Incentives for Public Goods Inside Organizations: Field Experimental Evidence[[1]](#footnote-21)

Andrea Blasco

Olivia S. Jung

Karim R. Lakhani

Michael Menietti

Last updated: 03 March, 2018

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

JEL Classification: D23; H41; M52.

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

# Introduction

Workers employed by firms are often expected to make contributions that go beyond an effective and efficient execution of their own tasks. Such contributions are essential to generate positive externalities for their co-workers, supervisors, and customers; but could be hardly spurred via formal evaluation and reward systems alone. This situation presents thus employees with the recurrent dilemma of taking actions for the benefit of others or behaving opportunistically hoping for the other members of the organization to step in. Given the importance of this kind of decisions for the functioning of organizations, in this article we ask a basic question: What motivates employees to self-select into unpaid tasks and volunteer work for the benefit of the organization? And how managers should design environments that foster such employee behaviors?

An extensive literature in economics points to volunteering at work as a way to receive extrinsic benefits, such as career advancements or a positive reputation.[[2]](#footnote-27) But there is also a growing literature that links this kind of volunteering to a form of *altruism* that goes either from the worker to the other members of the organization (Bandiera, Barankay, and Rasul 2005; Rotemberg 2006) or from the worker to customers/clients who might benefit from the activities of the organization (Delfgaauw 2005; Delfgaauw and Dur 2008; Prendergast 2007). This second view is usually predominant in organizations in which workers may feel a sense of duty that acts as a motivation (Akerlof and Kranton 2005; Besley and Ghatak 2005), like those producing collective goods (e.g., hospitals, schools, government agencies). However, the lack of systematic empirical evidence on employee behavior makes difficult to assess the strength of these different motivations and predict in which direction these might affect the dilemma faced by workers inside organizations.[[3]](#footnote-28)

In this study, one primary goal is to examine how employees respond to a public good dilemma in a field setting where both extrinsic and intrinsic motives are likely to be present. In particular, we examine empirically how over 1,200 staff members (doctors, nurses, and administrative staff) of an elite medical organization in the United States respond to an organization-wide call to submit proposals for improving the organization to an *internal innovation contest*. Such innovation contest solicited employees to submit project proposals describing an existing problem and providing a solution to address the problem, in the spirit of “open innovation” contests discussed in Terwiesch and Xu (2008); Lakhani et al. (2013); and Glaeser et al. (2016). The entire staff was then invited to read and rate each proposal on a five-point scale. Based on these ratings and other criteria, the winning proposal would receive funding for implementation, implying additional costs and responsibilities from making a winning proposal (e.g., providing further guidance or a direct involvement in implementation).

One reason to focus on employee behaviors associated with an innovation contest, rather than some other kind of incentive scheme, is practical: such competitions are easy to run and very common to most organizations.[[4]](#footnote-29) From another perspective, it is relatively inexpensive for the organization (and the researchers) to alter the communications around the contest in a way to separately identify channels associated with extrinsic and intrinsic reasons for participating. For example, a personalized messaging strategy may either convey information about the presence of personal awards for the winners, or omit that information. In this sense, a contest represents a very convenient setting for experimental interventions to study how employees self-select themselves into extra tasks that generate public good effects for the organization while they are running the normal operations.

Another key element of our study is the focus on the health care delivery context, which is important for two main reasons. First, the need for organizational improvement and innovation is vastly noted (Cutler, Wikler, and Basch 2012). Second, health care professionals are commonly seen as willing to step beyond the boundaries of their contractual duties to offer better care (Delfgaauw 2005), which makes the comparison of different incentives towards a public good especially relevant and interesting.

Within this empirical context, we conduct a natural field experiment that isolates and provides evidence on four possible incentives that underlie employee participation in the contest: (1) the presence of financial awards for the winners; (2) the opportunity of funding most valuable contributions; (3) the potential positive impact on the other workers; and (4) the potential positive impact on their patients. Identification comes from a personalized email marketing campaign with four different *solicitation treatments,* which were assigned at random to each staff member to isolate one of the above four different motives.

This experimental intervention allows us to obtain causal estimates of the effect of different solicitation strategies on two main outcomes: (a) employee participation measured by the decision to submit a proposal and engage in an organizational improvement task and (b) the quality of the submissions as measured by (over 12,000) peer ratings and by the management organizing the contest.

Testing the presence of both participation and quality effects is important as the presence of systematic quality differences associated with different motivations would substantially complicate the problem of incentives for the organization. For example, a higher employee participation may represent a desirable outcome for the organization but it may also be a waste of resources if the increased participation level is mainly driven by employees with low-quality proposals who would have stayed away from the contest otherwise.

Another key issue is whether the solicited contributions will be representative of the needs of the entire organization, or just a smaller group of people. The presence of sorting effects based on the gender, for example, may impact on the extent and type of public goods provided, complicating the analysis of the incentives substantially. So, another aim of our study is to test to see whether differences in the motives to participate are attributable to factors associated with the employee’s profession, gender, and position inside the organization.

The analysis reveals that our solicitations achieved overall significant effects on employee participation, with small and insignificant effects on the quality of the proposals. More specifically, our findings suggest that: (1) the opportunity of winning a prize dominates all other incentives, leading to significant increase in employee participation rates in the innovation contest; (3) the increase in participation rates associated with the announcement of a prize is without lowering the quality of submissions; (3) calibration of a simple linear public-good model, reveals further that this increase in participation rates may go beyond the extrinsic value of the prize, consistently with our theory of prizes as means to internalize public goods.

We also find that: (4) the opportunity of leading implementation of one’s own submitted project proposal is the least effective incentive, leading to a lower employee participation to the innovation contest relative to all other solicitation treatments.

In addition, by looking at the sorting patterns by gender, profession, and position inside the organization, we find that: (5) solicitation treatments with mission-oriented incentives alone may result in responses that appear sensitive to the gender of the solicited person (women’s response to solicitations for improving patient care is higher than men’s); and (6) gender differences in preferences, such as competitive inclinations or risk aversion, may not exert great influence on responses of workers to the competition-for-prizes incentive (women’s and men’s response to solicitations for prizes are the same).

We discuss the implications of these results for the provision of public goods inside organizations in Section [summary-and-conclusions].

# Literature

Our work adds to the literature that analysis the problem ofr free-riding inside organizations, e.g., Alchian and Demsetz (xxxx). Much of xxx, however, is focused on compensation schemes for team production, such as internal contests (e.g., Erev et al., 1993), incentives based on group performance (Holmstrom and Milgrom, 1990), specialized task structures (Itoh et al., 1991), and reputation and peer monitoring (Che and Yoo, 2001). But the focus in xxx is more on team incentives, where peer monitoring could solve. In our study, the mechanism of peer monitoring is prevented by the innovation contest xxxx. We add to these literature by focusing on xxx outside of teams.

This work adds to the empirical research on the role of prizes in the workplace. Existing studies, however, have focused on contests where xxx xxx and contest organizers xxxx. Much less attention has been given to contests where output benefits contestants and competitors are affecting one another other than via competition for prizes. There exists literature on prize-mechanisms for public goods, but these have not been studied inside organizations. Less studies have looked at the winning of grants. And framing? But not in terms of the public goods. There exists an extensive literature in economics on the use of contests as a source of incentives inside firms [xxxx]. Much of the existing theoretical literature, however, presumes that agents are motivated to compete on the basis of the utility derived by winning personal awards. Less attention has been devoted to situations in which a competitor’s performance generates public good effects for the other competitors, and the whole organization (a notable exception is XXX). By focusing on incentives to workers to do organizational tasks (tasks with public good effects for the organization), we improve the existing literature in this direction.

Economists have long recognized that prize-based competitions are an important source of incentives inside organizations (Lazear and Rosen 1981; Green and Stokey 1983; Nalebuff and Stiglitz 1983; Mary, Viscusi, and Zeckhauser 1984). Much of the existing theoretical literature in labor economics, however, presumes that agents are motivated to compete solely on the basis of the utility derived by winning one of the prizes. Less attention has been devoted to situations in which a competitor’s performance generates public good effects for the other competitors, and the whole organization; a notable exception is XXX. By focusing on incentives to workers to do organizational tasks (tasks with public good effects for the organization), we improve the existing literature in this direction.

for which there exist consistent findings across many different empirical settings, including sport competitions (**???**), production competitions in firms (**???**; Terwiesch and Xu 2008), and more recently online competitions (**???**; **???**).

To be sure, free riding incentives inside organizations have been widely studied in labor economics, especially in the context of team production (**???**; **???**; **???**; **???**). However, our study differs from much of the existing literature in that it focuses on an individual competition where the team component is missing. That is, the public good dilemma comes from externalities towards anyone in the organization, not just a set of identified team members. It follows that one can remove from consideration conventional team dynamics such as peer pressure, monitoring, reciprocity among team members, and other kinds of social interactions that have been shown to affect behavior in the presence of free riding incentives.

Our study is also related the literature in public economics that studies prize-based mechanisms to foster the provision of public goods. (**???**) appears to be the first to note that fixed-prize lotteries – a special case also known as “Tullock” contest – are widely used tools among non-profit fundraising firms, showing conditions under which these may increase the provision relative to voluntary contributions. This insight has spurred much attention in public economics with several studies testing this idea empirically (see **???** for a survey). However, as noted by Vesterlund, the existing evidence on the profitability of lotteries for charities is only mixed. Our work extends the existing literature on the topic by focusing on an organizational setting where monetary contributions are replaced by effort and the “greater good” is helping the organization achieve its goals. Within this context, we find evidence that fixed-prize contests are a profitable tool to foster public good effects inside firms.

Finally, our work provides support to the incentive effect of mission-oriented preferences – inner satisfaction from helping the organization achieve its goals – (Akerlof and Kranton 2005; Besley and Ghatak 2005; Delfgaauw 2005; Delfgaauw and Dur 2008; Prendergast 2007; Rotemberg 2006) and social preferences at work (Bandiera, Barankay, and Rasul 2005; **???**; **???**; **???**). According to this perspective, workers are motivated agents. They do their work because they care about their co-workers, employers, and customers. Theoretical models suggest different ways in which managers can exploit these intrinsic motivations to raise individual levels of participation and productivity. Here, we use announcing an internal contest for organizational improvements to make these motivations salient. We find that emphasizing mission-oriented motivations has countervailing effects: positive for women and negative for men. While this finding is consistent with altruism being an important driver of effort inside organization, it also suggests that people are sensitive to the framing and in ways that may be difficult to predict ex-ante.

# Conceptual 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 the public good constitutes the sum of innovation projects to improve the organization.[[5]](#footnote-32) 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 and of low quality with probability . 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.

Let consider first the simplest case where the probability so any project is of high quality for sure. We assume a linear model of the utility of a typical employee who contributes and benefits from total contributions of :

The benefits of contributing derive from three sources. First, there is an altruistic benefit from the improved workplace, . 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, , where, for simplicity, all efforts have an equal chance of being selected as the winner, as in Morgan (2000). The personal reward 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 . This includes the case in which workers may derive a personal satisfaction from contributing personally to the organization, often called warm glow preferences for giving (**???**). 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 .

Contributors incur some cost from developing and submitting a proposal, . If there are employees the public goods dilemma arises when . 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 and receive utility of

where denotes the expected level of contributions and is the probability of having total contributions. Or they can contribute nothing and receive utility of

If there are employees, then the unique symmetric mixed-strategy equilibrium is for each employee to contribute a proposal with probability . After using the binomial probability for , the payoff-equating condition to find a mixed-strategy equilibrium is:

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

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., ).
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 ().

Now suppose that the probability is less than one. The equilibrium public good is not deterministic but follows a binomial distribution with average , where the equilibrium probability can be derived as before with the only difference being that it is also an increasing function of the probability . This leads to the following prediction.

1. 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 . 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., when , and vice versa). This leads the final prediction.

1. 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).

# Context, experimental design, data

## Context

The empirical context for the experiment is the Massachusetts General Hospital’s (MGH) Corrigan Minehan Heart Center, or the “Heart Center” for short. The Heart Center is a prominent medical organization in the United States and a teaching hospital of the Harvard Medical School. 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 is in cooperation with the Heart Center’s launch of the Health-care Transformation Lab (HTL),[[6]](#footnote-35) an initiative aimed at developing innovative health care process improvements to enhance the health care safety and delivery of the hospital. The launch of the HTL was accompanied by the announcement of an internal *innovation contest*, called the Ether Dome Challenge[[7]](#footnote-36) that sought to engage all staff members to participate.

The communication around the innovation contest highlights 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 website reads:

“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.”

![Timeline of the innovation contest](data:application/pdf;base64,)

Timeline of the innovation contest

The innovation contest itself can be divided into three main phases (Figure [timeline]): submission, project evaluation, and implementation.

First, there is a four-week submission phase in which all staff members are encouraged to identify one or more organizational problems and submit proposals addressing them. Employee participation is voluntary and all project submissions can be done online via the website of the contest. The number of project proposals to submit is unlimited and proposals could cover any issue within the organization. The only constraint is that each proposal is limited to approximately 300 words to lower the costs of entry and encourage broader participation. Another constraint is that team submissions are not permitted to ensure that treatment effects could be isolated, identified, and matched to individual participants. Beyond matching each submitter’s characteristics to the randomly assigned treatment, limiting submissions to individual participation also lowers the incentives to communicate or exchange information with other employees. For the same reason, the website is designed to provide no feedback information about the status of the contest during the submission period; so that individual decisions could not be easily influenced by the perceived popularity of the contest or previous submissions.

It follows a two-week project evaluation phase in which all staff members are invited to rate the merit and potential of submitted proposals on a five-point rating scale. All evaluations are done online on the website of the contest. After signing up, each evaluator is shown a list of anonymized proposals to read and rate. Proposals are presented at random in batches of 10 each. Each proposal is described by a title, a main description of the problem to solve, and the proposal. Voting is then introduced by the following text: “Rate this idea” followed by the rating scale: 1-low; 2; 3; 4; 5-high. Ratings would be kept confidential and the website is designed to provide no feedback or any other kind of information that would influence individual judgment until this project evaluation phase is over. Evaluators are left free to decide how many (and which) proposals to rate. However, since these are presented in a random order, every proposal had on average the same exposure to people asked to rate its quality. As a compensation for their effort in rating, evaluators get a chance to win a limited edition T-shirt.

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

## Experimental design

Within this context, we designed our experiment that was conducted during the normal operations of the Heart Center.[[8]](#footnote-39). The basic idea of the experiment was to randomize the content of the communication announcing the innovation contest to all staff members. The start of the submission phase was indeed announced to everyone in a series of personalized emails. A direct message was sent to each contact in the list of employees’ emails from our subject pool.

The content of this communication with a placeholder for our solicitation treatment is reported below:[[9]](#footnote-40)

Dear Heart Center team member,

**Submit your ideas to [TREATMENT HERE]**

The Ether Dome Challenge is your chance to submit ideas on how to improve the MGH Corrigan Minehan Heart Center, patient care and satisfaction, workplace efficiency and cost. All Heart Center Staff are eligible to submit ideas online. We encourage you to submit as many ideas as you have: no ideas are too big or too small!

Submissions will be reviewed and judged in two rounds, first by the Heart Center staff via crowd-voting, and then by an expert panel. Winning ideas will be eligible for project implementation funding in the Fall of 2014!

The first paragraph of the above message was randomized into *four* different solicitation treatments (the exact words are in Table [experimental-design]), thus creating as many treatment groups of equal size (Table [experimental-design]). The first group was given a solicitation treatment (PRIZE) announcing the innovation contest as an opportunity to win individual prizes (iPad mini’s) for top submissions. The second group was given a solicitation treatment (FUND) announcing the contest as an opportunity to win a $20,000 budget for developing their project proposals. The other groups received solicitation treatments announcing the contest as an opportunity to improve the health care of their patients (PCARE) or the workplace (WPLACE).

@lp5cm>rr  
  
& &  
(lr)2-2(lr)3-4 & & freq. & %  
  
PRIZE & Submit your ideas to win an Apple iPad mini & 312 & 25  
[1.8ex] FUND & Submit your ideas to win project funding up to $20,000 to turn your ideas into actions & 308 & 25  
[1.8ex] PCARE & Submit your ideas to improve patient care at the Heart Center & 310 & 25  
[1.8ex] WPLACE & Submit your ideas to improve the workplace at the Heart Center & 307 & 25  
[1.8ex] Total & & 1237 & 100

A sample size of more than 300 units for each treatment group ensured a sufficiently high statistical power based upon standard power calculations on the difference of proportions. In testing the difference of proportions between any two treatments, the probability of type-I errors was slightly below for *small* differences at 5 percent significance level but higher than for *medium* and *large* differences at the more stringent 1 percent significance level.[[10]](#footnote-41)

Also, note the lack of a traditional “control” treatment in this study. Since the experiment was run in a workplace, we were constrained to carry out treatments having equal chances of being successful. This prevented us from having a ‘null’ treatment with no personalized incentives messaging as a control group. Indeed, the analysis focused on multiple comparisons of several unordered discrete treatments (e.g., prizes vs funding vs framing).[[11]](#footnote-42)

These solicitation messages were sent three times: at the launch of the submission phase, eight days from the launch and two days before the end of the submission phase of the challenge.

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 [experimental-design]).

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

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

## Data

Our subject pool is the entire population working at the Heart Center as of the end of 2014, a total of 1,237 individuals. For each individual, we have administrative data on the gender, the type of profession, and whether they had a fixed office location or not. Additional, complementary data are available for a limited group of 378 employees (31 percent). These extra data have 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.

We report summary statistics for the different variables by solicitation treatment (Table [summary-statistics]), showing that these are statistically balanced across groups. These also show that the large majority (72 percent) of employees in our sample are 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).

Nursing workers constitute about half of the sample, and the rest is split almost equally between physicians and administrative workers. Though we do not have data on income, there exist large differences in earnings across these professions. 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). It follows that, if staff members are motivated by the extrinsic value of the prize alone, one should expect large differences in participation rates across profession.

Finally, it is also important to remark that only half of the employees ( percent) have fixed office locations, as they may be on duty in multiple wards. However, more senior staff tend to have a fixed location. So, within each profession, this measure can be viewed as a proxy for the employee’s position or status inside the organization.

@lccccccc  
  
& &  
(lr)2-5(lr)6-7 & FUND & PCARE & WPLACE & PRIZE & % & Obs. & P-value  
  
Other & 30 & 30 & 26 & 32 & 29 & 362 & 0.84  
MD/Fellow & 19 & 18 & 18 & 18 & 18 & 226 &  
Nursing & 51 & 52 & 56 & 51 & 52 & 649 &  
Female & 69 & 70 & 75 & 75 & 72 & 890 & 0.16  
Male & 31 & 30 & 24 & 26 & 28 & 347 &  
[1.86ex] No office & 50 & 46 & 47 & 45 & 47 & 577 & 0.56  
Office & 50 & 54 & 52 & 56 & 53 & 660 &  
[1.86ex] Age\* &&&&&&  
18-25 & 6 & 8 & 8 & 6 & 6 & 24 & 1.00  
26-35 & 29 & 29 & 31 & 26 & 29 & 107 &  
36-45 & 18 & 19 & 24 & 16 & 22 & 81 &  
45 & 44 & 46 & 51 & 45 & 42 & 157 &  
[1.86ex] Tenure\* &&&&&&  
 10 & 40 & 31 & 36 & 37 & 36 & 132 & 0.89  
10-20 & 26 & 29 & 38 & 28 & 30 & 111 &  
20-30 & 12 & 19 & 15 & 10 & 14 & 50 &  
30-40 & 10 & 16 & 15 & 12 & 13 & 48 &  
40 & 10 & 4 & 8 & 8 & 8 & 28 &

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

# Analysis

## Employee participation

@lcccc  
  
&  
(lr)2-5 Submission & FUND & PCARE & WPLACE & PRIZE  
  
No & 301 & 296 & 291 & 289  
Yes & 7 & 14 & 16 & 23

We begin by focusing on the causal effect of our experimental intervention on employee participation (Table [submit]), which is defined by the percentage of employees who made project submissions within the four-week submission period of the contest.

Overall employee participation is 5 percent and the rate of participation differs significantly across the four treatment groups (a Fisher’s exact test gives a p-value of 0.026). This means that, although overall participation is moderate, our experimental intervention produces significant group differences in participation.

Pairwise comparisons among individual solicitation treatments further reveal that: (1) the PRIZE solicitation treatment generates xxx, xxx, and xxx times higher participation rates than the WPLACE, PCARE, and FUND, respectively; (2) the PCARE and WPLACE solicitation treatments generate basically identical participation rates; and (3) the FUND solicitation treatment generates xx and xx times less participation rates than xx and xxx.

@lccc  
  
& FUND & PCARE & WPLACE  
  
PCARE & 0.124 & &  
WPLACE & 0.055 & 0.688 &  
PRIZE & 0.003 & 0.132 & 0.269

*Note: This table reports the p-values of pairwise comparisons of proportions among solicitation treatments.*

To test to see whether these pairwise differences are statistically significant, we use pairwise two-sample tests of proportions (Table [pairwise]). The analysis reveals that: the positive difference in participation rates between the PRIZE and FUND solicitation treatments is statistically significant (p=0.003); the positive difference between the PRIZE and PCARE solicitation treatments is marginally significant (p=0.132); whereas the positive difference between the PRIZE and WPLACE solicitation treatments is insignificant (p=0.269); the negative difference between the FUND and WPLACE solicitation treatments is significant (p=0.055); and the negative difference between the FUND and the PCARE solicitation treatments is marginally significant (p=0.124). Overall, these findings are consistent with employee participation being higher under a solicitation with personal awards incentives; and lower under a solicitation with funding incentives.

To complement the above univariate analysis, we use a multiple linear regression model that explicitly controls for observable differences across staff members. Let denote employee making a submission, and otherwise. We assume the conditional probability of an employee making a submission is given by:

where is a constant, is the causal effect of the solicitation treatment assigned to an employee (), controlling for the employee’s profession (), the gender (), and a dummy for office location () indicating whether the employee had a permanent office instead of being assigned to a ward. Notice that, in our context, having a fixed office location is highly correlated with the type of profession.[[12]](#footnote-46) Nurses, for instance, 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. This variable is hence potentially controlling for income and hierarchical differences occurring within each profession (more than just the effect of having a fixed office location per se).

@lccccc  
  
&  
  
&  
  
& (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 a percentage point change in the probability of submitting. Solicitation treatment dummies are coded to indicate deviations from the overall probability of submitting. The asterisks , , indicate significance at 1, 5 and 10 percent level, respectively.

The results of the regression model above (Table [participation ols]) give estimates of the solicitation treatment differences relative to the overall mean participation and controlling for baseline characteristics. We find that, at the 95 level of statistical significance, employees in the PRIZE solicitation treatment are cf.full[treatment2PRIZE] percentage points *more* likely to submit compared to the overall mean, whereas employees in the FUND solicitation treatment are -cf.full[treatment2FUND] percentage points *less* likely to do so.[[13]](#footnote-47) Treatment differences appear thus fairly large (e.g., a 2 percentage increase is about half of the increase in participation associated with a higher organizational position, as captured by the office location proxy).

Following the existing literature on gender-based differences in preferences towards competition, one may anticipate participation rates to be lower for women, holding constant the employee’s profession. Contrary to these expectations, our results show the participation rate is higher for women than for men, although the effect is largely insignificant. Similarly, one may expect differences in participation associated with the employee’s job. These may be driven, among various things, to differences in income, education, and the opportunity cost of effort (although no explicit barriers to enter the contest based on education or income are present). We find that the participation rate for nurses is higher relative to the residual category of other workers (e.g., admins), whereas it is lower for physicians. However, these effects, which are consistent with sorting, are not statistically different from the residual category of other workers, as well as from one another. Finally, we find a positive effect associated with the worker having a fixed office location, as opposed to being assigned to a ward (and the effect size doubles after controlling for the profession and gender). This evidence suggests that differences in the employee’s hierarchical position inside the organization, as captured by our office-location proxy, may be a stronger driver of participation relative to differences in gender and profession as sometimes assumed by the literature.

![Evolution of employee participation over time. Employee participation in the PRIZE treatment is higher than the other solicitation treatments at all periods. By contrast, employee participation in the FUND treatment is lower at all periods. The plot also shows little convergence of the participation rates over time.](data:application/pdf;base64,)

Evolution of employee participation over time. Employee participation in the PRIZE treatment is higher than the other solicitation treatments at all periods. By contrast, employee participation in the FUND treatment is lower at all periods. The plot also shows little convergence of the participation rates over time.

Next, we turn to examining participation dynamics (Figure [dynamics]). Though our data may not allow for a complete analysis of participation dynamics, looking at the overall submission patterns can be useful for the following reason. If employees assigned to different solicitation treatments were sharing (either face-to-face or electronically) the content of their solicitation with others, one should expect participation rates to converge over time, yielding estimates of the causal effects of a solicitation treatment biased towards zero. Contrary to these expectations, we find no evidence of convergence. Submissions in the PRIZE solicitation treatment are constantly higher than in the other treatments (except perhaps in the final week); at the same time, submissions in the FUND treatment are constantly low. These patterns are hence consistent with communication effects having little, or no, consequences on our findings, a topic we will discuss in greater detail later (Section [discussion]).

### Interactions

Following xxx literature (xxxx), we check whether gender was a factor driving participation in the contest (with men relatively more willing to sort into the competition). But we find no evidence (xxx gives a p-value). Similarly we check whether the presence of effects associated with unobservable characteristics that vary by job, like differences in income and education. But we find again no evidence in favor of this hypothesis (xxxx). We find instead evidence that having an office location is xxx associated with participation. This result, however, is confounded by other factors because more senior staff tend to have a fixed location but, at the same time, having a fixed office location is highly correlated with the type of profession (e.g., nurses xxxx).

We now turn to examining treatment interactions involving the employee’s gender and profession (Figure [interactions]).[[14]](#footnote-50) We hypothesize gender interactions to 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, we expect treatment interactions associated with the employee’s profession to occur because, for example, the prize opportunity (i.e., the PRIZE treatment) could be relatively less effective for employees with a higher income, such as doctors, than the others.

![Employee participation by gender or profession and solicitation treatment](data:application/pdf;base64,)

Employee participation by gender or profession and solicitation treatment

Examining the proportion of submissions conditional on the gender (Figure [interactions], panel a) shows that women are more likely (about 5 percentage points) to participate than men in the PCARE solicitation treatment. And examining the same proportion conditional on the profession (Figure [interactions], panel b) shows instead that doctors are as likely to submit as any other worker in PRIZE solicitation treatment; thus suggesting little sorting based on income or other characteristics associated with a given profession.

@lccc  
  
&  
  
&  
  
& (1) & (2) & (3)  
  
PRIZEfemale & 2.99 & 2.95 & 2.84  
& (1.68) & (1.79) & (1.78)  
& & &  
PCAREfemale & 1.25 & 1.21 & 1.08  
& (1.57) & (1.61) & (1.61)  
& & &  
FUNDfemale & 2.91 & 2.95 & 2.79  
& (1.06) & (1.20) & (1.19)  
& & &  
WPLACEfemale & 0.49 & 0.52 & 0.62  
& (1.35) & (1.44) & (1.43)  
& & &  
PRIZEmale & 1.37 & 1.42 & 1.40  
& (2.44) & (2.51) & (2.50)  
& & &  
PCAREmale & 3.75 & 3.72 & 3.64  
& (1.15) & (1.16) & (1.16)  
& & &  
FUNDmale & 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. Solicitation treatment dummies are coded to indicate deviations from the overall probability of submitting. The asterisks , , indicate significance at 1, 5 and 10 percent level, respectively.

To isolate gender and profession effects, we employ a version of model with gender-treatment interactions.[[15]](#footnote-52) The regression results (Table [tab: probability submitting interactions]) show similar results to the simple comparison of proportions. That is, after gradually adding profession and office controls, interaction coefficients remain stable across all specifications: the response of men under the PCARE solicitation treatment 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 (), which is consistent with our previous analysis. Thus, and overall, men respond less than women in the PCARE solicitation treatment, controlling for the profession and office location. This effect could be due to gender differences in preferences, as suggested by the literature, and we will return on this topic in the discussion of the results.

## Employee participation in peer evaluation

We now turn to examining the outcomes of the peer evaluation that follows the submission period. In this phase, 113 project proposals ended up being rated by a total of 178 employees (14 percent of our sample) who volunteered for the task. Their effort yielded a total of 12,219 evaluator-proposal pairs, providing a very sensitive test for differences in project quality across our solicitation treatments.

@lcccc  
  
&  
(lr)2-5 *Peer evaluation:* &FUND & PCARE & WPLACE & PRIZE  
  
% making no evaluations & 88 % & 84 % & 84 % & 86 %  
& (272) & (261) & (257) & (269)  
% making evaluations & 12 % & 16 % & 16 % & 14 %  
& (36) & (49) & (50) & (43)

Interestingly, we note (Table [ratings]) more employees volunteering to evaluate proposals in the PCARE and WPLACE than in the PRIZE and FUND solicitation treatments. However, the differences attributable to our experimental intervention are statistically insignificant (a Fisher’s exact test gives a p-value of 0.339). So, we find no evidence indicating treatment effects on employee participation in evaluation.

We further check with linear regression (results not shown) whether the self-selected sample of staff rating proposals is representative of the whole organization, or just a subset. Testing for statistical significance of the coefficients for the profession, gender, ond office location shows that evaluators are broadly representative of the organization as a whole, with no differences based on gender and profession albeit with a significantly higher participation from staff members with an office location. This is consistent with our previous results about participation incentives. So, overall, the collected ratings appear a broadly representative sampling of opinions inside the organization.

## 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 treatment which generated the fewest submissions (FUND) also produced the highest quality submissions. A quality versus quantity trade-off would increase the complexity of choosing optimal incentives for employees. We examine the issue with the assessments of quality made by peers in the evaluation phase of the contest and, subsequently, by the management.

*Quality assessed by peers.* To check whether differences in the quality of the submissions can be explained by the solicitation treatments of the submitter, we first look at differences in the distribution of ratings obtained from peers. Overall, a project proposal is given the “neutral” point (i.e., a rating of 3) 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. This rating pattern does not change much when we condition the data to the solicitation treatments of the proponent (Figure [ratings]); suggesting an equal distribution of good and bad quality projects across the solicitation treatments.

![Aggregated differences in proposal quality by the solicitation treatment of the proponent](data:application/pdf;base64,)

Aggregated differences in proposal quality by the solicitation treatment of the proponent

To formally test this hypothesis, we aggregate the mean ratings for each proposal and regress these aggregate measures on solicitation treatment dummies. The regression results (not reported) show only an insignificant relationship between ratings and solicitation treatments. The treatment coefficients are all insignificantly different from zero, with the linear model not significantly different from a constant model (an overall F-test gives a p-value of ftest$p.value).

The above analysis on the aggregate ratings does not hold in general.[[16]](#footnote-56) 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 percent difference between the proportion of 5’s in the WPLACE treatsent versus the other distributions, 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.[[17]](#footnote-57)

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

The vetting process conducted by the HTL staff resulted in sum(z$score>0) proposals being scored (from 1 to 100 points) with the best sum(hc$finalist) proposals invited to submit implementation plans. The remaining sum(z$score==0) proposals were excluded (and received a score of zero) either because flagged as inappropriate for funding or because the proponent manifested no intention to participate in the implementation phase (a ft$method finds no association between proposals excluded and treatments with a p-value of ft$p.value).

The Spearman’s rank correlation coefficient between the scores given by the HTL staff and the average peer ratings was relatively high (spear), indicating good agreement between our two measures of quality. Indeed, as before, we find no treatment effects on quality using the scores (a kt$method gives a p-value of kt$p.value). We also find no treatment differences in the percentage of submitters being selected and invited by HTL staff to present additional implementation plans (a ft$method gives a p-value of ft$p.value). Although not significant, employees who made project proposals in the FUND solicitation treatment are 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.

## The content of the project proposals

The goal of the challenge was to improve Heart Center operations by identifying problem areas and potential solutions. The proposed projects broadly conformed to the stated goals of the contest, aligning with improving the work processes within the organization or providing high-quality patient care. For example, one project proposal that received high peer ratings was to create a platform for patients to electronically review and update their medicine list in the office prior to seeing the physician. Another was to develop a smartphone application showing a patient’s itinerary for the day providing a guide from one test or appointment to another. Nevertheless, other contest organizers may have varying goals and be concerned about different aspects of the submissions.

In order to examine additional dimensions of submission content, we now study the area of focus of the submissions. Of particular interest is understanding whether different wordings used in the general encouragement solicitations (either towards improving the workplace or targeting the wellbeing of patients) induce employees to concentrate on different categories.

Members of the HTL categorized each project proposal into one of seven “areas of focus” (Table [tab: area-of-focus]): 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.

We test overall association between these categories and the solicitation treatments with a ft$method. Results show a marginally significant (p=ft$p.value) association, which means that our solicitation treatments have indeed an effect on the content of the submitted proposals.

To test which areas of focus was affected by our treatment, we regress the probability of a project proposal being in a given category against solicitation treatment dummies. We use an F-test where the null hypothesis tested is that all the treatment effects have a zero effect on the probability of the proposal being in a given category. The results of these F-tests of overall significance (Table [areas of focus]) reveals significant differences in the “Quality and Safety” and “Information and access” categories, which we view as improvements centered around patients (as opposed to workplace improvements). The first significance result is due to project proposals in the PCARE solicitation treatment being less likely to fall in the “Quality and Safety” category. The second result is due to project proposals in the FUND solicitation treatment being less likely to fall in the “Information and access” category.

Although it is difficult to interpret these results because our model does not provide any prediction on the content of proposals, they indicate a possible trade-off between stimulating participation via solicitations and inducing selection in the type of contributions to the public good, which complicates the analysis of incentives for public goods inside organizations beyond what the current literature anticipates.

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

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

# Summary and conclusions

We report results of a natural field experiment conducted at a medical organization that held an innovation contest seeking contribution of public goods (i.e., projects for organizational improvement) from its more than 1200 employees. The experiment tested incentives for contributing by manipulating the content of emails soliciting staff participation. We presented different incentives to participate in the contest, such as a prize (PRIZE) for winning submissions, improving patient care (PCARE), improving the workplace (WPLACE), and funding for implementation (FUND). Each staff was randomly assigned to receiving an email containing one of the four incentives.

We find that the PRIZE solicitation treatment boosts participation by about 40 percent relative to the WPLACE and PCARE solicitation treatments. The FUND solicitation treatment is the least effective. It generates not only about three times less submissions than the PRIZE solicitation treatment, but also less submissions than the WPLACE and PCARE solicitation treatments.

These participation differences, we find, are without changing the quality of the submissions as judged by peers and the management.[[18]](#footnote-60) The higher employee participation in the PRIZE solicitation treatment does not seem to be driven by low-quality submissions. Similarly, the lower employee participation in the FUND solicitation treatment does not seem to be driven by high-quality submissions. In other words, treatments that attracted more (or less) participation resulted in proposals of comparable quality and content.

Taken together, these findings suggest that (1) the competition-for-prizes incentive dominates mission-oriented incentives and (2) the opportunity to lead implementation of one’s own submitted project proposal is a poor incentive. In addition, these effects combined with the small (extrinsic) value of the prize relative to the median income of the participants, the long odds of winning the prize, the lack of differences in participation among professions, and the foreseeable additional costs of winning may suggest that (3) the effect of a prize competition on participation goes beyond the actual value of the prize itself, suggesting workers have, in fact, internalized some of the benefits of participating in an organizational task.

We also fined that, although the WPLACE and PCARE solicitation treatments are equally effective on average, responses appear sensitive to the gender of the solicited person. Women’s participation is greater when emphasizing the patient care whereas men’s participation is significantly lower, controlling for the profession and position inside the organization. This finding suggests that gender may be an important factor influencing sensitivity of responses to solicitations concerning the organizational mission.

At the same time, only an insignificant gender-based differences with respect to participation in the PRIZE solicitation treatment was found: women’s participation was slightly higher but not significant than men’s, all else being equal. This evidence indicates that gender differences in preferences, such as competitive inclinations or risk aversion, may not exert great influence on responses of workers to contests inside organizations.

We believe these results have three main implications for comparable organizations and, more broadly, the internal provision of public goods.

The first implication is that announcing a competition for an individual prize foster workers’ participation in organizational tasks beyond the value of the awarded prize itself. That is, prizes generate two opposing externalities that help workers internalize the public good effects of their organizational contributions. This result is important because it highlights a relatively less understood function of contests that is to mitigate the free riding incentives on organizational tasks.

A second implication is that offering the opportunity to lead collective projects can exacerbate the free riding incentives. This result may appear contrary to intuition. In theory, one may benefit more from leading a project than winning an iPad. For instance, one may use the opportunity to signal project management skills to the management aiming for a career advancement; or steer some of the resources towards assets or problems that are relatively more beneficial to his or her situation compared to the rest of the organization. If so, why a negative result? We believe that the private benefits from winning were negligible in our setting. First, the opportunity for a career advancement is small because medical staff gets promoted on the basis of other parameters (e.g., the quality of care provided). Second, the peer evaluation and vetting by the management ensure that the winning contributions yield as distributed benefits as possible. These aspects may have eliminated the possible private benefits from leading a project, resulting in poor participation rates. This result is important because projects need to be lead by someone and making more resources available does not seem to increase volunteers.

A third implication is that participation in organizational tasks is sometimes triggered by mission-based preferences. Although we find evidence that these preferences can be an effective incentive, we also find gender-based selection effects that are difficult to predict ex-ante. Our experiment does not provide any insights to better interpret these differences. But a large literature has investigated gender-based difference in preferences (see **???** for a review) or difference in self-stereotypes (**???**) that could explain some of these effects. Yet, more experimentation is needed to understand the different drivers in the field inside organizations.

A few limitations of this study deserve consideration. The first is that the validity of our causal interpretation of the results rests on a few conventional assumptions (**???**). These include the “no interference between units” assumption. In our study, it is possible that communication among staff assigned to different treatment arms could have influenced decisions to participate. The magnitude of this interference would depend on intensity of staff communication and the density of social interactions. Both of which should be small because (i) an individual competition may provide only weak incentives for information sharing and (ii) the staff members are scattered across multiple buildings on the hospital campus. Even so, a potential inference bias may alter the results towards a null effect as differences in employee participation should converge towards zero when communication spreads the content of the different email solicitations. This goes against our results. Moreover, by looking at the temporal dynamics of submissions, we find no indication of a convergence in the participation rates. Hence, the assumption of no interference seems appropriate.

Another potential limitation is that staff members may have left the solicitation email that was sent to them unopened or unread, thus non-complying with the assigned solicitation treatment. As this kind of noncompliance is almost entirely unobserved,[[19]](#footnote-61) the analysis follows an *Intention-To-Treat* (ITT) approach, discarding entirely any information about the solicitation treatment actually received. The main drawback of an ITT analysis is that it does not answer questions about causal effects of the content of the solicitation itself, only about causal effects of the assignment to a solicitation treatment.

Finally, our results have implications that extend beyond the specific organization under study. While the choice of focusing on health care workers may limit the generalizability of our results in some respect, it should be noted that in the US alone health care spending accounts for 17 percent of the GDP (in 2015). And, more generally, our study results are also directly applicable to a variety of other professions exposed to a public good dilemma (e.g., teachers, public servants, researchers). In all these settings, our study suggests that contests soliciting employee contributions and awarding an individual prize to the winning contribution appear an effective way to foster the internal provision of public goods inside organizations.

Akerlof, George A, and Rachel E Kranton. 2005. “Identity and the Economics of Organizations.” *Journal of Economic Perspectives*.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul. 2005. “Social Preferences and the Response to Incentives: Evidence from Personnel Data\*.” *The Quarterly Journal of Economics*.

Besley, Timothy, and Maitreesh Ghatak. 2005. “Competition and Incentives with Motivated Agents.” *The American Economic Review*.

Cohen, Jacob. 1992. “A power primer.” *Psychological Bulletin*.

Cutler, David, Elizabeth Wikler, and Peter Basch. 2012. “Reducing administrative costs and improving the health care system.” *New England Journal of Medicine*.

Delfgaauw, Josse. 2005. “Dedicated Doctors: Public and Private Provision of Health Care with Altruistic Physicians.” *New England Journal of Medicine*.

Delfgaauw, Josse, and Robert Dur. 2008. “Incentives and Workers’ Motivation in the Public Sector\*.” *Economic Journal*.

Gibbons, Robert, and Michael Waldman. 1999. “Careers in organizations: Theory and evidence.” In *Handbook of Labor Economics*.

Glaeser, Edward L, Andrew Hillis, Scott Duke Kominers, and Michael Luca. 2016. “Predictive Cities Crowdsourcing City Government: Using Tournaments to Improve Inspection Accuracy.” *The American Economic Review*.

Green, Jerry R, and Nancy L Stokey. 1983. “A Comparison of Tournaments and Contracts.” *The Journal of Political Economy*.

Kreps, David M. 1997. “Intrinsic Motivation and Extrinsic Incentives.” *The American Economic Review*.

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. 2013. “Prize-based contests can provide solutions to computational biology problems.” *Nature Biotechnology*.

Lazear, Edward P, and Sherwin Rosen. 1981. “Rank-Order Tournaments as Optimum Labor Contracts.” *The Journal of Political Economy*.

Mary, O’Keeffe, W Kip Viscusi, and Richard J Zeckhauser. 1984. “Economic Contests: Comparative Reward Schemes.” *Journal of Labor Economics*.

Morgan, John. 2000. “Financing public goods by means of lotteries.” *Review of Economic Studies*.

Nalebuff, Barry J, and Joseph E Stiglitz. 1983. “Prizes and incentives: towards a general theory of compensation and competition.” *The Bell Journal of Economics*.

Prendergast, Canice. 2007. “The Motivation and Bias of Bureaucrats.” *The American Economic Review*.

Rotemberg, Julio J. 2006. “Altruism, reciprocity and cooperation in the workplace.” In *Handbook of the Economics of Giving, Altruism and Reciprocity*.

Terwiesch, Christian, and Yi Xu. 2008. “Innovation contests, open innovation, and multiagent problem solving.” *Management Science*.

1. Blasco: Harvard Institute for Quantitative Social Science, Harvard University, 1737 Cambridge Street, Cambridge, MA 02138 (email: ). Jung: Harvard Business School, Soldiers Field, Boston, MA 02163 (email: ), Lakhani: Harvard Business School, Soldiers Field, Boston, MA 02163, and National Bureau of Economic Research (email: ). Menietti: Harvard Institute for Quantitative Social Science, Harvard University, 1737 Cambridge Street, Cambridge, MA 02138 (email: ). 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. [↑](#footnote-ref-21)
2. When career advancements depend on this kind of behavior (and not on productivity alone), then the prospect of a promotion provide motivations (Kreps 1997). See Gibbons and Waldman (1999) for a survey of the theoretical literature on careers in organizations. [↑](#footnote-ref-27)
3. An extensive literature has looked at settings that mimic the employee-employer relationship in the laboratory [see XXX for example]. [↑](#footnote-ref-28)
4. Among the many examples of internal contests that have appeared in the news are the Apple’s 2016 contest among its store employees seeking ideas on how to improve the way it sells iPhones (“Apple seeks ‘pie in the sky’ ideas for innovation,” Computerworld, 2013); Xerox’s internal contest seeking employees ideas on how to make a more environmentally friendly workplace environment (“Xerox employees green ideas save company $10.2 million,” The Guardian, 2010); and AT&T’s ideation contests seeking employee ideas about new products (“AT&T develops employee ideas for innovation,” The Wall Street Journal, 2014). [↑](#footnote-ref-29)
5. Instead of using a summation, we could have used different functional forms for the collective benefits (e.g., the max). However, the presence of free-riding incentives does not crucially depend on this specific assumption. [↑](#footnote-ref-32)
6. See the website at: www.healthcaretransformation.org [↑](#footnote-ref-35)
7. The name is taken from a historical place on MGH’s main campus where the first public surgery using anesthesia was demonstrated in 1846. [↑](#footnote-ref-36)
8. Following standard taxonomy, the experiment is formally a *natural field experiment.* [↑](#footnote-ref-39)
9. An image of the exact email is in the Appendix. [↑](#footnote-ref-40)
10. The definition of small, medium and large differences is given by Cohen (1992); e.g., a difference of 5 percentage points of the pair is considered a small effect: see Cohen (1992) p. 158. [↑](#footnote-ref-41)
11. 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. [↑](#footnote-ref-42)
12. Much of the clinical staff might be mobile and only half of the employees ( 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. [↑](#footnote-ref-46)
13. Subtracting these two effects gives cf.full[treatment2PRIZE] - cf.full[treatment2FUND] which is the difference in the probability of submitting between PRIZE and FUND treatments. [↑](#footnote-ref-47)
14. We find no significant differences for interactions with office location, which we do not report for space limitation. [↑](#footnote-ref-50)
15. We also run a model with profession-treatment interactions and results are simular to those shown in Figure [fig: interactions]. [↑](#footnote-ref-52)
16. It crucially relies on the assumption that an increment in a proposal’s quality as measured by an increase in ratings from to is the same for any value . [↑](#footnote-ref-56)
17. 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, comparing 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. [↑](#footnote-ref-57)
18. We find good agreement (high positive correlation) in the assessed quality of proposals between peer ratings and the evaluations conducted by the management; thus suggesting incentives being sufficiently aligned. [↑](#footnote-ref-60)
19. The email was sent using the internal messaging system of the Heart Center, which, at the time, was not collecting individual analytics. [↑](#footnote-ref-61)