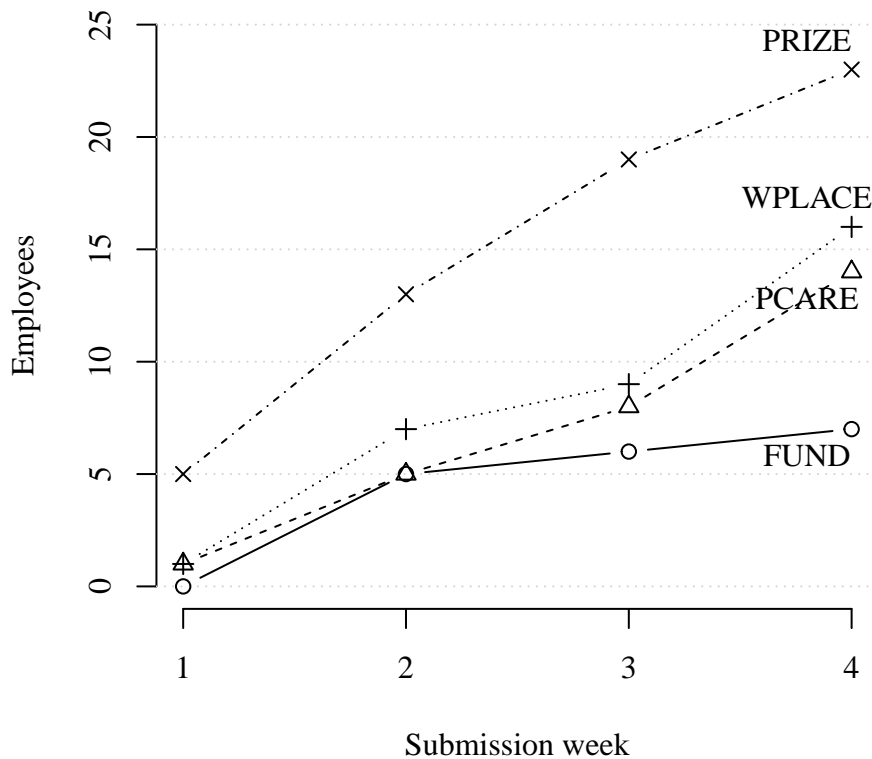
Notes: This figure plots the staff submitting proposals over the four weeks of the submission period in each condition.

We now turn our attention to the role of differences in the opportunity cost of participation (following Hypothesis H5 in Section 3). Though staff may benefit from organizational improvements in similar ways, the opportunity cost of contributing time and effort to improvement will likely vary with the gender, profession, and organizational role of the employee. For example, the opportunity cost of time may reflect the sharp hourly wage differences by profession (e.g., physicians should face higher opportunity costs than other staff members). Differences in preferences between the genders can also have an impact on participation. For example, the willingness to volunteer effort for the benefit of others might be different between men and women. To study these hypotheses,

we now model the conditional probability of submitting proposals as follows:

$$\Pr(\mathrm{SUBMIT}_{ij}) = \alpha + \tau_j + \mathrm{JOB}_i + \mathrm{MALE}_i + \mathrm{OFFICE}_i, \qquad (6)$$

where the dependent variable SUBMIT_{ij} is 1 if the employee i in treatment j has made a submission, and zero otherwise; the parameter τ_j denotes a change associated with the treatment j controlling for the employee's profession (JOB_i), the gender (MALE_i), and a dummy for office location (OFFICE_i) that indicates whether the employee had a permanent office instead of being assigned to a ward.[12]

Table 4 reports the estimation results. To simplify interpretation, coefficients are multiplied by 100 to indicate the percentage point change in the probability of submitting and treatment coefficients must sum to zero to indicate deviations in the average probability.[13] First, note that treatment differences do not change because of the individual controls, which is reassuring given the randomization. Then, by looking at the results of the full model (Column 5), employees in the PRIZE treatment were 2.4 percentage points *more* likely to submit compared to the average, whereas employees in the FUND treatment were 2.4 percentage points *less* likely to do so. Subtracting these two effects gives 4.8 which is the difference in the probability of submitting between PRIZE and FUND treatments. In a similar way, the difference in the probability of submitting between WPLACE and FUND treatment is 2.7 percentage points, which is mildly significant (p=.083).

In columns 2 and 5, we examine effects associated with the profession of the employee. One might expect employees to sort by profession because differences in income between hospital employees can be sharp.[14] The coefficient for nurses is indeed positive and negative for physicians, consistent with sorting. However, these effects are not statistically different from the residual category of other workers, as well as from one another.

In columns 3 and 5, we examine possible differential effects on participation between men and women. Although a large literature in economics and psychology (Croson and Gneezy, 2009) has documented a lower propensity of women to become involved in competitive activities (Niederle and Vesterlund, 2007), we do not find evidence of such a difference. In our setting, women are as likely as men to submit proposals.

In columns 4 and 5, we show a positive effect on participation associated with the worker having a fixed office location, as opposed to being assigned to a ward. In our context, having a

---

[12]Much of the clinical staff might be mobile and only half of the employees (53 percent) had fixed office locations, as they may be on duty in multiple wards. More senior staff tend to have a fixed location. So, within each profession, this measure can be viewed as a proxy for status inside the organization.

[13]This is just a normalization; results are not affected by using a different parameterization such as using one treatment as a specific reference category.

[14]As mentioned before, the median wage of a physician is about 40 percent higher than the that of a registered nurse.

Table 4: Probability of submitting proposals

			Dependent variable:		
			$SUBMIT_{ij} = 1$		
	(1)	(2)	(3)	(4)	(5)
PRIZE	2.53**	2.53**	2.52**	2.46**	2.45**
	(1.21)	(1.21)	(1.21)	(1.21)	(1.21)
WPLACE	0.37	0.37	0.35	0.38	0.30
	(1.09)	(1.09)	(1.10)	(1.09)	(1.10)
FUND	−2.57***	−2.57***	−2.55***	−2.49***	−2.38***
	(0.86)	(0.86)	(0.85)	(0.86)	(0.85)
Job (nursing)		0.14			1.85
		(0.82)			(1.23)
Job (MD)		−0.31			−1.14
		(1.03)			(1.24)
Male (yes)			−0.54		−0.42
			(1.33)		(1.64)
Office (yes)				2.79**	4.56***
				(1.20)	(1.60)
Constant	4.84***	4.78***	5.00***	3.35***	1.97
	(0.61)	(0.66)	(0.73)	(0.75)	(1.25)
Log Likelihood	-5545	-5545	-5545	-5542	-5540
Observations	1,237	1,237	1,237	1,237	1,237

Note: This table reports OLS estimates with heteroskedasticity robust standard errors in parenthesis. All coefficients are multiplied by 100 to indicate the percentage point change in the probability of submitting. Treatment coefficients indicate the percentage point deviation from the overall probability of submitting (there is no specific reference category). The asterisks ***, **, * indicate significance at 1, 5 and 10 percent level, respectively.

fixed office location is highly correlated with the type of profession. For example, nurses are more likely to being assigned to a ward than physicians or administrative workers, due to the nature of their job. Within each profession, however, having a fixed office location is usually correlated with the hierarchical position inside the organization. Hence, this variable is potentially controlling for income and hierarchical differences occurring within each profession.

Viewing these results through our theoretical model, it appears that the contribution cost, c, may not change much between different categories of workers. This interpretation makes sense because everyone could have an idea on how to improve the organization, regardless of profession or background skills. Proposals were also required to be short and nontechnical in order to keep individual costs of participation small for everyone. On the other hand, the cost appears systematically lower for those with a fixed office; and one may speculate that these are employees higher up in the hierarchy with more experience of existing organizational problems and the available solutions and, therefore, lower costs for contributing project proposals.

6.1.1 Interactions

We now turn to examine treatment interactions involving the employee's gender and profession.[15] Following extensive literature on differences in preferences between men and women (Croson and Gneezy, 2009), gender interactions might occur as a result of three main factors: differences in risk taking, social preferences (willingness to contribute to public goods), and competitive inclinations. If women prefer to work on activities that are less risky, more pro-social (e.g., aiming at improving people's health) and where competition is less intense, then we should observe significant treatment interactions. Similarly, one may also expect treatment interactions associated with the employee's profession since the information of a fixed-value prize (i.e., the PRIZE treatment) could be relatively less effective for employees with a higher income, such as doctors, than the others.

As shown in Figure 5 (left panel) men were significantly less likely (about 5 percentage points) than women to submit proposals in the PCARE treatment, while there was no gender difference in the other treatments. Figure 5 (right panel) also shows that there was no difference associated with the profession: doctors are as likely to submit as any other worker in each treatment.[16]

To isolate gender and profession effects, we now employ a version of model (6) with gender-treatment interactions.[17] Estimates are shown in Table 5. After gradually adding profession and office controls, interaction coefficients remain stable across all specifications. The response of men under PCARE is about 3 times the magnitude and in the opposite direction of the women's

---

[15]We also look at interactions with office location without finding any significant difference.

[16]As before, bootstrap resampling and the Holm-Bonferroni correction yield very similar results (see the Appendix).

[17]We also run a model with profession-treatment interactions and results are simular to those shown in Figure 5.

Figure 5: Differences in the probability of submitting proposals by gender and profession
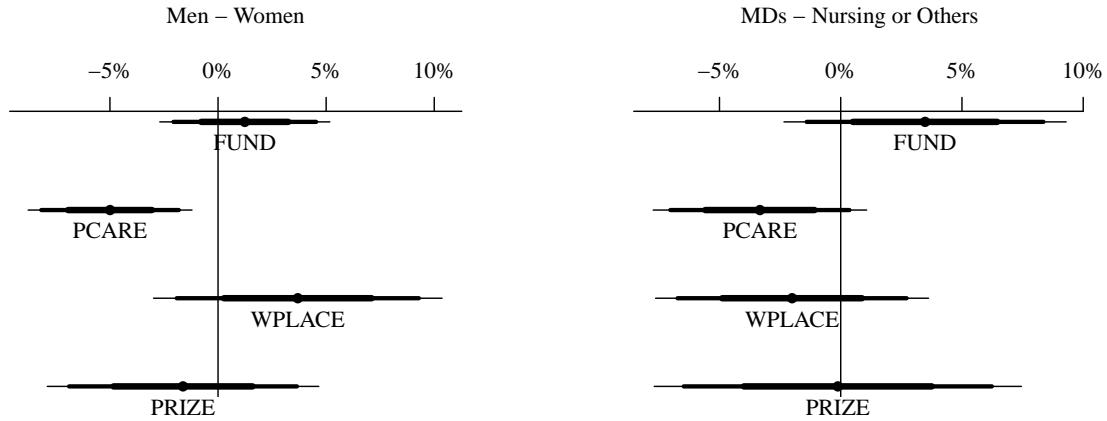
Notes: This figure plots the point estimates of the difference between men and women (left panel) and between doctors and other workers (right panel) in each treatment plus ± 1, ± 1.6, and ± 2 standard errors.

response. By subtracting these two coefficients, we find a significant difference between men and women of about 5 percentage points ($p = .018$), which is consistent with our previous analysis. Thus, and overall, we find that men responded less than women in the PCARE treatment. This effect could be due to gender differences in preferences and we will return on this in the discussion of the results.

6.2 Rating project proposals

A total of 178 employees (14 percent of our sample) ended up rating the project proposals, with each evaluator rating a median of 65.5 out of 113 project proposals (58 percent) yielding a total of 12,055 evaluator-proposal pairs.[18] Unlike the preceding submission phase, the WPLACE treatment had the highest participation (Table 6, left panel), followed by the PCARE, the PRIZE, and the FUND. However, we find no statistically significant (using a Fisher's Exact Test for Count Data gives a p-value of 0.339) relationship between rating proposals and the treatments. Likewise, the differences in the count of rated proposals (Table 6, right panel) were not statistically significant (a Kruskal-Wallis rank sum test gives a p-value of 0.286). Thus, and overall, our data indicate no prolonged effects of the treatments on both the extensive and intensive margin. This result is consistent with the general propensity of the effects from nudging and framing interventions to vanish over time.

---

[18]The projects were 118 in total but, due to a technical problem in uploading the proposals on the website for evaluation, five proposals ended up with no ratings. This problem was independent of the treatment. A Fisher's exact test rejects any association between the missed proposals and the treatment of its proponent ($p = .7$).

Table 5: Probability of submitting proposals

	Dependent variable:		
	$SUBMIT_{ij} = 1$		
	(1)	(2)	(3)
PRIZE×female	2.99*	2.95*	2.84
	(1.68)	(1.79)	(1.78)
PCARE×female	1.25	1.21	1.08
	(1.57)	(1.61)	(1.61)
FUND×female	−2.91***	−2.95**	−2.79**
	(1.06)	(1.20)	(1.19)
WPLACE×female	−0.49	−0.52	−0.62
	(1.35)	(1.44)	(1.43)
PRIZE×male	1.37	1.42	1.40
	(2.44)	(2.51)	(2.50)
PCARE×male	−3.75***	−3.72***	−3.64***
	(1.15)	(1.16)	(1.16)
FUND×male	−1.67	−1.65	−1.48
	(1.70)	(1.65)	(1.66)
Constant	4.80***	4.79***	1.87*
	(0.69)	(0.70)	(1.10)
Job	no	yes	yes
Office	no	no	yes
Log Likelihood	-5542	-5542	-5538
Observations	1,237	1,237	1,237

Note: This table reports OLS estimates with heteroskedasticity robust standard errors in parenthesis. All coefficients are multiplied by 100 to indicate the percentage point change in the probability of submitting. Treatment coefficients indicate the percentage point deviation from the overall probability of submitting (there is no specific reference category). The asterisks ***, **, * indicate significance at 1, 5 and 10 percent level, respectively.

One may find counterintuitive that there was less (although not significant) participation in the evaluation phase from employees in the PRIZE treatment than in the other treatments, given the greater participation in the submission phase. However, this result is not entirely surprising because only 70 percent of employees who made submissions resolved to rate proposals as well (we detect no difference in the propensity of submitting and rating proposals between the treatments); so, even a difference of 2 percentage points in submitting will shrink to about 1 percentage point in the rating phase. In other words, we were not expecting self-rating to affect evaluation much.

Table 6: Outcomes of the peer evaluation phase

|	Rating proposals:			Rated proposals:		
	No	Yes	% yes	Total	Mean	Median
PRIZE	269	43	13.8	2457	57.1	43
FUND	272	36	11.7	2484	69.0	68
PCARE	261	49	15.8	3413	69.7	73
WPLACE	257	50	16.3	3701	74.0	82
Total	1059	178	14.4	12055	67.7	66

6.3   The quality of the project proposals

The treatment interventions may not have only impacted the propensity to make a submission, but the quality of the submission as well. Of particular interest is any indication of a quantity versus quality trade-off. For example, if the FUND treatment which generated the fewest submissions also produced the highest quality submissions. A quality versus quantity trade-off would increase the complexity of choosing optimal incentives for employees.

The ratings collected in the peer evaluation phase of the challenge provide our main measure of quality. Figure 6 shows the distribution of the ratings received by a proposal conditional on treatment of its proponent. In each treatment, a proposal was given a rating of 3, the "neutral" point, on a five-point scale about 30 percent of the times with employees being more likely to give high (4-5) rather than low (1-2) ratings.

Figure 6 reveals that the probability of a proposal receiving a given rating was about the same in each treatment. And indeed, by aggregating the mean rating for each proposal, we do not identify any significant treatment effect (a Kruskal-Wallis rank sum test gives a p-value of 0.416). Similarly, a linear regression of mean ratings on treatment dummies does not reveal any relationship between ratings and treatments. The treatment coefficients are not significant, with the linear model not significantly different from a constant model (an overall F-test gives a p-value of 0.611).

Figure 6: Probability of a project proposal receiving a given rating in each treatment
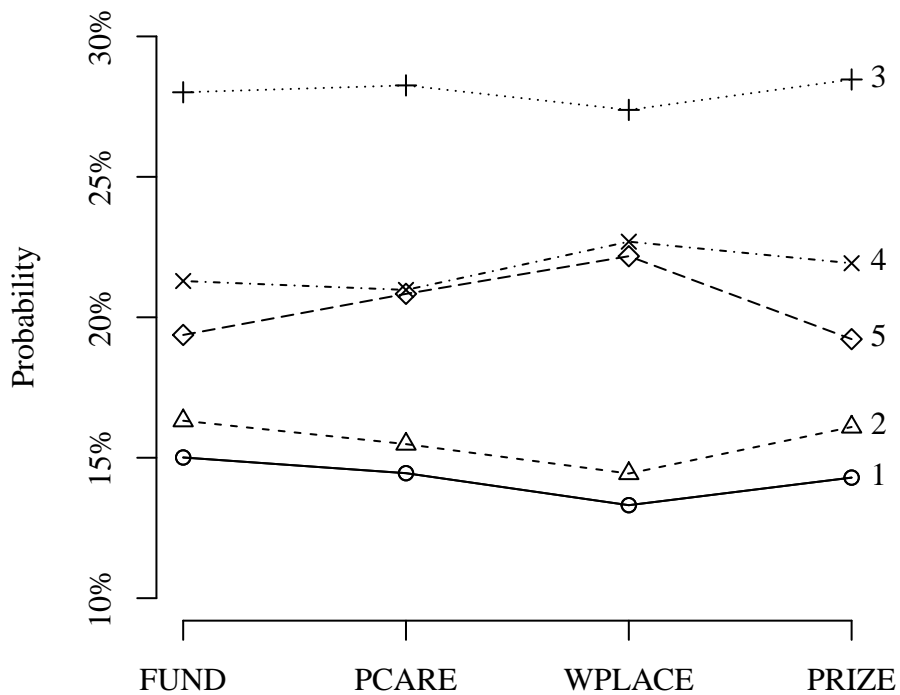
Notes: This figure plots the distribution of the ratings given to a proposal conditional on the treatment of its proponent. Each curve presents point estimates of the probability of a project proposal receiving a given rating on a five-point scale (1=Low and 5=High). Flat, non-intersecting curves indicate that there were small differences across treatments for each rating.

The above analysis on the aggregate ratings does not hold in general. It crucially relies on the assumption that an increment in a proposal's quality as measured by an increase in ratings from v to $v + 1$ is the same for any value v. So, we also examine the distribution of ratings as generated by treatments with no aggregation. We have over 12,000 ratings, providing a very sensitive test for differences across treatments. Using a Pearson's Chi-squared test we find that the hypothesis of dependence between the distribution of ratings and the treatments is *not* quite significant at the 10 percent level (p-value of 0.103). Driving the p-value is a less than 2 percent difference between the proportion of 5's in the WPLACE treatment versus the other distributions (Figure 6), which is probably due to outliers (the winning proposal was in the WPLACE treatment). Taken together with the fact that our sample is large, we have strong evidence suggesting that there are no (economically meaningful) differences in the quality of project proposals across treatments and in particular no evidence of a quantity versus quality trade-off up to the resolution of the five-point scale.[19]

One potential limit of assessing quality only on the basis of peer ratings is that the employees might have a different view of a proposal's quality than executives (due, for instance, to a misalignment of incentives). Indeed, to ensure alignment between managerial goals and the peer assessment, all project proposals were further vetted by the HTL staff before being considered for implementation funding. So, we now focus on the outcomes of this vetting process to investigate more broadly the presence of treatment effects on the quality of project proposals.

The vetting process conducted by the HTL staff resulted in 93 proposals being scored (from 1 to 100 points) with the best 29 proposals invited to submit implementation plans. The remaining 20 proposals were excluded (and received a score of zero) either because flagged as inappropriate for funding or because the proponent manifested no intention to participate in the implementation phase (a Fisher's Exact Test for Count Data finds no association between proposals excluded and treatments with a p-value of 0.652).

The Spearman's rank correlation coefficient between the scores given by the HTL staff and the average peer ratings was relatively high (0.198), indicating good agreement between our two measures of quality. Indeed, as before, we find no treatment effects on quality using the scores (a Kruskal-Wallis rank sum test gives a p-value of 0.437). We also find no treatment differences in the percentage of submitters being selected and invited by HTL staff to present additional implementation plans (a Fisher's Exact Test for Count Data gives a p-value of 0.652). Although not significant, employees who made project proposals in the FUND treatment were less likely to be

---

[19]One may worry that such binning is a fairly coarse measure of quality. In particular, effects concentrated in the upper tail of the distribution may not be detected. For example, compare the ratings of proposals A, B, C and D with hypothetical true qualities of 3, 4, 5, and 10 stars respectively. Under a five-point scale rating system, proposals A and B can be distinguished, but C and D cannot be distinguished. Hence, one needs to be very cautious in interpreting these results as evidence against quality effects in general.

selected as finalist than the others (only 1 out of 7 in the FUND treatment were selected and invited by the HTL staff), providing additional evidence of a no quantity versus quality trade-off, as discussed before.

6.4 The content of the project proposals

The goal of the challenge was to improve Heart Center operations by identifying problem areas and potential solutions. The proposed projects broadly conformed to the stated goals of the contest, aligning with improving the work processes within the organization or providing high-quality patient care. For example, one project proposal that received high peer ratings was to create a platform for patients to electronically review and update their med list in the office prior to seeing the physician. Another was to develop a smartphone application that allows the patient to see the itinerary for the day providing a guide from one test or appointment to another. Nevertheless, other contest organizers may have varying goals and be concerned about different aspects of the submissions. In order to examine additional dimensions of submission content, we now study the area of focus of the submissions. Of particular interest is understanding whether the framing intervention induced employees to concentrate on different categories. For example, while staff in the WPLACE treatment focused on improvements for the workplace, those in the PCARE treatment concentrated on interventions directly targeting the patients.

Table 7: Project proposals by area of focus

	FUND	PCARE	WPLACE	PRIZE	Total
Information and access	0	4	8	11	23
Patient support	2	8	7	6	23
Care Coordination	1	9	3	7	20
Staff workflow	4	5	4	5	18
Workplace	3	6	3	5	17
Quality and safety	0	0	5	5	10
Surgical tools and support to research	1	1	0	0	2
Total	11	33	30	39	113

Notes: The areas of focus were manually identified by the HTL staff at the end of the competition. Due to a technical problem five proposals ended up with no classification.

Members of the HTL categorized each project proposal into one of seven "areas of focus" (Table 7): three categories ("Care coordination", "Staff workflow", "Workplace") identified improvements for the workplace, other three ("Information and access", "Patient care", and "Quality and Safety") focused on improvements centered around patients, and another one ("Surgical tools

and support to research") categorized projects developing tools to support scientific research.

Using a Fisher's Exact Test for Count Data with simulated p-value (based on 50000 replicates), we find a mildly significant (p=0.089) association between these categories and the treatments.[20] The analysis of pairwise differences between treatments (Figure 7) reveals that this result is driven by differences in the "Quality and Safety" and "Information and access" categories. Project proposals in the PCARE treatment were less likely to fall in the "Quality and Safety" category. Similarly, project proposals in the FUND treatment were less likely to fall in the "Information and access" category.

It is difficult to interpret these effects because our model does not provide any prediction on the content of proposals. One possibility is that framing induced participants to concentrate on different areas of interventions but our data do not appear consistent with this story.

Figure 7: Differences in the probability of proposals being in a given area of focus in each treatment
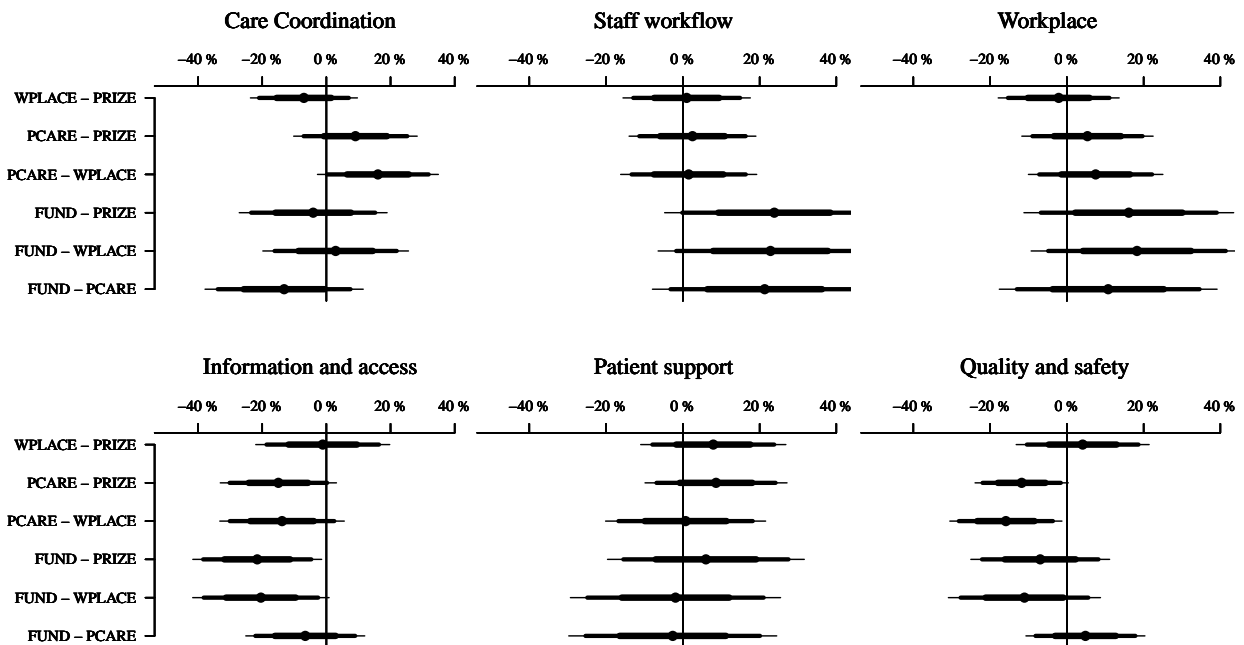
Notes: This figure plots the point estimates of the difference plus ± 1, ± 1.6, and ± 2 standard errors. Estimates have been adjusted for the small counts of the data (Agresti and Caffo, 2000) resulting in more conservative confidence intervals.

We also look at differences in the underlying complexity of the project proposal as captured by differences in the length (i.e., the word count) of a submission. Submissions were below 200 words in most cases with little differences between the treatments. Indeed, testing for a significant linear regression relationship between the length of submissions and treatment dummies returned an overall insignificant result (p=.43, F-test).

---

[20]Simulations are used to reduce the computational burden.

As a result, based on the analysis of the areas of focus and the length of the submissions, we do find only little evidence of differences in submission content across treatments. However, submission content is not a well-defined concept and could be characterized in many dimensions. While content does not vary in the dimensions we selected, we have not exhausted all possible dimensions.

6.5 Estimating social preferences

In this section, we calibrate the theoretical model developed in Section 3 with the experimental data to get a sense of the magnitude of underlying preferences for contributing to the organization. Following, the mixed-strategy equilibrium of the model, the theoretical probability of contributing must be proportional to the expected value of winning, R, the underlying preferences towards the public good, γ, the marginal costs of contributing, c, and the number of agents, n.

We assume the cost of making a submission c is the same in each treatment,[21] and the individual preferences are constant, being predetermined to our intervention. Then we derive a structural relationship between the observed difference in the probability of contributing Δp and the difference in the expected rewards from winning ΔR between the treatments. That is,[22]

$$\Delta p \approx \frac{\Delta R}{n(c - \gamma)}. \tag{7}$$

(Throughout this section we will consider $\delta = 0$ ignoring the distinction between impure and pure altruism.) By solving for γ, we get

$$\gamma \approx c - \Delta R / (n \Delta p). \tag{8}$$

This implies that the parameter capturing individual preference for the public good (that is consistent with our data) must be proportional to the ratio between the difference in rewards and the difference in the probability of submitting. Although we do not observe the levels of R in each treatment, we approximate the difference of rewards between the PRIZE and the other conditions by the pecuniary value of the reward, which has its upper bound in the highest price that can be paid for an iPad mini ($350).[23] We further calibrate the cost of submitting a proposal c to $40

---

[21]This seems a reasonable assumption, given everyone is asked to perform the same task (identical submission procedure, same word limit, etc.).

[22]This equation can be obtained by following these steps. First, we approximate the profit equating condition (4) to a linear function by noticing that the $1/(1 - (1 - p)^n)$ approximates one for n large enough and p sufficiently small. Second, we solve for p and we simplify using the definitions of Δp and ΔR.

[23]The price paid by the Heart Center was $239 at the end of 2014 (including shipping cost). Other popular models (those with cellular data and large storage) could cost as high as $350. Agents, however, were not aware of the specific model used for the competition and of the price paid. So, the value of $350 is very conservative.

which is the median income per hour of a Nurse Practitioner according to the Bureau of Labor Statistics; we assume the number of competitors n to be 30 percent of the entire sample to take into account rational expectations about the actual number of participants in the contest.[24] Finally, by substituting these calibrated values into equation (8) along with the empirical difference in participation rates between the PRIZE and the other treatments ($\Delta p = 0.037$), we get an estimate of the magnitude of the social preferences towards the organization which is $\hat{\gamma} = \$12$. As shown in Figure 8, this value is equivalent to about 30 percent reduction in the cost of contributing. Hence, increasing the prize by $100 is expected to raise the probability of submitting by 1 percentage points. This increase can be compared to the corresponding increase of 0.7 that one will obtain by assuming no social preferences $\gamma = 0$ at all.

A few remarks are in order here. To get confidence around these estimates one need to consider several sources of uncertainty. First, there is the uncertainty of estimating the probability of submitting in our sample (standard errors can be computed directly from the data). Another source of uncertainty is due to the calibration of the marginal cost or the number of competitors. As shown in Figure 8, the fraction of costs explained by social preferences increases monotonically in the number of competitors (going up to 80 percent of costs if employees expected to compete against every Heart Center staff member); and decreases monotonically in the calibrated cost of making a submission. Finally, another important source of uncertainty is regarding the main behavioral assumptions of the model, as we discuss in the next section.

7 Discussion

We report results of a natural field experiment conducted at a medical organization that held an innovation contest seeking contribution of public goods (i.e., projects for organizational improvement) from its more than 1200 employees. The experiment tested incentives for contributing by manipulating the content of emails soliciting staff participation. We presented different incentives to participate in the contest, such as a prize (iPad mini) for winning submissions, improving patient care, improving the workplace, and funding for implementation. Each staff was randomly assigned to receiving an email containing one of the four incentives. Our data show that offering a prize for winning submissions boosted participation by 85 percent without affecting the quality of the submissions. The effect was consistent across gender and job type. We posit that the allure of a prize, in combination with mission-oriented preferences, drove participation. Using a simple model, we estimate that these preferences explain about a third of the magnitude of the effect.

---

[24]This choice is our best guess of the number of active staff members at the Heart Center and is based on the number of employees who voluntarily took a survey before the experiment (378 people). Assuming greater participation would lead to artificially increasing the estimates of underlying incentives. In fact, staff members may have rational expectations about the actual number of potential participants, which may be less than the entire population.

Figure 8: Estimated value of social preferences ($\hat{\gamma}$)
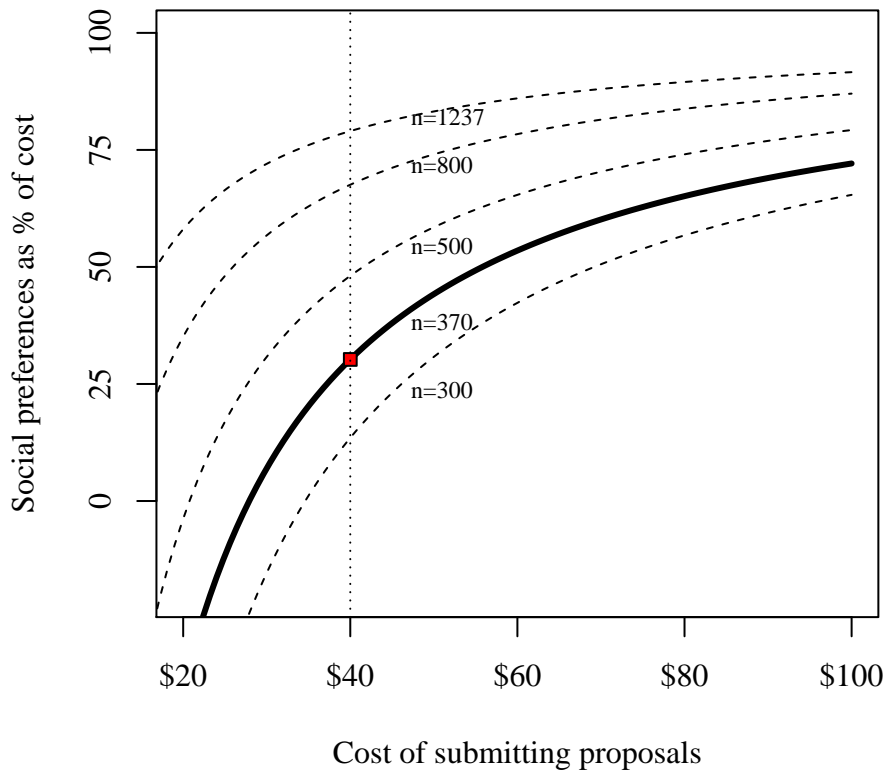
Notes: This figure plots the theoretical relationship between the cost of participation and the the social preferences parameter γ (in percentage of the costs) which is consistent with our experimental data. Different curves represents different assumptions on the number of competitors.

These results were sensitive to the solicited person's gender. Women's participation was greater than men's when email solicitations emphasized patient care, controlling for profession. One possible explanation for this finding is that female staff may be more altruistic and more responsive to the issue of improving patient care than male staff. However, the existing economic literature on gender differences in altruistic preferences (see Croson and Gneezy, 2009) does not fully support this explanation. Another possible reason for women's greater responsiveness to patient care involves gender-based stereotypes, which are pervasive in health care organizations (Evans, 2002). Male staff could have judged the contribution of an innovation project towards patient care improvement as a female-typed activity. Similarly, Coffman (2014) report that, in laboratory experiments, women were less likely to contribute ideas to groups when the topic falls in male-typed domains, e.g., sports, and vice versa.

An interesting non-finding is the lack of gender-based differences in the PRIZE treatment; we found that women's participation was equal to men's. An extensive literature in economics has revealed women's tendency to be risk-averse (Borghans et al., 2009) and a distaste for competition (Niederle and Vesterlund, 2007), suggesting that a contest for prize may discourage women's participation. In our experiment, however, women participated the most in the PRIZE treatment. We suspect that the unique characteristics of our subject pool (e.g., careers in health care) could have led to this finding.

While not observed, there could have been negative interactions between the prize incentive and the contest that encouraged staff members to contribute to public good by suggesting an organizational improvement opportunity. A negative interaction occurs when a particular incentive might crowd out the motive to contribute; for example, providing a self-serving incentive might negatively interact with a self-sacrificing activity and crowd out the motivation to participate in this activity. Experiments in the context of blood donations (Lacetera et al., 2013, 2014), or public "bads" in daycare pick-ups (Gneezy and Rustichini, 2000) indicate crowd-out effects. Pecuniary incentives may not have the same effect if in-kind gifts are used in place of currency (e.g., Kube et al., 2012), or if the setting already involves an employer-employee relationship (e.g, Fehr et al., 1998).

Ulterior motivations such as the prospect of a promotion (Baker et al., 1994; Gibbs, 1995), professional prestige, or peer recognition (Kosfeld and Neckermann, 2011; Blanes i Vidal and Nossol, 2011) could have propelled participation. However, the likelihood of career advancement through participating in the contest seems marginal as promotions in academic hospitals are typically based on the number of publications, tenure, and professional training. In addition, our email solicitations contained incentives that are fairly neutral to potential career advancement opportunities associated with the contest. Recognition for authors of winning proposals included announcement during a hospital-wide public event and listing on the hospital website. These opportunities to be

acknowledged equally affected all potential participants.

The Stable Unit Treatment Value Assumption (SUTVA) for causal inference (Rubin, 1974) is relevant for all randomized experiments. In our study, it is possible that communication among staff assigned to different treatment arms could have influenced decisions to participate. The magnitude of the resulting effects would depend on the level of interference (i.e., intensity of staff communication) and the density of staff' network of social interactions. Although this bias could severely distort our estimates, we expect the bias to cause differences in participation across treatments to converge to zero (a bias towards null effect) as communication spreads the content of the different email solicitations. Our findings exhibit little evidence of convergence. Communication among staff during the contest was likely limited as the Heart Center staff are scattered across multiple buildings on the hospital campus and competition provides weak incentives for information sharing.

Staffs' poor understanding of the costs arising from winning the contest could explain their participation. In the same vein, a large experimental literature on contests (Dechenaux et al., 2014) shows that similar misconception could lead to higher effort levels than predicted. In particular, those in the PRIZE treatment could have underestimated the costs of participating as the iPad mini accentuated the appeal of an immediate reward and diminish the costs of winning the contest and becoming involved in implementation. Across treatments, however, we find that most of the staff submitted a detailed proposal on their project when invited to do so. Although participants might have had incorrect beliefs about expected costs from winning, these are unlikely to explain the significant differences.

Finally, while the choice of focusing on health care workers may limit the generalizability of our results in some respect, it should be noted that in the US alone health care spending accounts for 17 percent of the GDP (in 2015) and, more generally, our study results are also directly applicable to a variety of other professions exposed to a public good dilemma (e.g., teachers, public servants, researchers).

8   Conclusions

Our results have implications that extend beyond the specific organization under study. Using a contest for prizes appears to be a more profitable way for firms to encourage contribution toward a public good among workers than currently acknowledged in the traditional tournament theory literature. The incentive effect of offering a prize for winning the contest interacts with pro-social motivations of workers to contribute to the public good.

References

A. Agresti and B. Caffo. Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures. *The American Statistician*, 54(4):280–288, 2000.

G. A. Akerlof and R. E. Kranton. Identity and the economics of organizations. *Journal of Economic Perspectives*, 19 (1):9–32, 2005.

J. Andreoni. Warm-glow versus cold-prickle: The effects of positive and negative framing on cooperation in experiments. *The Quarterly Journal of Economics*, 110(1):1–21, 1995.

G. Baker, M. Gibbs, and B. Holmstrom. The internal economics of the firm: Evidence from personnel data. *The Quarterly Journal of Economics*, 109(4):881–919, 1994.

O. Bandiera, I. Barankay, and I. Rasul. Social preferences and the response to incentives: Evidence from personnel data. *The Quarterly Journal of Economics*, 120(3):917–962, 2005.

T. Besley and M. Ghatak. Competition and incentives with motivated agents. *The American Economic Review*, 95(3): 616–636, 2005.

J. Blanes i Vidal and M. Nossol. Tournaments without prizes: Evidence from personnel records. *Management Science*, 57(10):1721–1736, 2011.

L. Borghans, J. J. Heckman, B. H. Golsteyn, and H. Meijers. Gender differences in risk aversion and ambiguity aversion. *Journal of the European Economic Association*, 7(2-3):649–658, 2009.

K. J. Boudreau, N. Lacetera, and K. R. Lakhani. Incentives and problem uncertainty in innovation contests: An empirical analysis. *Management Science*, 57(5):843–863, 2011.

K. J. Boudreau, K. R. Lakhani, and M. Menietti. Performance responses to competition across skill levels in rank-order tournaments: field evidence and implications for tournament design. *The RAND Journal of Economics*, 47 (1):140–165, 2016.

C. Bull, A. Schotter, and K. Weigelt. Tournaments and piece rates: An experimental study. *Journal of Political Economy*, 95(1):1–33, 1987.

G. Charness, D. Masclet, and M. C. Villeval. The dark side of competition for status. *Management Science*, 60(1): 38–55, 2013.

K. B. Coffman. Evidence on self-stereotyping and the contribution of ideas. *The Quarterly Journal of Economics*, 129 (4):1625–1660, 2014.

J. Cohen. A power primer. *Psychological bulletin*, 112(1):155, 1992.

R. Croson and U. Gneezy. Gender differences in preferences. *Journal of Economic Literature*, 47(2):448–474, 2009.

D. Cutler, E. Wikler, and P. Basch. Reducing administrative costs and improving the health care system. *New England Journal of Medicine*, 367(20):1875–1878, 2012.

D. J. Dale. Charitable lottery structure and fund raising: Theory and evidence. *Experimental Economics*, 7(3):217–234, 2004.

E. Dechenaux, D. Kovenock, and R. M. Sheremeta. A survey of experimental research on contests, all-pay auctions and tournaments. *Experimental Economics*, pages 1–61, 2014.

J. Delfgaauw. Dedicated doctors: Public and private provision of health care with altruistic physicians. *New England Journal of Medicine*, 353(15), 2005.

J. Delfgaauw and R. Dur. Incentives and workers' motivation in the public sector*. *The Economic Journal*, 118(525): 171–191, 2008.

S. Della Vigna, J. A. List, U. Malmendier, and G. Rao. Estimating social preferences and gift exchange at work. Technical report, National Bureau of Economic Research, 2016.

R. Drago and G. K. Turnbull. The incentive effects of tournaments with positive externalities among workers*. *Southern Economic Journal*, 55(1):100, 1988.

R. G. Ehrenberg and M. L. Bognanno. Do tournaments have incentive effects? *The Journal of Political Economy*, 98 (6):1307–1324, 1990.

S. Erat and U. Gneezy. Incentives for creativity. *Experimental Economics*, pages 1–12, 2015.

I. Erev, G. Bornstein, and R. Galili. Constructive intergroup competition as a solution to the free rider problem: A field experiment. *Journal of Experimental Social Psychology*, 29(6):463–478, 1993.

T. Eriksson. Executive compensation and tournament theory: Empirical tests on danish data. *Journal of Labor Economics*, 17(2):262–280, 1999.

J. A. Evans. Cautious caregivers: gender stereotypes and the sexualization of men nurses' touch. *Journal of advanced nursing*, 40(4):441–448, 2002.

E. Fehr, G. Kirchsteiger, and A. Riedl. Gift exchange and reciprocity in competitive experimental markets. *European Economic Review*, 42(1):1–34, 1998.

B. S. Frey. Not just for the money: An economic theory of personal motivation. *Books*, 1997.

M. Gibbs. Incentive compensation in a corporate hierarchy. *Journal of Accounting and Economics*, 19(2):247–277, 1995.

M. Gibbs, S. Neckermann, and C. Siemroth. A field experiment in motivating employee ideas. *Review of Economics and Statistics, Forthcoming*, 2015.

E. L. Glaeser, A. Hillis, S. D. Kominers, and M. Luca. Predictive cities crowdsourcing city government: Using tournaments to improve inspection accuracy. *The American Economic Review*, 106(5):114–118, 2016.

U. Gneezy and A. Rustichini. A fine is a price. *Journal of Legal Studies*, 29:1, 2000.

J. R. Green and N. L. Stokey. A comparison of tournaments and contracts. *The Journal of Political Economy*, 91(3): 349–364, 1983.

B. H. Hamilton, J. A. Nickerson, and H. Owan. Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation. *Journal of political Economy*, 111(3):465–497, 2003.

T. Hellmann and V. Thiele. Incentives and innovation: A multitasking approach. *American Economic Journal: Microeconomics*, 3(1):78–128, 2011.

B. Holmstrom and P. Milgrom. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics and Organization*, 7:24–52, 1991.

F. Hong, T. Hossain, and J. A. List. Framing manipulations in contests: a natural field experiment. *Journal of Economic Behavior & Organization*, 2015.

T. Hossain and J. A. List. The behavioralist visits the factory: Increasing productivity using simple framing manipulations. *Management Science*, 58(12):2151–2167, 2012.

C. R. Knoeber and W. N. Thurman. Testing the theory of tournaments: An empirical analysis of broiler production. *Journal of Labor Economics*, 12(2):155–79, 1994.

M. Kosfeld and S. Neckermann. Getting more work for nothing? symbolic awards and worker performance. *American Economic Journal: Microeconomics*, 3(3):86–99, 2011.

S. Kube, M. A. Maréchal, and C. Puppe. The currency of reciprocity: Gift exchange in the workplace. *The American Economic Review*, 102(4):1644–1662, 2012.

N. Lacetera, M. Macis, and R. Slonim. Economic rewards to motivate blood donations. *Science*, 340(6135):927–928, 2013.

N. Lacetera, M. Macis, and R. Slonim. Rewarding volunteers: a field experiment. *Management Science*, 60(5): 1107–1129, 2014.

K. R. Lakhani, K. J. Boudreau, P.-R. Loh, L. Backstrom, C. Baldwin, E. Lonstein, M. Lydon, A. MacCormack, R. A. Arnaout, and E. C. Guinan. Prize-based contests can provide solutions to computational biology problems. *Nature biotechnology*, 31(2):108–111, 2013.

C. E. Landry, A. Lange, J. A. List, M. K. Price, and N. G. Rupp. Toward an understanding of the economics of charity: Evidence from a field experiment*. *The Quarterly journal of economics*, 121(2):747–782, 2006.

A. Lange, J. A. List, and M. K. Price. Using lotteries to finance public goods: Theory and experimental evidence*. *International Economic Review*, 48(3):901–927, 2007.

E. P. Lazear. Pay equality and industrial politics. *Journal of Political Economy*, 97(3):561–80, 1989.

E. P. Lazear and S. Rosen. Rank-order tournaments as optimum labor contracts. *The Journal of Political Economy*, 89 (5):841–864, 1981.

S. D. Levitt and J. A. List. What do laboratory experiments measuring social preferences reveal about the real world? *The Journal of Economic Perspectives*, 21(2):153–174, 2007.

J. A. List. The market for charitable giving. *The Journal of Economic Perspectives*, 25(2):157, 2011.

J. A. List and D. Lucking-Reiley. The effects of seed money and refunds on charitable giving: Experimental evidence from a university capital campaign. *Journal of Political Economy*, 110(1):215–233, 2002.

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

O. Mary, W. K. Viscusi, and R. J. Zeckhauser. Economic contests: Comparative reward schemes. *Journal of Labor Economics*, 2(1):27–56, 1984.

J. Morgan. Financing public goods by means of lotteries. *The Review of Economic Studies*, 67(4):761–784, 2000.

J. Morgan and M. Sefton. Funding public goods with lotteries: experimental evidence. *The Review of Economic Studies*, 67(4):785–810, 2000.

M. Niederle and L. Vesterlund. Do women shy away from competition? do men compete too much? *The Quarterly Journal of Economics*, 122(3):1067–1101, 2007.

C. Prendergast. The motivation and bias of bureaucrats. *The American Economic Review*, 97(1):180–196, 2007.

J. Reeve and E. L. Deci. Elements of the competitive situation that affect intrinsic motivation. *Personality and Social Psychology Bulletin*, 22:24–33, 1996.

D. B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688, 1974.

C. Terwiesch and K. T. Ulrich. *Innovation tournaments: Creating and selecting exceptional opportunities*. Harvard Business Press, 2009.

C. Terwiesch and Y. Xu. Innovation contests, open innovation, and multiagent problem solving. *Management science*, 54(9):1529–1543, 2008.

L. Vesterlund. Voluntary giving to public goods: moving beyond the linear vcm. In J. H. Kagel and A. E. Roth, editors, *The Handbook of Experimental Economics*, volume 2, in preparation. Princeton University Press, 2012.