Contributing to Public Goods Inside Organizations: Field Experimental Evidence[1]

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

Last updated: 09 January, 2017

[1]Blasco: Harvard Institute for Quantitative Social Science, Harvard University, 1737 Cambridge Street, Cambridge, MA 02138 (email: ablasco@fas.harvard.edu). Jung: Harvard Business School, Soldiers Field, Boston, MA 02163 (email: ojung@hbs.edu), Lakhani: Harvard Business School, Soldiers Field, Boston, MA 02163, and National Bureau of Economic Research (email: k@hbs.edu). Menietti: Harvard Institute for Quantitative Social Science, Harvard University, 1737 Cambridge Street, Cambridge, MA 02138 (email: mmenietti@fas.harvard.edu).
We gratefully acknowledge the financial support of the MacArthur Foundation (Opening Governance Network), NASA Tournament Lab, and the Harvard Business School Division of Faculty Research and Development. This project would not have been possible without the support of Eric Isselbacher, Julia Jackson, Maulik Majmudar and Perry Band from the Massachusetts General Hospital's Healthcare Transformation Lab.

Abstract

We investigate the factors driving workers' decisions to contribute public goods inside an organization through a randomized solicitation of innovation project proposals in a medical center with over 1200 employees. We find that a contest awarding pecuniary prizes for the best contributions generates an 85 percent increase in participation, with no difference in the quality of contributions. We show that the effect on participation is partly due to workers having positive social preferences towards the organization, rather than just the value of the prize. Participation is also increased by a solicitation appealing to improving the workplace. However, emphasizing the patient mission of the organization led to countervailing effects on participation based on gender differences. Overall, these results are consistent with workers having multiple underlying motivations to contribute to public goods inside the organization consisting of a combination of pecuniary and non-pecuniary incentives associated with the mission of the organization.

JEL Classification: H41; D23; D03.

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

Notes

Dynamics. Fix problem with ”interference” 24
One possible problem concerning the above regression analysis is the relatively small number of responses per treatment compared to the sample size (response rates below 10 percent are usually seen as rare events). The main problem is that asymptotic confidence intervals may not be fully accurate [@king2001logistic]. Logistic regression models allow testing this issue by direct methods that deal with rare events, such as exact inference. Using logistic regression, we find the same results indicating these are robust under exact inference (tables available on request). 24
”xxxxx” 29
”xxxx” 29

Contents

1 Introduction **3**

2 Analytical framework and predictions **9**

3 Experimental Design **11**
3.1 The context 11
3.2 The design 14

4 Data **16**

5 Results **17**
5.1 Submitting project proposals 17
5.1.1 Interactions 23
5.2 Rating project proposals 24
5.3 The quality of the project proposals 26
5.4 The content of the project proposals 29
5.5 Estimating social preferences 31

6 Discussion **34**

```
## Warning: package 'knitr' was built under R version 3.2.5
```

1 Introduction

Inside organizations, employees often make choices between working on their regular assignment versus developing privately observed opportunities to improve the organization, such as addressing specific problems, providing business ideas, and innovating. Given the importance of these activities for the organization, one key concern is that they may decide to focus on the assigned tasks regardless of the potential for improving the organization. This generally happens because of misspecified incentives leading to multitasking problems (Holmstrom and Milgrom, 1991; Hellmann and Thiele, 2011) or the higher risk from working on non-standard tasks (Manso, 2011). But complications can also arise because employees may behave strategically and, being members of the same organization, try to avoid tasks that benefit everybody in the hope to "free-ride" on the efforts of others.

Competition in the form of promotions or bonuses can be seen as a response to this problem. An internal prize-based competition can indeed 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, the literature by large presumes that the outcomes will be enjoyed exclusively by the sponsor of the competition (i.e., the manager, entrepreneur) with little returns for the competitors beyond the prizes given to the winners. An assumption that seems unlikely when the competition aims at improvements for the organization and everyone can anticipate benefits from the work of others (e.g., solutions addressing common issues). As such, the use of contests under these circumstances raises important questions. Will prizes be effective in fostering participation and contributions to internal public goods? To what extent employees "spontaneously" (in the absence of explicit incentives) internalize the need of improving the organization?

In the present study, we design and execute a natural field experiment to address these questions. In particular, we compare two dominant perspectives on the problem. The first perspective borrows from the seminal work of Morgan (2000), who stresses that awarding prizes in lottery-like contests will mitigate or, in fact, eliminate the incentive to free riding. At the margin, employees choose their effort levels where expected gains and costs are equal. Inside organizations, such gains may include both the value of winning a personal reward and the value of improving the organization as well. Therefore, so long as employees care about their organization, even small prizes will boost participation be-

yond one would expect from the value of the prize alone. Another perspective is that workers have prosocial preferences towards the mission of the organization, as discussed by Besley and Ghatak (2005), Prendergast (2007), Delfgaauw and Dur (2008), among others. This is probably caused by a favorable process of self-selection into jobs by which employees routinely devote time and volunteer effort to perform tasks in the best interest of their organization.[1] In this case, awarding small prizes can be ineffective — it will only affect less motivated workers and lower opportunity costs to participate — or even have a negative impact. Competition may indeed evoke mixed feelings in workers with different competitive inclinations, a disposition that is potentially different for men and women (Croson and Gneezy, 2009); it may also inadvertently promote unethical behavior (Lazear, 1989; Charness et al., 2013); weaken intrinsic motivations (Reeve and Deci, 1996; Frey, 1997); and even kill the creativity of innovative work (Erat and Gneezy, 2015).

In this study, we contrast these two perspectives in the field. The natural field experiment was in collaboration with the Massachusetts General Hospital's Corrigan Minehan Heart Center ("Heart Center"), a prominent medical organization in the United States and a teaching hospital of the Harvard Medical School. The context of the experiment was an internal "innovation contest" aimed at improving the operations of the organization, in the spirit of open innovation initiatives as in Terwiesch and Xu (2008), Guinan et al. (2013), and Lakhani et al. (2013). The subject pool was the entire population of staff members with over 1,200 employees including physicians, nurses, and administrative workers. [This is great because xxxx as vastly known]

Our main intervention was in altering the content of personalized solicitations to participate in the innovation contest among randomly selected staff members. By doing so, we were able to get causal estimates of the effect of different incentives on two main outcomes: (a) the decision to submit and engage in implementation activities of a project proposal and (b) the quality of the proposal submitted, as measured by over 12,000 ratings made by peers. Using data on profession and gender, we were also able to characterize the heterogeneity of responses to the treatment, testing the implications of our intervention for the organization.

We designed four treatments related to our research questions. In a first treatment (PRIZE), we nudged employees to participate in the innovation contest by announcing a personal reward (an Apple iPad mini) for those having made the best proposals. In a second treatment (FUND), we nudged employees to participate by announcing the budget

[1]Everyday evidence for this type of behavior comes from blood donations, charitable giving, social workers. There is also a sizable scientific evidence on this topic that comes from a series of studies in the laboratory based on economic games (see Levitt and List, 2007, for a review) and field studies showing evidence on prosocial preferences at work (Bandiera et al., 2005; Della Vigna et al., 2016).

available for realizing the innovation (an explicit commitment by Hearth Center executives to pay expenses up to $20,000). In the remaining two treatments, we "framed" the solicitation by emphasizing the opportunity to improve the healthcare of their patients (PCARE) or their workplace (WPLACE).

There are various reasons why this environment represents a unique opportunity to study incentives for public goods inside organizations. First, workers had no prior experience with the innovation contest that marked the first time for the Heart Center and MGH. Thus, they learned about the opportunity primarily via our individualized communication strategy. Second, staff were asked to make individual submissions, thus helping us to exclude common reasons for collaboration such as peer pressure, peer monitoring, or being part of a team (Kandel and Lazear, 1992; Hamilton et al., 2003). Third, staff could anticipate additional costs and responsibilities from making a winning proposal (providing further guidance or a direct involvement in special implementation teams). These additional costs were not matched by the value of winning the prize, which was relatively small compared to the costs. However, the management had also expectations of high untapped value from frontline workers to innovate the workplace. [[NEED TO FIX THIS]]

Our empirical results show that workers were eager to step beyond the boundaries of their standard duties to contribute to improving the organization: 196 employees (16 percent) across the entire organization took part in the initiative, with 5 percent of our sample submitting project proposals.[2] About half of the proponents followed up with detailed implementation plans after an invitation by Heart Center executives, which culminated in two proposals receiving full funding.

Our main finding is that simply announcing a personal reward (PRIZE) produces participation rates about 85 percent higher than the average. This effect appears too large to fit a pure contest environment without public good incentives. The high incomes of our subject pool, the low chance of winning, and the additional costs of being selected for implementation suggest that very few would find it advantageous to participate. We discuss several possible explanations for this observation. We conclude that awarding a prize increased participation by helping motivated employees (i.e., with prosocial preferences) internalize the effects of their active participation for improving the organization, as in Morgan (2000). We then use a simple model to estimate employees' underlying so-

[2]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.

cial preferences, showing that these preferences can account for 30 percent of individual costs of submitting.

While we find effects on participation, we fail to find differences in the intensity of participation as measured by the count of projects proposals per proponent and in the distribution of the quality of proposals by treatments. The lack of quality effects could indicate there was no crowding-out effect, as the higher propensity of workers in the prize treatment does not seem to be driven by low-quality submissions. We also find little evidence of difference in content other than peer-evaluated quality, such as the areas of focus and length of the proposals. Overall, these findings suggest no negative trade-off between quantity and quality: treatments that attracted more participants 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 alsof find that responses are sensitive to the gender of the solicited person: First, women's participation was greater with a prize (PRIZE) than men's, although not significantly so. Second, women's participation was greater when emphasizing the patient care mission (PCARE) whereas men's participation was significantly lower, controlling for the profession. The first finding ensures that [[ADD EXPLANATION]] differences in competitive inclinations are not strong. The other finding suggests that female workers may perceive the mission of the organization differently than male workers. Women may derive higher utility from contributing to the mission of the workplace, i.e., helping patients. Another reasonable explanation is that gender-based self-stereotypes have affected individual decisions to contribute proposals, as Coffman (2014).

Finally, announcing a commitment of $20,000 in project funding (FUND) made employees less likely to submit compared to all other treatments. This evidence confirms that the being in charge of the project was not perceived as a prize in itself, and could not be the predominant motivation to participate and submit proposals. We do not view the commitment per se as having a negative effect. However, it is also possible that such a commitment provided workers with a reference point on which they formed their beliefs resulting in an increased incentive to free-ride. Both explanations are consistent with our view on the role of awarding immediate personal rewards to spur the provision of public goods inside organizations.

The present study contributes to the literature on the use of relative incentives (prizes) in the workplace, which finds its theoretical foundations in 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 activities that produce organizational improvements in the field. 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,[3] 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, 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); see Vesterlund (2012) for an excellent literature review. Our work provides further support to this theory. Furthermore, we generalize the results of past studies to an organizational setting, where individual contributions are non-monetary but consists of time and effort in putting forward (and implementing) a proposal whereas the public good consists of the potential improvements for the organization derived by implementing the innovation project. Under such circumstances, we find that prize mechanisms spur more participation compared to voluntary mechanisms and appear overall more profitable for the organization.[4]

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. Everyday evidence for this type of behavior comes from blood donations, charitable giving, social workers.[5] There is also a sizable scientific evidence on this topic that comes from a series of studies in the laboratory based on economic games (see Levitt and List, 2007, for a review). However, there are only a few field studies providing evidence supporting the same effects of prosocial preferences at work (Bandiera et al., 2005; Della Vigna et al., 2016). Bandiera et al. (2005) shows that workers internalize preferences of co-workers and may reduce effort under relative incentives. Della Vigna

[3]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.

[4]By contrast, the existing evidence on profitability of contest-type mechanisms for raising money for other public goods, such as charity, is mixed (Vesterlund, 2012).

[5]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 2 percent of gross domestic product for the United States. Lacetera et al. (2014) further reports that: "27% of Americans volunteer with formal organizations, for a total of about 8 billion hours per year."

et al. (2016) studies social preferences in the field by hiring freelance workers to fold envelopes for a charity and shows a positive effect of mission-oriented preferences, also called "vertical social preferences,"[6] on the level of effort at work. 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.[7]

This study also contributes to a growing literature exploring the role of incentives and competition in creative works (Gibbs et al., 2015; Erat and Gneezy, 2015). Within this literature, Gibbs et al. (2015) is perhaps the study most closely related to ours. In a consulting firm, Gibbs et al. (2015) finds that piece rate incentives increase participation and the quality of workers' ideas compared to flat wages. In that setting, however, managers seek ideas that would "directly benefit the client, as compared to ideas that improve internal processes." Hence, an employee's idea is successful only when it is adopted by the client and has no direct impact on the organization. Our study is complementary to Gibbs et al. (2015) as we consider a competition (with and without personal rewards) instead of piece rates. The main novelty is that we examine an innovation task directed at improving the organization and its "internal processes." So, workers have an incentive to free ride not present in Gibbs et al. (2015).[8] Overall, personal rewards have a positive effect on participation but we find no (negative) quality effects. We argue this could result from the competition stopping workers from submitting innovation projects below a certain level of quality.

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 prosocial motivations to raise individual productivity; in the current study, we use "framing" to make particular motivations salient. We find that emphasizing prosocial

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

[7]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").

[8]Another difference is that Gibbs et al. (2015) studies teams instead of individual workers. This implies that also in Gibbs et al. (2015) there may be a problem of free-riding within teams. Yet, formal teams are believed to solve internally the free-riding problem through peer pressure, monitoring, reciprocity within the team, etc. Hamilton et al. (2003) finds poor empirical evidence of free-riding in teams. In this sense, by focusing on individual workers, our results center on the problem of free-riding isolated from team dynamics.

motivations has countervailing effects on participation; negative for men and positive for women.[9] 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.

2 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 also 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).

[9] 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.

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 effort 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).

3 Experimental Design

3.1 The context

Our natural field experiment was conducted at the Heart Center from July 2014 to November 2014. 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 Healthcare Transformation Lab (HTL), an initiative aimed at developing innovative health care process improvements to enhance the health care safety and delivery of the hospital.[10] 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 anesthetic 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.

The innovation contest can be 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.

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

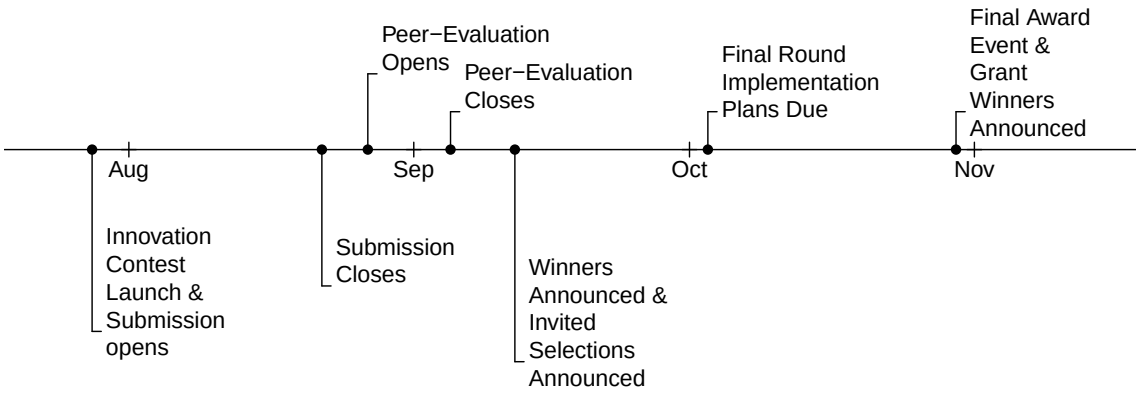


Figure 1: Timeline of the innovation contest

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 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.

3.2 The design

The main intervention was to alter the content of the communication that announced the innovation contest. The start of the submission phase was indeed announced to all staff members 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 body 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,
>
> [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: FUND, PRIZE, PCARE and WPLACE. Thus, creating four treatment groups of equal size (Table 3).

In the first two groups (FUND and PRIZE), the solicitation nudged employees to participate by announcing individual prizes to be won (PRIZE), i.e., an Apple iPad mini, or a $20,000 budget for developing their project proposals (FUND). In the remaining two groups (PCARE and WPLACE), the solicitation "framed" the contest as an opportunity to improve the healthcare of their patients (PCARE) or the workplace (WPLACE). The exact words are in Table 3. In all groups, employees were not told that they were part of an experiment.

	Obs.	%	First paragraph
PRIZE	312	25	Submit your ideas to win an Apple iPad mini
[1.8ex] FUND	308	25	Submit your ideas to win project funding up to $20,000 to turn your id

	Obs.	%	First paragraph
[1.8ex] PCARE	310	25	Submit your ideas to improve patient care at the Heart Center
[1.8ex] WPLACE	307	25	Submit your ideas to improve the workplace at the Heart Center
[1.8ex] Total	1237	100	NA

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.[11]

Also, note the lack of a traditional "control" treatment in this study.[12] Indeed, the analysis focused on multiple comparisons of several unordered discrete treatments (e.g., prizes vs funding vs framing).[13]

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 3).

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. To further create a "safe" environment for employees submitting proposals. We made 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

[11]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.

[12]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.

[13]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.

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.

4 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 **??** presents summary statistics showing that the variables in the four treatment groups were statistically balanced.

	FUND	PCARE	WPLACE	PRIZE	%	Obs.	P-value
Other	29.9	29.5	26.1	31.6	29.3	362	0.844
MD/Fellow	19.0	18.2	17.9	17.9	18.3	226	NA
Nursing	51.1	52.3	55.5	51.0	52.5	649	NA
[1.86ex] Female	68.8	70.0	75.3	75.2	71.9	890	0.159
Male	31.0	30.4	24.2	26.0	28.1	347	NA
[1.86ex] No office	49.7	45.9	47.1	45.0	46.6	577	0.556
Office	50.0	54.2	52.3	55.8	53.4	660	NA
[1.86ex] 18-25 years old*	5.7	8.0	8.0	5.7	6.5	24	1.000
26-35 years old*	29.0	29.0	31.2	25.8	29.0	107	NA
36-45 years old*	18.3	19.2	24.0	16.3	22.0	81	NA
>45 years old*	44.0	46.4	51.2	45.2	42.5	157	NA
[1.86ex] < 10 years tenure*	39.8	31.2	35.6	36.9	35.8	132	0.891
10-20 years tenure*	25.8	28.8	38.1	28.4	30.1	111	NA
20-30 years tenure*	11.5	19.0	14.8	9.7	13.6	50	NA
30-40 years tenure*	9.5	15.9	15.1	11.5	13.0	48	NA
>40 years tenure*	10.2	4.3	7.7	8.3	7.6	28	NA

	FUND	PCARE	WPLACE	PRIZE	%	Obs.	P-value

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).

5 Results

5.1 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").

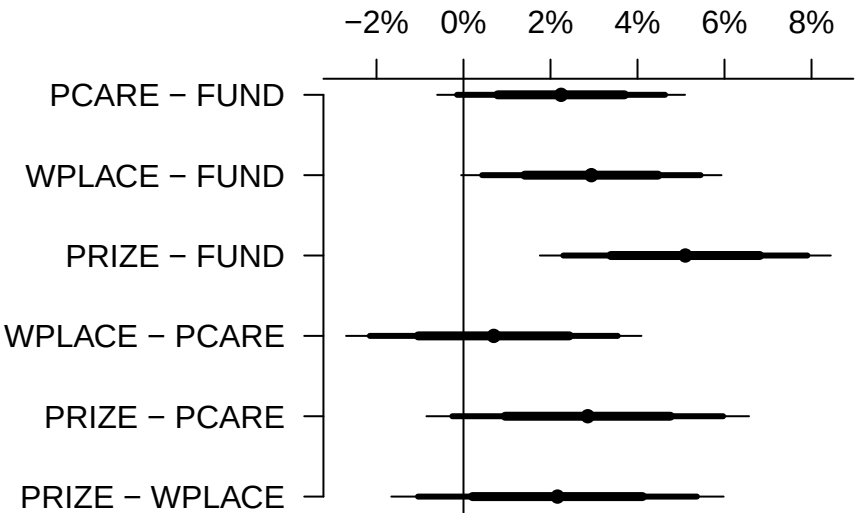
A pairwise comparison of the probability of submitting project proposals (Figure 2) reveals that employees in the PRIZE treatment were significantly more likely (5 percentage points) to submit than those in the FUND treatment. We also find a significant positive difference (about 3 percentage points) 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

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

same results (see the Appendix).[14]

Figure 2: Difference in the probability of submitting project proposals



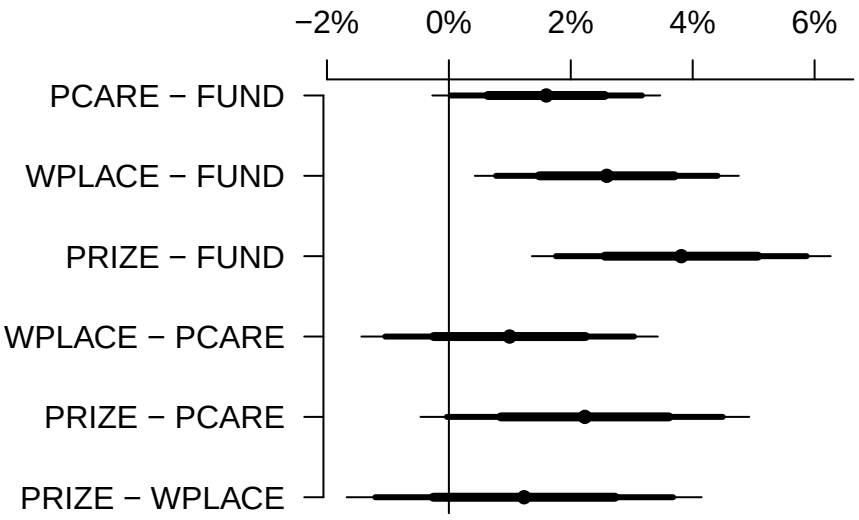
Notes: This figure plots the point estimates of the difference plus ± 1, ± 1.6, and ± 2 standard errors. Bootstrap resampling and confidence intervals based on the more conservative Holm-Bonferroni method yields very similar results (see the Appendix).

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

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

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.

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

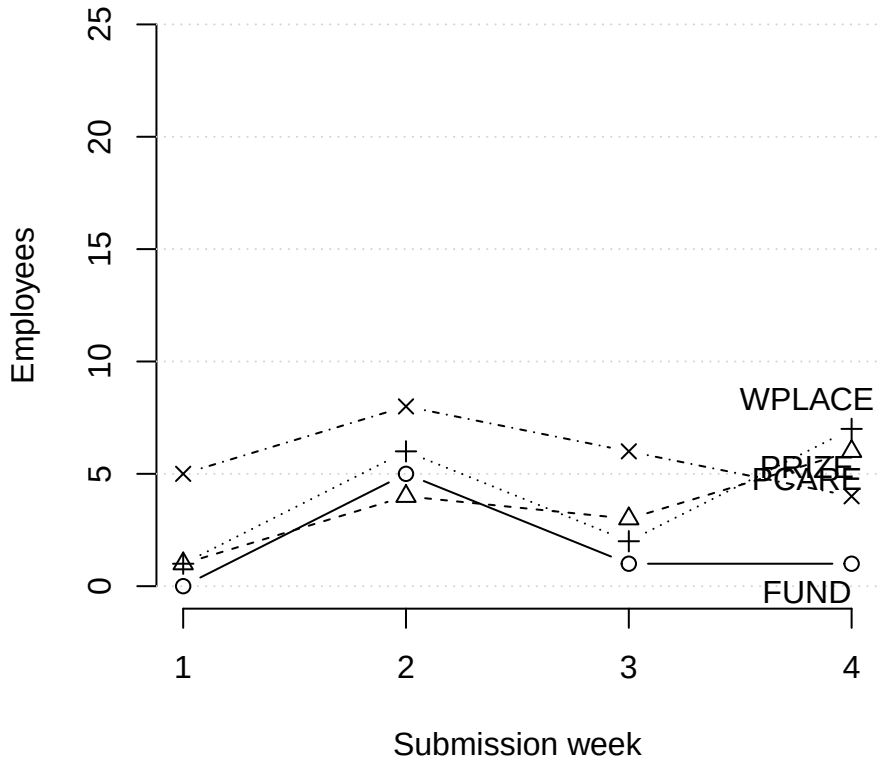


Notes: This figure plots the point estimates of the difference plus ± 1, ± 1.6, and ± 2 standard errors. Estimates have been adjusted for the small counts of finalists (Agresti and Caffo, 2000) resulting in more conservative confidence intervals. Bootstrap resampling and confidence intervals based on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

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 6. 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 incentives would be the same in each condition and participation rates should 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 all weeks but the last, where participation in the WPLACE and PCARE had a little boost. Thus, if anything, contamination occurred at the very end of the competition. And even so, it might only have biased downwards (instead of inflating) the estimated positive effect of prizes on participation, which would only

confirm our view of prizes being quite effective.

Figure 4: The dynamic of submissions



Notes: This figure plots the percentage of staff submitting proposals over the four weeks of the submission period in each condition.

Although the contest was "unbiased" in the sense that it sought to engage employees at all levels of the organization, one may anticipate participation rates to vary across staff due to differences in individual costs of participation (as our Hypothesis H5 in Section 2). To study this hypothesis, we model the conditional probability of submitting proposals as

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

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.[15]

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.[16] 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.[17] 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 working having a fixed office location, as opposed to being assigned to a ward. In our context, having a 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.

[15]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.

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

[17]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.

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.

5.1.1 Interactions

We now turn to examine treatment interactions involving the employee's gender and profession.[18] 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.[19]

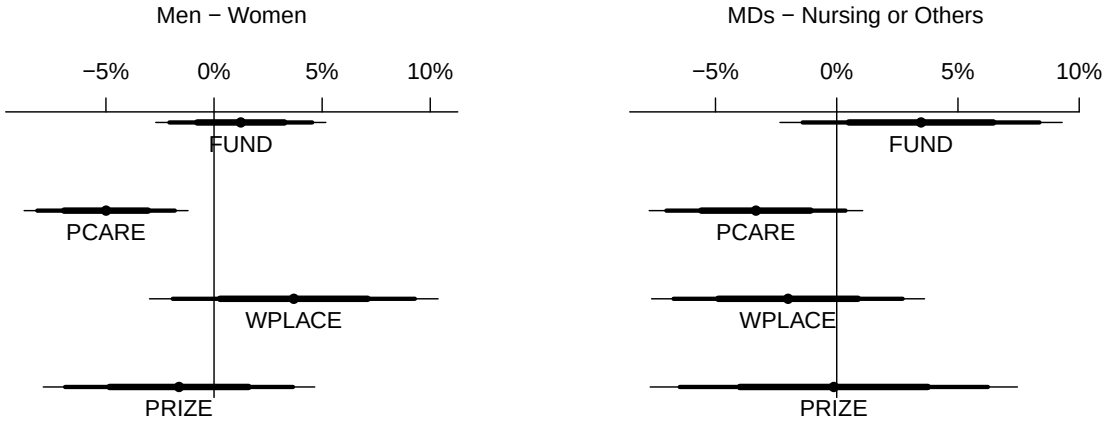
To isolate gender and profession effects, we now employ a version of model (6) with gender-treatment interactions.[20] Estimates are shown in Table 5. After gradually adding profession and office controls, interaction coefficients remain pretty stable across all specifications. The response of men under PCARE is about 3 times the magnitude and in the opposite direction of the women's response. By subtracting these two coefficients, we find a significant difference between men and women of about 5 percentage points ($p = .018$),

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

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

[20] 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.

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.

Dynamics. Fix problem with "interference"

One possible problem concerning the above regression analysis is the relatively small number of responses per treatment compared to the sample size (response rates below 10 percent are usually seen as rare events). The main problem is that asymptotic confidence intervals may not be fully accurate [@king2001logistic]. Logistic regression models allow testing this issue by direct methods that deal with rare events, such as exact inference. Using logistic regression, we find the same results indicating these are robust under exact inference (tables available on request).

5.2 Rating project proposals

The project proposals were then rated by 178 employees (14 percent of our sample), with the evaluators rating a median of 65.5 out of 113 project proposals (58 percent) yielding a total of 12,055 evaluator-proposal pairs.[21] 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, using a Fisher's Exact Test for Count Data, we find no statistically significant (p=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

[21] 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.

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.

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. In terms of statistical inference, 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 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. Nevertheless, the difference in the probability of staff rating proposals between the PRIZE and the WPLACE or the PCARE treatments being negative might also suggest a slight (although not significant) motivation crowding-out effect.

Table 6: Outcomes of the peer evaluation phase

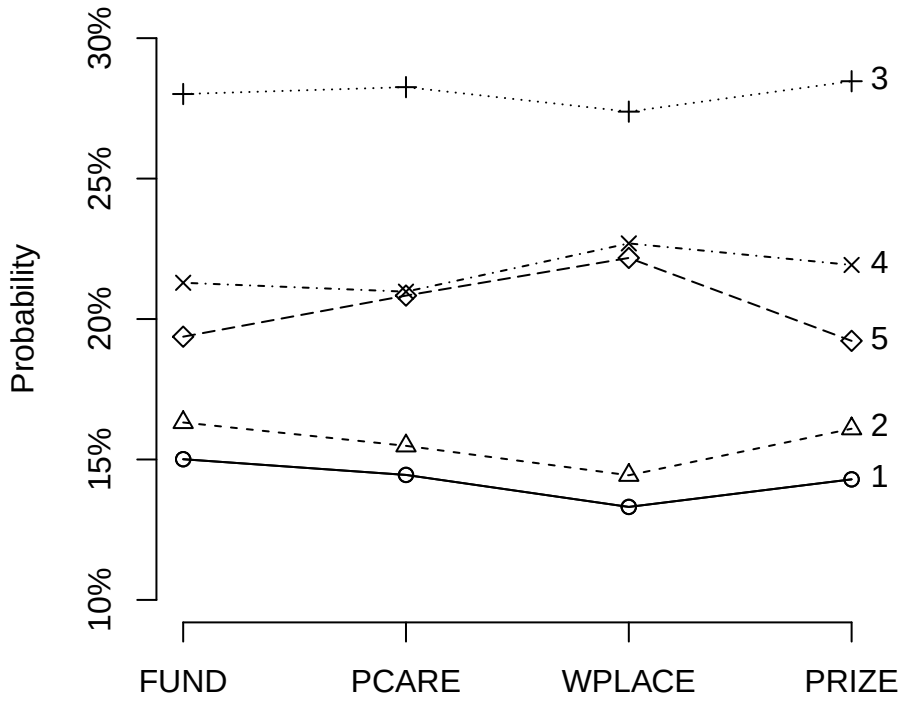
	Rating proposals:			*Rated proposals:*		
	No	Yes	% yes	Total	Mean	Median
PRIZE	269	43	13.8	2457	57.1	43
FUND	272	36	11.7	2484	69.0	68
PCARE	261	49	15.8	3413	69.7	73
WPLACE	257	50	16.3	3701	74.0	82
Total	1059	178	14.4	12055	67.7	66

5.3 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: 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.

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).

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.[22]

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). And 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 intention to

[22]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.

not 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 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.

5.4 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 . Another was to develop a smartphone application that . However, 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.

"xxxxx"

"xxxx"

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.

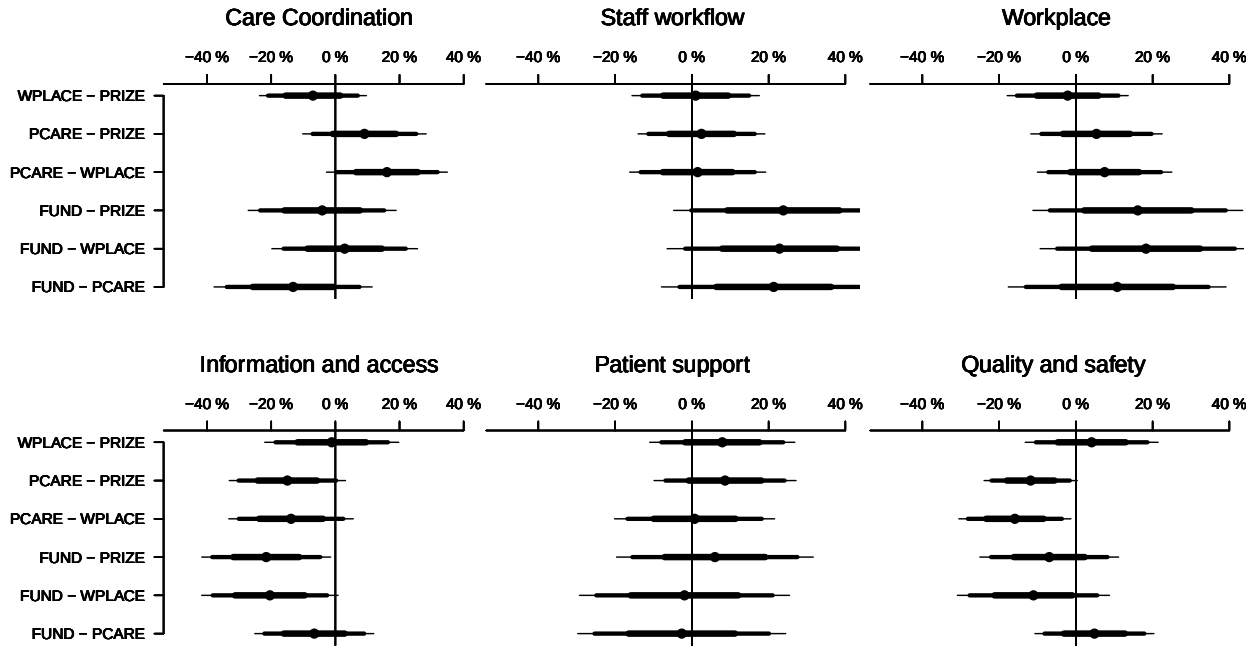
Using a Fisher's Exact Test for Count Data with simulated p-value (based on 50000 replicates), we find a mildly significant (p=0.089) association between these categories and the treatments.[23] The analysis of pairwise differences between treatments (Figure 7) reveals that this result is driven by differences 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 on how our intervention will affect the content of proposals. One possibility is that the framing intervention induced employees to concentrate on different categories. For example, while staff in the WPLACE treatment might have focused on improvements for the workplace, those in the PCARE treatment might have concentrated on interventions directly targeting the patients. However, our data do not provide any evidence consistent with this story.

Another possibility is that of differences in the underlying complexity of the project proposal that was not captured by our measure of quality. To address this issue, we now turn to examine differences in the length of a proposal as measured by the word count of a submission. Submissions are 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).

Hence, 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.

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

Figure 7: Differences in the probability of proposals being in a given area of focus in each treatment



5.5 Estimating social preferences

In this section, we calibrate the theoretical model developed in Section 2 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,[24] 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,[25]

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

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

[25]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.

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

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

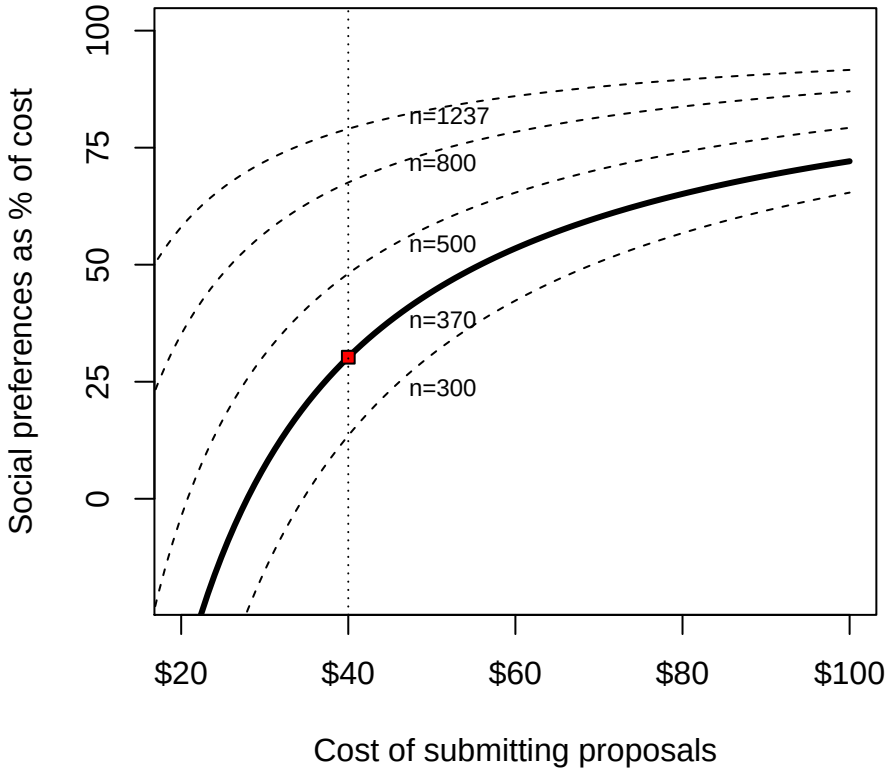
This implies that the parameter capturing individual preference for the public good (that is consistent with our data) must be proportional to the ratio between the difference in rewards and the difference in the probability of submitting. Although we do not observe the levels of R in each treatment, 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).[26] We further calibrate the cost of submitting a proposal c to $40 which is the median income per hour of a Nurse Practitioner according to the Bureau of Labor Statistics; we assume the number of competitors n to be 30 percent of the entire sample to take into account rational expectations about the actual number of participants in the contest.[27] Finally, by substituting these calibrated values into equation (8) along with the empirical difference in participation rates between the PRIZE and the other treatments ($\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,

[26] 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.

[27] 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 ($\hat{\gamma}$)



another important source of uncertainty is regarding the main behavioral assumptions of the model, as we discuss in the next section.

6 Discussion

In this article, we report the results of a natural field experiment conducted during an innovation contest held at the MGH Heart Center. Different incentives for participation were announced by manipulating the personal messaging to employees. Messaging emphasized either available prizes (iPad mini's) for top submissions, motives for improving patient care and improving the workplace, or a funding opportunity alone. From the results, we find that: simply announcing a prize for top contributions significantly increases participation rates by 85 percent compared to the other treatments, announcing a funding commitment significantly reduces participation, and these incentive effects vary by gender. Using a simple model of public good contribution, we show that about 30 percent of the magnitude of the positive effect of prizes on participation can be explained by underlying social preferences of workers. We also find that the announcements appear to have no effect on the quality or content of submissions.

The increase in participation due to the announcement of available prizes is consistent with our hypotheses (H1) and (H2), implying a positive relationship between the value of the prize and the propensity to perform a task that is improving the organization. However, if the increase was due to a calculation of the odds of winning the prize alone, then one would expect the worker's level of income to affect the extent of this effect since richer workers would be less incentivized by the opportunity of winning the prize. Our results show that this is not the case. Despite the presence of large differences in income by profession, participation is not significantly affected by the profession of the worker. This evidence indicates that the extent of the increase may be too large to fit a pure contest environment without public good incentives (e.g., Bull et al., 1987; Ehrenberg and Bognanno, 1990; Knoeber and Thurman, 1994).

One objection to this interpretation is that workers might have perceived economic gains that exceeded the value of the prize without being necessarily related to their desire to contribute to the organizational public good. Workers would volunteer and perform unpaid tasks when there are career incentives (Baker et al., 1994; Gibbs, 1995) or an opportunity for acquiring higher social status, such as winning symbolic awards either at work (Kosfeld and Neckermann, 2011; Blanes i Vidal and Nossol, 2011) or in non-profit endeavors such as Wikipedia (Gallus et al., 2015). We acknowledge that these additional incentives might be at work, however, these are unlikely to explain our results because

the opportunity for a career advancement or a recognition lied in winning the contest, the attention derived from it, and the interactions with the executives in the implementation phase. Indeed, the employees having made the best proposals were publicly recognized as winners both online (on the MGH website) and in person in a public event held at the end of the innovation contest. And it would be hard to imagine any of our treatments to create any additional effect on reputation compared to such public recognition.

Another concern is that decisions could be influenced by word-of-mouth communication among workers. In particular, the exchanging information about the personalized solicitation may create a problem of "contamination" or interference between experimental units (Rosenbaum, 2007). Perhaps, the most problematic situation occurs when employees who did not pay attention to the solicitation at the launch of the contest might become interested later because they have been told about the opportunity of winning a prize. We are not very much worried about this problem for several reasons. First, XXXX. Second, xxxx. Third, xxxx.

Moreover, by surveying HTL staff a few months after the contest, we find no examples of interference (such as workers noticing or complaining about the difference in the content of the messages received).

Another finding that deserves further consideration is the negative effect on participation rates of announcing a precise commitment by the management. Our interpretation is that workers viewed the commitment as a "matching gift," a conditional commitment by the management to "match" the efforts of more ambitious projects, up to a maximum amount of $20,000. In the public goods context, Karlan and List (2007) finds a positive effect of matching gifts on participation and contributions to charitable giving. Here, instead, we find a negative effect on participation. One possible explanation is that the announcement of a large commitment produced a reference point on which workers built their expectations about the additional cost of effort conditional on winning, thus exacerbating the incentive to free ride. According to this interpretation, this finding indicates that interventions that resemble a "matching gift" may not work well in an organizational environment.[^We also rule out that employees viewed the commitment in the FUND treatment as "cheap," due to prior expectations of a higher commitment. As it was the first iteration of the challenge, no specific expectation from prior experience could have existed.]

Concerning the observed differences based on gender, the existing literature in economics indicates that women tend to have greater risk-aversion (Borghans et al., 2009) and a distaste for competition compared to men (Niederle and Vesterlund, 2007). So, one may be concerned that using a contest incentive structure to encourage the private

provision of a public good would discourage participation by women. However, in our experiment women participated most in the treatment emphasizing the contest prize, suggesting that this type of contest is robust to these effects. One possible explanation is that women have a higher fixed cost of competing than men, thus a competition for prizes may subsidize the higher cost of entry for women (compared to the same competition with only non-pecuniary rewards).

The difference associated with the patient care mission of the organization is consistent with several interpretations. One possibility is that female workers may perceive the mission of the organization differently than male workers. Women may derive higher utility or an inner satisfaction from contributing project ideas to improve the healthcare of their patients. But this finding is also consistent with an effect of gender stereotypes in the workplace that may affect individual decisions to contribute with a submission (see Evans, 2002, for examples of gender-based stereotypes in healthcare organizations). In this sense, one interpretation is that our study supports the findings of Coffman (2014). Coffman (2014) finds evidence in the laboratory that women are less likely to contribute ideas to groups when the topic falls in male-typed domains, e.g., sports, and vice versa. This result also echoes Delfgaauw et al. (2013) showing evidence, in a sales contest, of differential effects on effort due to the interaction between the employee's gender and the way a competition is announced or "framed."

While our results indicate strong treatment effects on participation, we do not find effects on the distribution of content, quality, and amount of proposals. In particular, the lack of quality effects holds with both peer voting and the projects selected by the management. This result is consistent with our hypothesis that every worker has the same probability of contributing a high quality project idea, and that increasing participation helps to raise the levels of the public good (hypothesis H4). This result is in apparent contrast with other studies on public goods in the field, such as Landry et al. (2006) that finds effects of prizes on both the amount of contributions and participation (intensive and extensive margins) in a natural field experiment on charitable giving. One problem in interpreting this incongruity is that we do not observe an accurate measure of the hours worked or the effort spent on each submission, and we do not observe the production function for quality. Thus, it is hard to tell whether the lack of quality differences is due to differences in contributions or, rather, it was a feature of a non-monotonic production function.

Concerning the role of prize incentives in creative endeavors, Gibbs et al. (2015) finds effects in both the number and quality of submissions. The contexts of the experiment were quite different. In Gibbs et al. (2015), the authors study a two-year,

pay-for-performance type of compensation scheme to encourage proposals for process improvements for clients of a consulting frame, whereas we focus on a four-week internal innovation contest that rewards best project proposals to improve the organization. Additionally, one goal of the innovation contest was to maximize engagement. To that end, messaging emphasized contributing project proposals of any scope. So, one may argue that the difference in the incentives could explain the no quality effect. However, the short time frame and nature of the contest do not give much scope for investing in proposal quality versus the environment in Gibbs et al. (2015).

Finally, one key concern when combining a pecuniary incentive with contributions to a public good is the interaction of incentives with the motive to contribute. A *negative interaction* occurs when incentives crowd-out the motive to contribute. Crowding out effects have been seen in other experiments in the context of blood donations (Lacetera et al., 2013, 2014), or the crowding in effects seen with public "bads" in daycare pick-ups (Gneezy and Rustichini, 2000). We do not observe evidence of a crowding-out in our environment, but differences in context make comparisons difficult. 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).

In conclusion, the results presented here have implications that go beyond the specific organization under study. Using a competition for prizes appears a profitable way for firms to encourage contributions among workers in situations resembling a private provision of public goods. In many settings, this approach can be more effective than what is acknowledged in the traditional tournament theory literature. The reason being that the incentive effect of prizes interacts with prosocial motivations of workers to exert effort. The management can appeal to internal motivations towards the mission of the organization to raise the level of voluntary contributions. However, this may be tricky in practice due to the heterogeneity of motivated agents. In particular, the evidence on gender differences in response to framing suggests female workers may perceive the goals of the organization differently from male workers. Investigating the causes of these gender-based differences is left as an avenue for future research.

References

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

Akerlof, George A and Rachel E Kranton, "Identity and the Economics of Organizations," *Journal of Economic Perspectives*, 2005, *19* (1), 9–32.

Andreoni, James, "Warm-Glow versus Cold-Prickle: The Effects of Positive and Negative Framing on Cooperation in Experiments," *The Quarterly Journal of Economics*, 1995, *110* (1), 1–21.

Baker, George, Michael Gibbs, and Bengt Holmstrom, "The Internal Economics of the Firm: Evidence from Personnel Data," *The Quarterly Journal of Economics*, 1994, *109* (4), 881–919.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul, "Social Preferences and the Response to Incentives: Evidence from Personnel Data," *The Quarterly Journal of Economics*, 2005, *120* (3), 917–962.

Besley, Timothy and Maitreesh Ghatak, "Competition and Incentives with Motivated Agents," *The American Economic Review*, 2005, *95* (3), 616–636.

Borghans, Lex, James J Heckman, Bart HH Golsteyn, and Huub Meijers, "Gender differences in risk aversion and ambiguity aversion," *Journal of the European Economic Association*, 2009, *7* (2-3), 649–658.

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

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

Bull, Clive, Andrew Schotter, and Keith Weigelt, "Tournaments and Piece Rates: An Experimental Study," *Journal of Political Economy*, 1987, *95* (1), 1–33.

Charness, Gary, David Masclet, and Marie Claire Villeval, "The dark side of competition for status," *Management Science*, 2013, *60* (1), 38–55.

Coffman, Katherine Baldiga, "Evidence on Self-Stereotyping and the Contribution of Ideas," *The Quarterly Journal of Economics*, 2014, *129* (4), 1625–1660.

Cohen, Jacob, "A power primer.," *Psychological bulletin*, 1992, *112* (1), 155.

Croson, Rachel and Uri Gneezy, "Gender Differences in Preferences," *Journal of Economic Literature*, 2009, *47* (2), 448–474.

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

Delfgaauw, Josse, "Dedicated Doctors: Public and Private Provision of Health Care with Altruistic Physicians," *New England Journal of Medicine*, 2005, *353* (15).

_ **and Robert Dur**, "Incentives and Workers' Motivation in the Public Sector*," *The Economic Journal*, 2008, *118* (525), 171–191.

_ **,** _ **, Joeri Sol, and Willem Verbeke**, "Tournament incentives in the field: Gender differences in the workplace," *Journal of Labor Economics*, 2013, *31* (2), 305–326.

Ehrenberg, Ronald G and Michael L Bognanno, "Do Tournaments Have Incentive Effects?," *The Journal of Political Economy*, 1990, *98* (6), 1307–1324.

Erat, Sanjiv and Uri Gneezy, "Incentives for creativity," *Experimental Economics*, 2015, pp. 1–12.

Eriksson, Tor, "Executive Compensation and Tournament Theory: Empirical Tests on Danish Data," *Journal of Labor Economics*, 1999, *17* (2), 262–280.

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

Fehr, Ernst, Georg Kirchsteiger, and Arno Riedl, "Gift exchange and reciprocity in competitive experimental markets," *European Economic Review*, 1998, *42* (1), 1–34.

Frey, Bruno S, "Not just for the money: An Economic Theory of Personal Motivation," *Books*, 1997.

Gallus, Jana et al., "Fostering Voluntary Contributions to a Public Good A Large-Scale Natural Field Experiment at Wikipedia," Technical Report, Center for Research in Economics, Management and the Arts (CREMA) 2015.

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

_ **, Susanne Neckermann, and Christoph Siemroth**, "A field experiment in motivating employee ideas," *Review of Economics and Statistics, Forthcoming*, 2015.

Gneezy, Uri and Aldo Rustichini, "A fine is a price," *Journal of Legal Studies*, 2000, *29*, 1.

Green, Jerry R and Nancy L Stokey, "A Comparison of Tournaments and Contracts," *The Journal of Political Economy*, 1983, *91* (3), 349–364.

Guinan, Eva, Kevin J Boudreau, and Karim R Lakhani, "Experiments in open innovation at Harvard Medical School," *MIT Sloan Management Review*, 2013, *54* (3), 45.

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

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

Holmstrom, Bengt and Paul Milgrom, "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics and Organization*, 1991, *7*, 24–52.

Hong, Fuhai, Tanjim Hossain, and John A List, "Framing manipulations in contests: a natural field experiment," *Journal of Economic Behavior & Organization*, 2015.

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

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

Kandel, Eugene and Edward P Lazear, "Peer pressure and partnerships," *Journal of political Economy*, 1992, *100* (4), 801–817.

Karlan, Dean and John List, "Does Price Matter in Charitable Giving? Evidence from a Large-Scale Natural Field Experiment," *The American Economic Review*, 2007, *97* (5).

Knoeber, Charles R and Walter N Thurman, "Testing the Theory of Tournaments: An Empirical Analysis of Broiler Production," *Journal of Labor Economics*, 1994, *12* (2), 155–79.

Kosfeld, Michael and Susanne Neckermann, "Getting More Work for Nothing? Symbolic Awards and Worker Performance," *American Economic Journal: Microeconomics*, 2011, *3* (3), 86–99.

Kube, Sebastian, Michel André Maréchal, and Clemens Puppe, "The Currency of Reciprocity: Gift Exchange in the Workplace," *The American Economic Review*, 2012, *102* (4), 1644–1662.

Lacetera, Nicola, Mario Macis, and Robert Slonim, "Economic rewards to motivate blood donations," *Science*, 2013, *340* (6135), 927–928.

_ , _ , and _ , "Rewarding volunteers: a field experiment," *Management Science*, 2014, *60* (5), 1107–1129.

Lakhani, Karim R, Kevin J Boudreau, Po-Ru Loh, Lars Backstrom, Carliss Baldwin, Eric Lonstein, Mike Lydon, Alan MacCormack, Ramy A Arnaout, and Eva C Guinan, "Prize-based contests can provide solutions to computational biology problems," *Nature biotechnology*, 2013, *31* (2), 108–111.

Landry, Craig E, Andreas Lange, John A List, Michael K Price, and Nicholas G Rupp, "Toward an Understanding of the Economics of Charity: Evidence from a Field Experiment*," *The Quarterly journal of economics*, 2006, *121* (2), 747–782.

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

Lazear, Edward P, "Pay Equality and Industrial Politics," *Journal of Political Economy*, 1989, *97* (3), 561–80.

_ and Sherwin Rosen, "Rank-Order Tournaments as Optimum Labor Contracts," *The Journal of Political Economy*, 1981, *89* (5), 841–864.

Levitt, Steven D and John A List, "What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?," *The Journal of Economic Perspectives*, 2007, *21* (2), 153–174.

List, John A, "The Market for Charitable Giving," *The Journal of Economic Perspectives*, 2011, *25* (2), 157.

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

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

Mary, O'Keeffe, W Kip Viscusi, and Richard J Zeckhauser, "Economic Contests: Comparative Reward Schemes," *Journal of Labor Economics*, 1984, *2* (1), 27–56.

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

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

Niederle, Muriel and Lise Vesterlund, "Do Women Shy Away From Competition? Do Men Compete Too Much?," *The Quarterly Journal of Economics*, 2007, *122* (3), 1067–1101.

Prendergast, Canice, "The Motivation and Bias of Bureaucrats," *The American Economic Review*, 2007, *97* (1), 180–196.

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

Rosenbaum, Paul R, "Interference between units in randomized experiments," *Journal of the American Statistical Association*, 2007, *102* (477).

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

_ **and Yi Xu**, "Innovation contests, open innovation, and multiagent problem solving," *Management science*, 2008, *54* (9), 1529–1543.

Vesterlund, Lise, "Voluntary giving to public goods: moving beyond the linear VCM," in John H Kagel and Alvin E Roth, eds., *The Handbook of Experimental Economics*, Vol. 2, in preparation, Princeton University Press, 2012.

Vigna, Stefano Della, John A List, Ulrike Malmendier, and Gautam Rao, "Estimating Social Preferences and Gift Exchange at Work," Technical Report, National Bureau of Economic Research 2016.