

Love the Job...or the Patient?

Task vs. Mission-Based Motivations in Health Care

Sheheryar Banuri

Philip Keefer

Damien de Walque



WORLD BANK GROUP

Development Research Group

Human Development Team

February 2018

Abstract

A booming literature has argued that mission-based motives are a central feature of mission-oriented labor markets. This paper shifts the focus to task-based motivation and finds that it yields significantly more effort than mission-based motivation. Moreover, in the presence of significant task motivation, mission motivation has no additional effect on effort. The evidence emerges from experiments with nearly 250 medical and nursing students in Burkina Faso. The students exert effort in three tasks, from boring to interesting.

In addition, for half of the students, mission motivation is present: their effort on the task generates benefits for a charity. Two strong results emerge. First, task motivation has an economically important effect on effort, more than doubling effort. Second, mission motivation increases effort, but only for mundane tasks and not when the task is interesting. Moreover, even for mundane tasks, the effects of mission motivation appear to be less than those of task motivation.

This paper is a product of the Human Development Team, Development Research Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The authors may be contacted at sbanuri@gmail.com; pkeef@iadb.org; and ddewalque@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Love the Job... or the Patient? Task vs. Mission-Based Motivations in Health Care

Sheheryar Banuri, Philip Keefer, and Damien de Walque¹

University of East Anglia
Inter-American Development Bank
Development Research Group, The World Bank

sbanuri@gmail.com
pkeef@iadb.org
ddewalque@worldbank.org

JEL Codes: C91; H83; J45

Keywords: public sector reform, civil service, intrinsic motivation, extrinsic motivation, performance

¹ Banuri: School of Economics, University of East Anglia, Norwich, UK, NR4 7TJ (e-mail: sbanuri@gmail.com); Keefer: Inter-American Development Bank, 1300 New York Ave, NW, Washington, DC (e-mail: pkeef@iadb.org); de Walque: World Bank, 1818 H St, NW, Washington, DC (e-mail: ddewalque@worldbank.org). The authors have no relevant or material financial interests that relate to the research described in this paper. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors and do not necessarily represent the views of the World Bank, the Inter-American Development Bank, their respective Boards of Directors, or the countries they represent. The authors are grateful for financial support from the World Bank. In addition, the authors are grateful to Dr. Maurice Ye and Bambio Yiriyibin, Ousmane Haidera, Paul Jacob Robyn, and participants of the CBESS seminar series at the University of East Anglia and the Netherlands Institute of Government annual work conference.

Introduction

Organizations make large investments to inspire nonpecuniary motivation. Unfortunately, they make these bets lacking three key pieces of information: which nonpecuniary motivations elicit greater effort; how the nonpecuniary motivations interact; and which employees are most susceptible to them. For example, nonprofit organizations emphasize mission motivation. They base recruitment decisions on mission dedication and spend millions to clarify and broadcast their mission, but do not know how the payoffs to these investments compare to or depend on task motivation.² In this paper, we report results from a novel lab-in-the-field experiment that, for the first time, distinguishes the contributions of two sources of nonpecuniary motivation to effort: the degree to which employees work harder for a mission that they value (mission-matching); and the extent to which the nature of the task itself motivates workers to exert extra effort (task motivation). We find that task motivation elicits significantly more effort, although mission motivation has received more attention in the literature. Furthermore, mission motivation contributes little to effort when task motivation is high. For mission-oriented organizations with task-motivated employees, therefore, significant investments in, and recruitment based upon, mission motivation are likely to be wasted.

The limitations of pecuniary compensation and the multiplicity of sources of intrinsic motivation confront organizations with the complex challenge of defining organizational objectives, designing tasks, and delineating human resource policies that optimally harness intrinsic and extrinsic motivation to elicit worker effort. Their challenge is further complicated by the fact that organizational arrangements that exploit one source of intrinsic motivation (e.g., the organization's mission) may have no effect on intrinsic motivation driven by another (e.g., worker interest in the task). A key consideration for organizations, therefore, is the relative importance of different types of intrinsic motivation in eliciting worker effort.

We focus on the behavior of 248 advanced medical and nursing students, for whom both task and mission motivation are likely to be salient, in a country, Burkina Faso, where extrinsic compensation for health workers is loosely related to actual effort. The between-subjects design randomly assigns subjects to undertake one of three possible tasks: two low motivation and one high motivation. The high motivation task closely reflects the real-world professional activities that these students have demonstrated, through their educational and career choices, that they prefer. Participation in the task is voluntary and independent of pay. Approximately half of the subjects are given a mission: their effort elicits donations to a poor primary school.

Subjects undertake the task, in two-minute intervals, as many times as they like, up to a maximum of 16 times (32 minutes). At the end of each interval, they are asked if they would like to continue the task or quit. If they quit, they complete a post-experiment survey, after which subjects are paid and free to leave. In one (low-motivation) task, subjects sit in front of a blank computer screen and do nothing (i.e. the task is simply a waste of time). The second (low motivation) task

² Expenditure by non-profits on marketing, branding and public relations may exceed \$5 billion per year. See http://www.huffingtonpost.com/tom-watson/consumer-philanthropy-non_b_36261.html and <http://adage.com/article/small-agency-diary/gooders-brands/127361/> for articles on this issue. Though pinning down the exact amounts is difficult, we know the sums are large because marketing agencies specialize in the non-profit sector: see <https://towerbrands.com/marketing-for-charities-not-for-profits-and-ngos/>. One objective of this spending is to shape the external image of the organization and to raise funds (Seo, Kim, and Yang, 2009). Another, however, is to enhance the organization's own productivity by strengthening the attachment of employees to the organization's mission.

asks subjects to move sliders on a computer screen (i.e. the task is boring). The third (high motivation) task, however, asks subjects to engage their medical knowledge: subjects view computer videos of a patient describing her (or her child's) medical conditions and then answer questions about how to treat the patient based on the information in the videos. In addition, half of the subjects in facing each task are provided a mission: engaging in the task generates donations to a poor school.

Not surprisingly, subjects engage in significantly more effort in the high motivation medical task. More surprisingly, the magnitude of the effect is large: effort more than doubles. Prior research finds that mission-orientation has a significant effect on subject effort (Banuri and Keefer, 2016). We replicate these results, but only for the low-motivation tasks. Subjects who engage in low-motivation tasks work significantly harder in the presence of a mission (i.e. when their effort benefits children in a poor school). However, subjects engaged in the high motivation task exert no greater effort when their effort benefits these children than when it does not: the quantity of effort is the same, regardless of whether the task has a mission.

These results have important implications for organizations. Prior research emphasizes potentially large payoffs, including lower wage costs, for organizations that can recruit workers who share their mission orientation. Our work suggests that mission orientation makes no contribution to effort when workers are highly task motivated. To the extent that mission- and task-motivation are not correlated, organizations are likely to incur a cost if they use mission criteria to place individuals with lower task motivation into tasks that elicit substantial effort among more task-motivated individuals. In contrast, organizations benefit significantly if they emphasize mission motivation in recruitment for tasks that elicit little task motivation. Our analysis also has implications for different strands of literature on intrinsic motivation; these contributions are discussed in the next section. We then describe the experimental design and present the results.

Contribution to prior research

Voluminous research addresses the impact of different nonpecuniary motivations on effort and the degree to which pecuniary motivation crowds out nonpecuniary motivation. Our study, by comparing two important nonpecuniary motivations, addresses two gaps in this literature. First, it quantifies the importance for effort of task relative to mission motivation; second, it examines whether motivation crowding theory (Frey and Jegen, 2001; Frey and Oberholzer-Gee, 1997) affects two nonpecuniary motivations in the same way as it does pecuniary and nonpecuniary motivations.

A wealth of studies has examined the relative contributions of extrinsic and intrinsic motivation to effort. In economics, Bénabou and Tirole (2006) analyze the effect of pecuniary incentives on effort in pro-social tasks. D'Adda (2011) does the same in the context of a field experiment examining forest conservation in Bolivia. Fehr and Gaechter (2000) and Fehr, Gaechter and Kirchsteiger (1997) find that pecuniary incentives crowd out intrinsic motivations to engage in reciprocal behavior. Reeson and Tisdell (2008) show that pecuniary incentives crowd out nonpecuniary motivations to contribute to public goods. In their analysis of 128 studies in the psychology literature, Deci, Koestner and Ryan (1999) conclude that the evidence supports the hypothesis that extrinsic motivation suppresses intrinsic motivation to exert on-the-job effort. However, the precise sources of intrinsic and extrinsic motivations are either heterogeneous or not identified. Judge, Thoresen, Bono, and Patten (2001) conclude that the hundreds of studies on the effects of job satisfaction on job performance pointed to a modest positive relationship, but that the literature is plagued by heterogeneity in the definition and measurement of these variables. Gagné

and Deci (2005) emphasize the importance of this ambiguity: the effects of pecuniary on non-pecuniary incentives would be more precisely identified if it were known whether workers found their tasks interesting, or whether they were motivated by the mission of their job. We disentangle these two motivations.

By focusing on task motivation and its interaction with mission incentives, we contribute to a substantial literature that focuses instead on the interaction between extrinsic factors and other forms of social/mission motivation. For example, Bandiera, Barankay, and Rasul (2005, 2007, 2000, 2010) report a series of field experiments manipulating extrinsic incentives and social motives among fruit-pickers. Social interactions are significant motivators, but less so when extrinsic incentives are high. Ashraf et al. (2014) introduce social incentives (using a tournament), which they find have a larger impact on effort than pecuniary incentives. The workers in these studies are engaged in low-skilled low-motivation tasks, raising the question of whether social motives also stimulate effort when task motivation is high. We address this question of external validity by focusing on the interaction of two types of intrinsic motives that are often found in high-skilled jobs, task and mission incentives.

A handful of papers, all in psychology, look specifically at the interaction of task and pecuniary motivation. Building on Fessler (2003), Bailey and Fessler (2011) find that pecuniary compensation has a smaller effect on subject effort the more interesting the task is to the subject.³ Pokorny (2008) also examines whether the effects of pecuniary incentives depend on task motivation. We advance research in this area both methodologically and substantively. Methodologically, we infer subjects' task motivation from their real-world choices: their investments in medical education. Previous work assesses the "attractiveness" to subjects of the task (e.g., assembling puzzles), using subjects' own ratings of the attractiveness of the puzzle picture. Substantively, we examine a different issue, the interaction between two types of intrinsic motivation, task and mission.

A central concern of the economics literature has been the effect of mission motivation on worker effort: to what extent does a strong match between the mission of an organization and the mission preferences of a worker increase worker effort? Francois (2000), Besley and Ghatak (2005), Benabou and Tirole (2006), Prendergast (2007) and Ellingsen and Johannesson (2008) are only a few of the many theoretical contributions in this area. The empirical literature has confirmed that mission matching leads to increased effort (e.g., Carpenter and Gong, 2016; Banuri and Keefer, 2016; Ashraf, Bandiera, and Jack, 2014). None of this work, either theoretical or empirical, considers task motivation and the relative magnitudes of the effort effects of task and mission motivation.

While no prior research examines the effects of mission motivation in the presence of task motivation, Ariely, Bracha and Meier (2009), and Carpenter and Myers (2010) examine trade-offs across extrinsic motivation, mission orientation and image motivation concerns, abstracting from task motivation. Ariely, Bracha and Meier (2009) find that pro-social effort declines when pecuniary incentives increase and effort is private information. Both Ariely, Bracha and Meier (2009) and

³ They are also concerned with task complexity, which they vary by setting the initial orientation of puzzle pieces such that puzzle assembly would be easier or more difficult for subjects. They find no effects of either salaries or task attractiveness when the task is complex. This could be the result of a small number of subjects in the treatment arm (they had 80 participants and 8 treatment arms). We find, in contrast, and in a much larger sample, that the effects of mission motivation are strongest when tasks are uninteresting and simple, and weakest when they are interesting and complex.

Carpenter and Myers (2010) find that when image motivation concerns are present – when effort is public – extrinsic incentives have less effect on effort.⁴ Friedrichsen and Engelmann (2017) find that subjects who care more about social approval are more likely to state a preference for fair trade when their statement is public knowledge, but only among those subjects who are not intrinsically motivated to buy fair trade products.

Experimental Design

We randomly allocate subjects to one of six treatments that vary with respect to task and mission motivation. Their utility from the task is a function of their salary, the intrinsic reward they receive from performing the task itself, and the intrinsic benefit they receive from the benefits that their task confers on others (through contributions made to a school attended by poor children). The literature on crowding-out hypothesizes that pecuniary rewards reduce the intrinsic reward from performing a task (Frey and Jegen, 2001; Frey and Oberholzer-Gee, 1997; Benabou and Tirole, 2003; Georgellis, Iossa, and Tabvuma, 2011; Ariely, Bracha, and Meier, 2009). No such behavioral hypothesis exists with respect to different types of intrinsic incentives, such as task and mission motivation.

We conjecture that there are diminishing returns to effort, such that, for a sufficiently large effect of one type of motivation, changes in the other type have a (relatively) small effect on effort. Worker utility is separable in the welfare improvements that they experience from engaging in tasks that are intrinsically rewarding or that satisfy their social preferences, and the disutility caused by the exertion that effort requires. Assume that worker contribution to output is given by $\ln(e_i)$, a function of effort, e_i . DellaVigna (2017) recommends that the functional form for the cost of effort allow for the elasticity of effort with respect to the “value” of effort to vary. The cost of effort is therefore given by the power function $\frac{e^\gamma}{\gamma}$, where γ is the value of effort (literally, the degree of curvature in the effort function, as in Bellemare and Shearer 2009). The cost of effort increases in effort, $\gamma > 1$. Worker utility is then given by:

$$(1) \quad u_i = w + (\varphi_i + \theta_i)\ln(e_{ij}) - \frac{e_i^\gamma}{\gamma}.$$

In (1), worker utility rises with the flat salary, w , which is independent of their effort. It rises with effort depending on workers’ task motivation φ_{ij} and mission, θ_i , but at a declining rate. The exertion required by additional effort similarly reduces utility, but at an increasing rate, $\frac{e^\gamma}{\gamma}$.

Maximizing utility with respect to effort yields optimal effort $e_{ij}^* = (\varphi_{ij} + \theta_i)^{\frac{1}{\gamma}}$. The two key comparative statics that we examine below are $\frac{\partial e_i^*}{\partial \theta_i} = \frac{1}{\gamma}(\varphi_i + \theta_i)^{\frac{1-\gamma}{\gamma}} > 0$ and $\frac{\partial e_i^*}{\partial \varphi_i} = \frac{1}{\gamma}(\varphi_i + \theta_i)^{\frac{1-\gamma}{\gamma}} > 0$: effort rises in motivation. However, crucially, the degree to which an increase in one type of intrinsic motivation increases effort is dependent on the contribution to utility of the other type of

⁴ As we do, Carpenter and Myers (2010) study the behavior of a group of individuals – volunteer firefighters - who might be expected to be particularly motivated by their task. However, their research is not concerned with this aspect of intrinsic motivation.

intrinsic motivation: $\frac{\partial e_i^*}{\partial \varphi_i \partial \theta_i}, \frac{\partial e_i^*}{\partial \theta_i \partial \varphi_i} < 0$. Introducing mission motivation, $\theta_i > 0$, into a task should have a smaller effect on effort the more motivating is the task (the greater is φ_i), and vice versa.⁵

We estimate the magnitude of task motivation on effort by comparing effort across three tasks with different motivation, $\varphi_{i1}, \varphi_{i2} < \varphi_{i3}$ and find that $e_{i3}^* > e_{i2}^*, e_{i1}^*$. In addition, we test whether the effort effects of mission motivation decline with task motivation and find that $\frac{\partial e_{i1}^*}{\partial \theta_i} > \frac{\partial e_{i2,3}^*}{\partial \theta_i}$: mission motivation significantly increases effort only when task motivation is low. The reverse is not true, however: task motivation varies little whether or not the mission is significant. These point to high task motivation φ_{ij} relative to mission motivation θ_i .

To test these effects, we designed an experiment with the following blocks (see Figure 1). To measure mission motivation, all subjects play a dictator game, where the beneficiary is a poor school in Burkina Faso (subjects are provided basic information on the school along with some pictures of the students and facilities). Next, in all treatments we measure subject ability to undertake one low motivation (slider) and one high motivation (medical) task.⁶ After the motivation and ability measures, subjects were randomly assigned to one of the three tasks (blank, slider, or medical task). Half of the subjects in each task (randomly assigned) are given a mission: their effort generated monetary donations to the poor school. For the remaining subjects, effort yielded no benefits for the school. Note that across all treatments, additional effort does not yield additional pecuniary benefits for the subject. Furthermore, subjects do not receive feedback about their performance in the task and are informed that they will not receive feedback (minimizing image concerns). Subjects can engage in the core effort task a maximum of 16 times (for a minimum two minutes each time, yielding a maximum of 32 minutes). At the end of the effort task, subjects are given an exit survey, are paid their earnings, and then are free to leave.

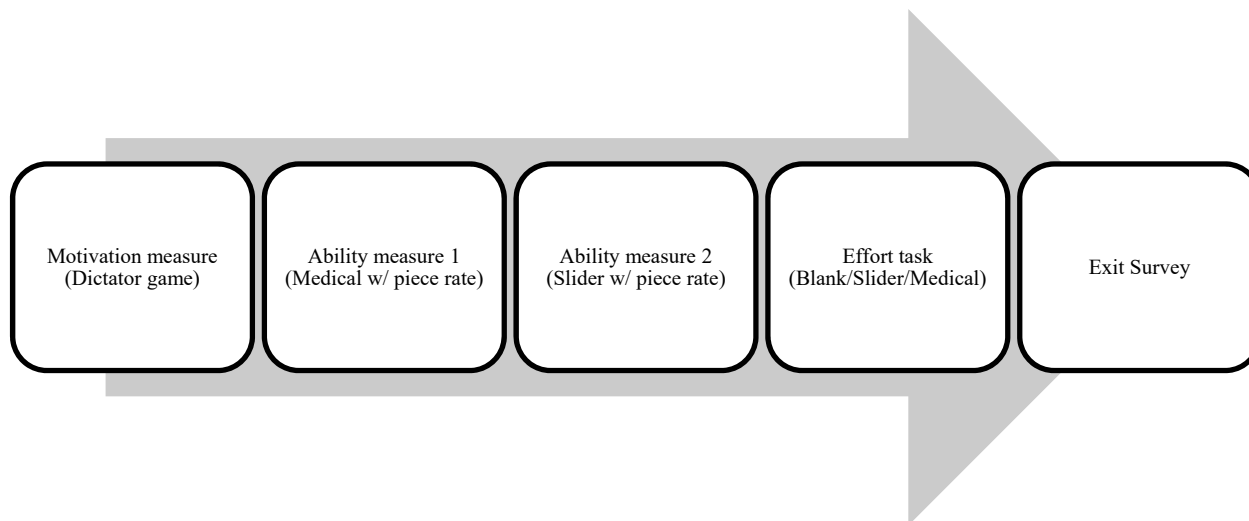
Subject effort might be influenced by expected differences in donation rates across the three tasks with a mission. We therefore took care to calibrate the link between effort and donations so that within-round donations would be similar across the three tasks. However, as tasks were different, and the nature of the effort in each task was also different, relating effort to payments was a challenge. Since the blank task had no real output, we implemented a piece rate to charity based on the number of times subjects chose to continue the task (200 CFA – \$0.42 – was donated for each time subjects continued the task). For the slider task, based on previous work in other contexts (see Banuri and Keefer, 2016), subjects could comfortably move 20 sliders per each 2-minute round. For this reason, we implemented a piece rate paid to the charity in the slider task of 10 CFA – \$0.02 – per slider (equating 200 CFA per round on average). Finally, from previous tests with the medical task, we knew that subjects had a 50 percent rate of accuracy on average in the medical task. As

⁵ Note that the same prediction emerges if we assume that there is a ceiling on effort. In this case, if one type of motivation is sufficient to induce effort close to the ceiling, we should observe no additional effect from other types of motivation. A ceiling effect would be particularly salient for measuring effort on the intensive margin – the amount of production per hour, for example. We focus on effort on the extensive margin (similar to Abeler et al. 2011, where subjects also choose when to stop working), however, the amount of time that individuals spend on their task, which in our experimental setting does not have a ceiling.

⁶ Note that the “Blank” task has no corresponding measure of ability, and hence no ability measure prior to conducting the task.

each 2-minute case contained 4 questions, we implemented a piece rate for charity of 100 CFA – \$0.21 – for each correct response (equating 200 CFA per round on average).

Figure 1: Structure of the Experiment



Assessing task motivation

We employ a novel approach to identify motivating tasks. Our subjects are medical students who, by their costly decision to enter nursing or medical school, revealed a strong preference for health-related tasks. The task that we judge to be the most motivating, because it matches the real-world choices of the subjects, is the one that involves analyzing patient reports of illness. We contrast effort under this task with effort under two other (low motivation) tasks, one requiring subjects to sit idly in front of a blank computer screen, and another that asks them to manipulate sliders on a computer screen.

Pokorny (2008) also assesses effort differences between two tasks, taking an IQ test or counting the number of “ones” and “sevens” in blocks of random numbers. The IQ test, plausibly offering greater task motivation, elicits greater effort. With our medical task, we can further buttress our claim that it offers the greatest task motivation by pointing to the correspondence between the task and the real-world choices of our subject pool.

Most prior research relies on surveys to establish motivation. Subjects are first asked to perform different tasks, and then they are asked how attractive or enjoyable the task was (e.g., Bailey and Fessler 2011). We do not rely on subject assessments of task interest, which could give rise to consistency bias if individuals who indicate that they prefer a task subsequently work harder on it for precisely that reason.

Measuring mission motivation

Mission motivation depends on the degree to which the task mission corresponds to the mission preferences of subjects. The greater is this match, the more effort subjects should exert. As in Banuri and Keefer (2016) and Ashraf, Bandiera and Jack (2014), we measure the mission motivation of subjects by asking them to play a modified version of the dictator “game”, with a poor

primary school as the beneficiary.^{7,8} Subjects were asked to donate as much as they liked out of an endowment of 1,250 CFA (\$2.60) to the primary school.⁹

Prior to making their decision, subjects were informed about the size of the school and the socioeconomic characteristics of its student body. To enhance the salience of the mission, subjects were shown a photograph of students sitting in a school classroom (see Appendix). Our measure is thus an ideal measure of mission motivation, since this same school was also the beneficiary of all mission-oriented tasks in the experiment.

Measuring effort

The key issue, here and in the literature, is the degree to which nonpecuniary motivation affects real effort. However, it is almost never possible to measure the exact mental, physical and emotional exertion entailed by real effort. Instead, researchers typically measure the output that subjects produce because of their effort. These measures are intrinsically noisy, reflecting not only subject motivation to exert effort, but also subjects' ability to undertake the task. An additional, important, challenge when analyzing task motivation, specifically, is that different tasks yield different outputs, making it difficult to compare effort across tasks.

We address the issue of comparability by creating a uniform measure of effort across tasks, the number of two-minute segments that subjects choose to spend on the task. This measure is homogeneous across tasks.¹⁰ The time that individuals spend on a task is only one of several types of exertion that effort could entail. For example, the effort required to spend time on a task may require a different type of exertion than the effort required to do the task well. However, when tasks are heterogeneous, effort exerted on quality is not comparable across tasks, unlike time spent.¹¹

⁷ In a typical dictator game, subjects are randomly assigned to groups of two, and one of them receives an endowment of \$10. The first player can transfer any proportion of the \$10 to the other player. Typically, individuals give on average about 10 percent of their endowment to the other player (Hoffman et al. 1994; Eckel and Grossman, 1996). We change the standard setup by replacing the second player with a poor primary school (Gampela 3) in the outskirts of Ouagadougou, the capital of Burkina Faso.

⁸ A large literature in behavioral economics uses the dictator game as its core measure of altruism and pro-social behavior (Forsythe et al 1994; Eckel and Grossman, 1996; Whitt and Wilson, 2007; among many others). Previous research has also replaced the recipient of the dictator game from a student to a charitable organization (Eckel and Grossman, 1996; Li et al, 2010; Carpenter et al. 2008, among others). Eckel and Grossman (1996) find, for example, that subjects give substantially more when the anonymous recipient is replaced with a charity (in their case, the American Red Cross).

⁹ The annual income per capita of Burkina Faso (in current US dollars) was approximately \$720 in 2014. The dictator endowment is approximately 134 percent of daily income per capita.

¹⁰ Unobserved differences in ability could have an indirect effect to the extent that higher ability individuals are more motivated by the task. This, however, is a source of noise in cross-task measurements of effort, not bias, since individuals are randomly assigned to tasks. Similarly, subjects may have unobserved differences in their opportunity costs of time. In practice, this unobserved difference refers to unobserved differences in the utility that subjects could gain by leaving the experiment 30 minutes early to study or chat with friends. Given the homogeneity of our subject pool, it is reasonable to assume that unobserved differences along these dimensions were small. Again, in any case, the random assignment of individuals to tasks attenuates any potential bias.

¹¹ This would again require comparing effort across tasks by measuring differences in output, but the outputs are heterogeneous and not comparable. One input into both quality and output is ability; our results, using time as a

In the first (low motivation) task (the “Blank” task), subjects sit in front of a blank computer screen. Only one measure of effort is relevant here, the number of two-minute segments that subjects undertake the task. The second (low motivation) task is the “Slider” task adapted from Gill and Prowse (2012). It demands real effort and some ability but is nevertheless dull. Subjects are shown 48 sliders on a computer screen. Each slider is set on the left, and the task for subjects is to move the slider precisely to the center of the slider bar. In each two-minute segment, subjects are asked to complete as many sliders as they can.¹² One frequently-used measure of effort with this task is the number of sliders that subjects move precisely to the center in every two-minute segment. However, because we want to compare effort across tasks, we instead simply count the number of two-minute segments that subjects choose to engage in the task.

Because we were working with subjects who have chosen medical careers, we determined that the third (high motivation), medical task would generally elicit a high level of task motivation. Standard approaches to measuring medical knowledge use survey vignettes, providing subjects with symptoms, and asking them to provide a diagnosis. We take a similar approach. Dr. Maurice Ye, of the Medical Research Center in Nouna, Burkina Faso, worked with us to create 20 cases of conditions that medical professionals in Burkina Faso would commonly encounter, focusing especially on maternal and early childhood care. They ranged from malaria and malnutrition to difficult pregnancies.

The development of each case entailed creating four multiple-choice questions and associated answers. The first question asked the subject to make a diagnosis; the second to identify the correct treatment; the third asked the subject to indicate if, and when, the patient should return for a follow-up visit; and the final question asked for the most appropriate follow-up treatment in case the initial treatment failed. Since medical cases were randomly assigned to subjects, cross-subject variation in the cases that they viewed could generate noise in the measurement of treatment effects. To reduce noise, great care was taken to ensure that the questions across cases were equally difficult to answer. For each question, the answers were designed so that one answer was correct, two were “almost correct” (e.g., they were consistent with most, but not all symptoms described), and two were entirely wrong.

We then worked with a film company to turn the cases into videos, hiring a screenwriter to develop scripts for each case and a well-known actress to play the role of the patient or mother of a patient.¹³ All of the videos were in French. Local languages are commonly used by patients, but French is also typical, is the language of instruction in nursing and medical school and was at least the second language of all the subjects. In the filming of the videos, care was taken to ensure that camera angles and the actress’ posture remained the same; to maintain interest, the video included

measure of effort, are robust to controlling for ability.

¹² The use of this computerized version of the “envelope-folding effort task” to simulate effort costs is common in the literature (Breuer, 2013; Georganas, Tonin, and Vlassopoulos, 2015; Ibañez and Schaffland, 2013; Banuri and Keefer, 2016; among others).

¹³ Here is the English transcript of one of the cases, in which the patient suffered from Mastitis: “Hello Doctor. I gave birth in your health center approximately one month ago. I'm back with another concern. Three weeks after my delivery, I started having pains in my right breast. My baby sucks a lot, it's too much. He nurses so much that I cannot close my eyes at night. From time to time, I get very hot. I have a fever and headaches; especially at night. Sometimes it gives me insomnia. I thus came to ask for your help in relieving my pain.” Additional information germane to the diagnosis was also provided, such as temperature, pulse, blood pressure, and additional notes from examining the patient. For more information on the development of the cases and videos, please see: <http://www.rbhealth.org/blog/measuring-quality-health-care-using-video-vignettes>.

close-ups, but camera movements were carefully controlled and homogeneous across cases. The videos and questions were then incorporated into a computer-based task. Each video lasted for 60 seconds. Subjects could re-watch the video, rewind, forward, or stop it entirely. Subjects could answer questions while watching the video as well. However, each segment lasted for 2 minutes, regardless of whether the subject had completed all the questions or not.¹⁴ As with the other two tasks, we use an effort measure that is unaffected by ability, the number of two-minute segments that subjects engaged in the task.

Measuring Ability

Note that in Figure 1 each treatment contains two measures of subject ability in the two tasks where ability matters: the slider and medical tasks. We use these to check the robustness of the main results to subject ability in the tasks. These ability measures were implemented across all treatments. First (ability measure 1), subjects were asked to undertake four medical cases, and were informed that they would be paid 100 CFA – \$0.21 – for each correct response. The ability measure was not timed, subjects spent 23.08 minutes (5.77 minutes per case) on average during the medical ability measure and had an accuracy rate of 45%. Hence, subjects earned 723 CFA – \$1.50 – on average in the medical ability measure.

Second (ability measure 2), subjects were asked to undertake four rounds of the slider task and were informed that they would be paid 10 CFA – \$0.02 – for every slider correctly positioned. In each round, subjects had a fixed amount of time (2 minutes) to move as many sliders as they could. Subjects correctly positioned 20.03 sliders (5.01 sliders per round) on average. Hence, subjects earned 200 CFA – \$0.42 – on average in the slider ability measure.

Additional Measures and Payment Procedures

Subjects received a flat wage of 4,000 CFA (\$8.32) for engaging in the final effort task but were also compensated for our motivation and ability measures described above. All sessions were conducted in May 2014. In addition, since subjects used a mouse to manipulate the sliders in the slider task, care was taken to utilize identical mice and computers at each location and to use the same screen resolution on the computers to minimize differences across samples. Since this was an individual task, multiple treatments took place within the same session. Subjects were randomly assigned to seats within the computer lab, and the actual treatments were randomly selected by the computer. 248 subjects participated in the specific experiments analyzed here; in the much larger project, of which these experiments were a part, 1,119 subjects participated. Table 1 presents the overall experimental design and the number of subjects in each treatment.

¹⁴ For more details on the medical cases, please see Banuri et al. (2017), also summarized here: <http://www.rbhealth.org/blog/measuring-quality-health-care-using-video-vignettes>.

Table 1: Number of subjects, by treatment

Task:	Blank	Slider	Medical
No mission (no donation)	37	46	48
Mission (donation to school)	38	35	44

After completing all the experimental tasks, subjects completed an extensive survey recording subject demographics. Each subject was informed that the total donation to the school generated by their actions would be put in sealed box and donated to the school at the end of the study period. Donations were contained in sealed plastic boxes typically used for collecting votes during elections. Subject watched as their donations were placed in the sealed container.

Fourth and fifth year medical students from the University of Ouagadougou (N=121) and third year nursing students from the National School of Nursing (École Nationale De Santé Publique; N=131) participated in the experiments. Subjects were recruited by posting flyers and through briefing sessions with representatives of student unions. The flyers indicated that participants would play games and be able to earn money. It did not reveal the nature or purpose of the experiments.

All earnings were expressed in tokens, with an exchange rate of 1.00 CFA per token.¹⁵ All subjects were paid in cash at the end of each session. Subjects were paid according to their decisions in the motivation measure (dictator game); the ability measure with sliders based on their performance (piece rate); the ability measure with medical cases based on their performance (piece rate); and the flat salary for the core effort task (independent of effort exerted). Donations to the school were generated based on the mission motivation measure (dictator game); and effort in the core effort task in the treatments with a mission. The average subject earned a sizeable amount, 5,742 CFA (\$11.94, or more than five times daily per capita income in the country). The average payment to charity, per subject, was 978 CFA (\$2.03). Table 2 presents the summary statistics for the entire sample.

¹⁵ We use “tokens” rather than cash to facilitate replication across cultures: currency focal points vary across countries. In implementing tokens, experimental protocols and instructions remain identical even when conducting experiments in different contexts. Though tokens reinforce the artificiality of the lab, replicability is a more important concern.

Table 2: Summary Statistics

	All treatments
Observations	248
Age (Years)	27.95 (5.15)
Female (%)	49%
Tokens donated to school (Dictator max: 1,250)	431.63 (334.53)
Score in slider task ability measure (Max: 192)	20.03 (16.73)
Score in medical task ability measure (Max: 16)	7.23 (1.96)
Risk preferences (% risk seeking) ¹⁶	35%
Current state of personal finances (% responding Excellent/Good)	11%
Confidence in payment to schools (% responding Agree/Strongly agree)	75%
Clarity of instructions (% responding Always/Most of the time)	75%
Qualifications studied for	
Nurses (%)	27%
Midwife (%)	25%
Doctor (%)	46%
Other (%)	0%
Years at institution	
First year (%)	1%
Second year (%)	0%
Third year (%)	50%
Fourth year (%)	23%
Fifth year (%)	25%

Notes: In order to select subjects at the appropriate level of education to participate in the medical task, we explicitly recruited final year (3rd year) students at the Nursing school, and 4th / 5th year students at the medical school. These students had the appropriate level of experience in terms of course- and fieldwork.

Results

The experiments shed light on two questions: How important is task motivation for effort? And does mission motivation crowd out the effects of task motivation on effort?

Task motivation and effort

The experimental design allows us to compare effort across the three diverse tasks with a simple metric: the number of times subjects chose to continue engaging in the task, each time for two minutes (fixed by design). At the end of each two-minute interval, subjects had the chance to end the task and continue to the exit survey (after which they were paid and were free to leave). To assess the effects of task motivation on effort, we use this measure of effort (number of rounds, which is directly translatable into the time that subjects spend on a task). We compare time spent

¹⁶ Risk preferences were measured using a survey question “In general, would you say that you are someone who takes risks, or do you avoid taking risks?” Responses were measured using a 5-point Likert scale with 1 = “Prefer to avoid risks”; 5 = “Prefer taking risks.”

across tasks holding constant private returns (the flat salary is the same in all treatments) and mission, or lack of mission (i.e. effort does not yield contributions to the poor school).

Figure 2: Effort and Task Motivation, No Mission Motivation

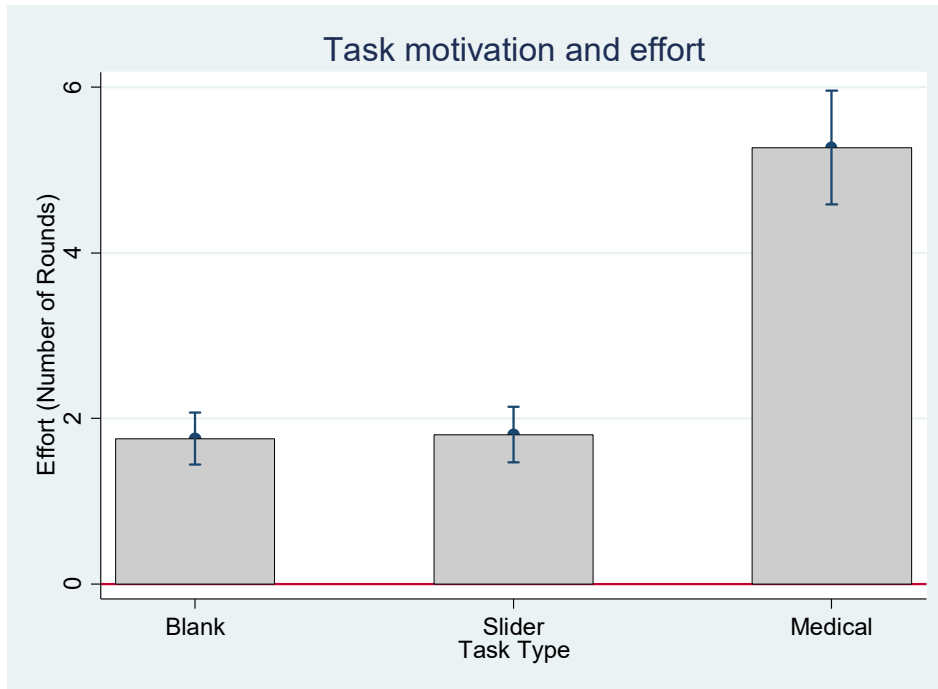


Figure 2 compares the number of rounds (i.e. time) spent by subjects across the three tasks. The direction of the results is entirely intuitive: subjects spent significantly more time on the most interesting task. However, the magnitude of the effect is striking: subjects spent more than twice as much time (more than five two-minute intervals compared to fewer than two two-minute intervals) on the medical task ($p < 0.01$), earning no additional reward, than they did on the slider or blank tasks. There was no significant difference in the number of rounds (time) spent between the two tasks with low task motivation.

We further investigate whether the effects identified in Figure 2 are driven by characteristics of the subject population. The relationship between task motivation and effort can then be identified in estimates of the following Tobit regression:

$$EFFORT_i = \alpha + \beta Slider_i + \gamma Medical_i + \delta CONTROLS_i + \varepsilon_i \quad (1)$$

where the dependent variable is the number of two-minute intervals spent on the task, the *Slider* dummy captures the motivational effects of the slider task relative to the omitted blank-screen task; and the *Medical* dummy yields an estimate of the task motivation of the health task relative to the blank screen. The *CONTROLS* consist of: occupation dummies (the medical, nursing, or midwife students, since these students might be differently motivated by the medical task, or differently demotivated by the other tasks); gender; age; years spent in the institution; current state of financial resources; risk preferences; and subject self-assessment of the clarity of instructions (which may be important given that the task is unusual).

Table 3: Effort and Task Motivation, Controlling for Observables

Dependent variable: Effort (Number of rounds)				
	I	II	III	IV
Treatment: Slider	0.048 (0.29)	-0.010 (0.31)	0.051 (0.24)	0.270 (0.24)
Treatment: Medical	3.592*** (0.58)	3.633*** (0.60)	3.718*** (0.62)	4.343*** (0.63)
Training: Midwife		-0.609 (1.35)	-0.331 (1.26)	-0.217 (1.44)
Training: Doctor		-1.772* (0.92)	-2.326** (1.04)	-2.642** (1.15)
Education level (years)		0.837 (0.59)	0.875 (0.61)	0.858 (0.63)
Female 1 = Female			-0.842 (0.58)	-0.617 (0.58)
Age (years)			-0.041 (0.05)	-0.066 (0.05)
Current state of personal finances 4 = Excellent			0.055 (0.46)	0.268 (0.42)
Risk preferences 5 = Risk seeking				0.695** (0.27)
Clarity of instructions 5 = Always clear				0.649*** (0.11)
Constant	1.757*** (0.38)	-0.345 (1.64)	1.088 (2.57)	-3.201 (2.97)
Sigma				
Constant	3.404*** (0.38)	3.375*** (0.39)	3.356*** (0.37)	3.210*** (0.36)
Pseudo R2	0.042	0.045	0.047	0.063
Log Likelihood	-342.6	-341.6	-340.8	-335.0
P	0.000	0.000	0.000	.
Observations	131	131	131	131
Right censored observations	4	4	4	4

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable is effort (number of rounds respondent chose to continue the task). Tobit specification with upper censors at 16 (the maximum rounds subjects continue prior to auto-exit), with clustered standard errors (by day) in parentheses. Table reports regression coefficients. The data are restricted to treatments with no mission motivation (consistent with figure 2). Results are robust to an OLS specification.

The results are displayed in Table 3. The central findings are in the first and second rows, which compare the effort expended by subjects on the slider or medical tasks relative to the (omitted) blank screen task. Across all specifications, subjects expended more effort on the high-motivation Medical task relative to the Blank task ($p < 0.01$ across all specifications). There are no differences in effort between the low-motivation Blank and Slider tasks ($p = 0.87$ in model I – no controls – and $p = 0.26$ in model IV – full set of controls).

The findings in Table 3 provide direct, behavioral evidence that task-motivated subjects do, in fact, provide more effort, even in the absence of extrinsic rewards. In addition, Table 3 estimates controls for whether subjects are medical, midwife or (the omitted category) nursing students. Doctors exert less effort on all tasks than nurses and midwives. This is consistent with the possibility that the opportunity costs of time of medical students are greater (e.g., because of the demands of their coursework). Risk-seeking subjects exert greater effort ($p < 0.05$). Interacting risk preferences with the treatment dummies reveals risk-seekers more likely to continue the Medical task ($p < 0.05$) rather than the Slider ($p = 0.16$) or Blank tasks ($p = 0.72$). Finally, instruction clarity is also significantly related to effort: subjects who found the instructions to be clearer were more likely to exert effort than those who did not.

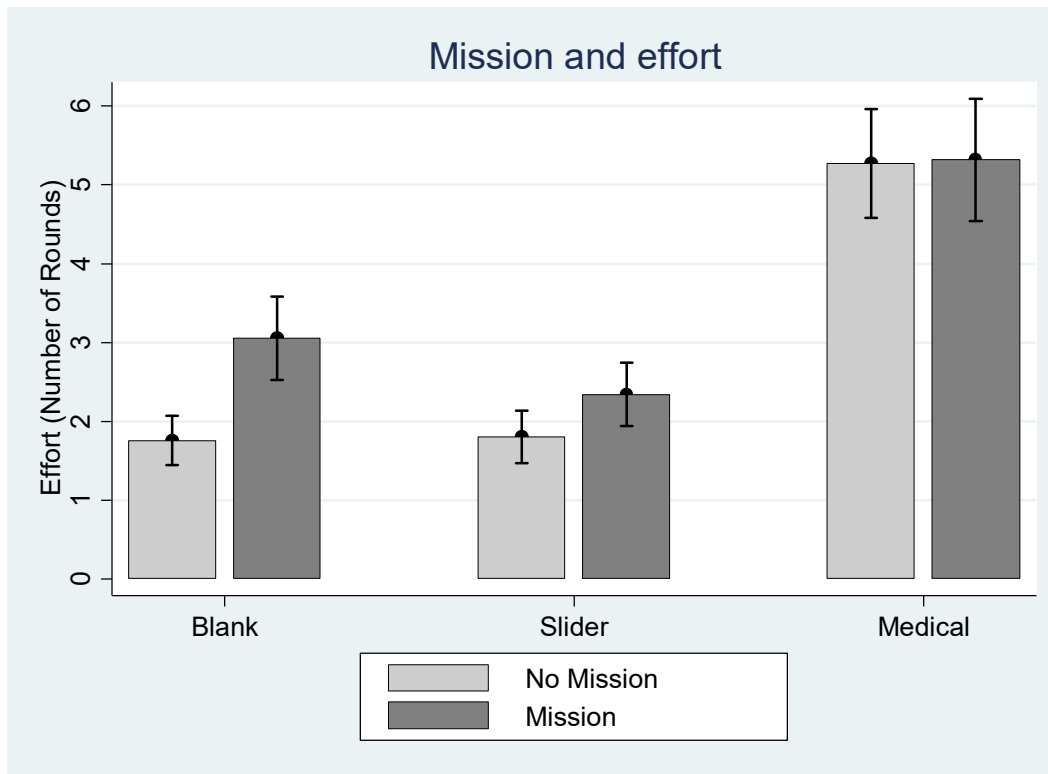
Mission vs. Task Motivation

The second question we address is whether the effort exerted by mission-motivated individuals depends on their level of task motivation. This question demands within-task, between subject comparisons: how does an increase in mission motivation affect subject effort across each of the three tasks? If task motivation reduces the returns to mission motivation, we expect that mission motivation should have a significant effect on effort in the low-motivation blank screen and slider tasks, but a smaller effect on the high-motivation medical task.

Figure 3 summarizes the answer to this question. Comparing the first two bars, mission motivation matters in the absence of task motivation: subjects who engage in the Blank task spend nearly 75 percent more time on the task when their effort benefits the poor school than when it does not ($p < 0.05$). Time spent in the slider task also increases by 30 percent when subject effort benefits the poor school, though this increase is not significant ($p = 0.30$). However, we earlier conjectured that if there are diminishing effects of intrinsic motivation, mission motivation should have little or no effect on effort in the high motivation task. This turns out to be the case. Effort on the medical task is indistinguishable across the mission and non-mission settings (the last two bars of Figure 3 indicate an increase of just 1 percent; $p = 0.96$).

Figure 3 also illustrates the relative magnitudes of the task and mission motivation effects: how much more effort does greater task motivation elicit in the absence of mission motivation compared to how much more effort mission motivation elicits in the absence of task motivation. The magnitude of the pure task motivation effect turns out to be much greater than the pure mission effect. Subjects spent 200 percent more time on the high-motivation medical task in the absence of mission motivation (the fifth bar) than on either of the low-motivation tasks in the absence of a mission (the blank task, the first bar; and the slider task, the third bar). This is high compared to the effects of mission motivation in the tasks with low motivation, which raised time spent on the task by 75 percent (blank task) and 30 percent (slider task).

Figure 3: Mission motivation by task



Again, features of the subject pool could account for these results. To control for these, we report results from Tobit regressions that take the following form:

$$EFFORT_i = \alpha + \varphi_2 \text{Blank with Mission}_i + \beta_1 \text{Slider, No Mission}_i + \beta_2 \text{Slider with Mission}_i + \gamma_1 \text{Medical, No Mission}_i + \gamma_2 \text{Medical with Mission}_i + \delta \text{CONTROLS}_i + \varepsilon_i \quad (2)$$

The omitted, benchmark category is the blank task with no mission. We expect effort to be higher in the mission than the non-mission task, but only in the low motivation tasks. Hence, the coefficients on φ_2 and β_2 should be positive, and β_2 should be larger than β_1 . However, γ_2 should not be different from γ_1 .

Table 4 presents results in line with predictions. In all specifications, treatment effects are relative to the treatment with no mission or task motivation, “Blank, No Mission” (the omitted treatment). First, no matter what controls are included in the specification, effort exerted under the high-motivation medical task is significantly greater than under the omitted treatment, with or without a mission ($p < 0.01$). Second, mission has a significant effect on effort when task motivation is low: the coefficient on the “Blank with mission” treatment dummy is significantly different from 0 across all specifications (with no controls, $p < 0.10$, in line with figure 3).

Table 4: Task and Mission Motivation, Controlling for Observables

Dependent variable: Effort (Number of rounds)				
	I	II	III	IV
Treatment: Blank with Mission	1.320*	1.290**	1.477**	1.553***
	(0.68)	(0.62)	(0.62)	(0.59)
Treatment: Slider, No Mission	0.048	0.071	0.247	0.373
	(0.29)	(0.33)	(0.31)	(0.29)
Treatment: Slider with Mission	0.586	0.714	0.921*	1.244***
	(0.41)	(0.46)	(0.49)	(0.45)
Treatment: Medical, No Mission	3.604***	3.862***	4.001***	4.418***
	(0.58)	(0.58)	(0.65)	(0.68)
Treatment: Medical with Mission	3.660***	3.692***	3.808***	4.099***
	(0.41)	(0.49)	(0.44)	(0.49)
Training: Midwife		-0.869	-0.425	-0.164
		(0.57)	(0.64)	(0.55)
Training: Doctor		-1.035	-1.492*	-1.638***
		(0.78)	(0.78)	(0.60)
Education level (years)		0.678	0.653	0.421
		(0.47)	(0.46)	(0.47)
Female			-1.152***	-1.005**
1 = Female			(0.41)	(0.39)
Age (years)			-0.040	-0.059*
			(0.04)	(0.03)
Current state of personal finances			0.264	0.376
4 = Excellent			(0.53)	(0.54)
Risk preferences				0.448***
5 = Risk seeking				(0.14)
Clarity of instructions				0.578***
5 = Always clear				(0.13)
Confidence in payment to schools				0.071
5 = Strongly Agree				(0.27)
Motivation (dictator)				0.001
CFA donated to school				(0.00)
Constant	1.757***	-0.136	1.147	-2.166
	(0.38)	(1.49)	(1.78)	(2.26)
Sigma				
Constant	3.691***	3.658***	3.623***	3.524***
	(0.35)	(0.35)	(0.33)	(0.35)
Pseudo R2	0.031	0.033	0.037	0.047
Log Likelihood	-666.2	-664.2	-661.8	-655.0
P	0.000	0.000	.	.
Observations	248	248	248	248
Right censored observations	9	9	9	9

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable is effort (number of rounds respondent chose to continue the task). Tobit specification with upper censors at 16 (the maximum rounds subjects continue prior to auto-exit), with clustered standard errors (by day) in parentheses. Table reports regression coefficients. Results are robust to an OLS specification.

Third, recall that effort in the slider task in the figure was not significantly different with and without a mission. When using a Tobit specification and clustering the standard errors, the difference in the Slider and Slider with mission treatments remains insignificant ($p = 0.11$, the first column of Table 4). Estimates of the effects of the slider treatment on effort are, however, less stable across specifications than those of the other two tasks. The coefficient estimates on the medical treatments do not vary more than 22 percent across specifications, whereas those on the slider treatments vary as much as 677 percent. This might be due to larger differences across subjects in slider task motivation, which are captured by controls for observables. When controlling for occupation (nurse, midwife or doctor in model II), mission has a significant effect on effort on the slider task ($p < 0.10$). This effect becomes even stronger with the full set of controls in model IV ($p < 0.05$). Hence, the presence of the mission increases effort when task motivation is low. When task motivation is high, the difference is not significant with ($p = 0.61$) or without controls ($p = 0.94$). These results are broadly consistent with theory: introducing the mission increases effort, but only in low-motivation tasks (such as the Blank and Slider tasks).

Furthermore, the model allows us to conduct a difference-in-difference on the effects of mission across task types. The effect of mission motivation on the blank task is significantly larger than the effect of mission motivation on the medical task¹⁷ ($p < 0.01$ using coefficients reported in model IV – full set of controls). When comparing the effects of mission on the slider task relative to the medical task, the difference is not significant¹⁸ ($p < 0.15$ using coefficients reported in model IV – full set of controls). These results provide additional support for the claim that mission motivation has a smaller effect on effort when task motivation is high.

In addition to the treatment effects, we also find that women provide significantly lower effort ($p < 0.05$), which may indicate a higher opportunity cost of time (for example, greater household responsibilities in addition to time needed at school). Consistent with the opportunity costs of time, we find that doctoral students and older students provide significantly lower effort. As before, risk seekers and subjects that found the instructions to be clear were likely to exert higher effort overall.

Our effort measure focuses on the extensive margin: how much time do subjects invest in the task? It has the great advantage that it is homogeneous across tasks with heterogeneous outputs. However, it is perhaps generally true, and it is certainly true in our experimental design, that ability plays a larger role in the successful performance of higher motivation tasks. Medical knowledge is essential to the correct diagnosis and treatment prescriptions of the medical task; subjects who are more dexterous in the manipulation of sliders will be able to correctly position a larger number of sliders precisely at the midpoint of the scale in any two-minute period. Ability could account for the

¹⁷ A joint F-test was conducted testing whether the difference between coefficients of Medical with Mission and Medical, No Mission was different from the coefficient on Blank with Mission

¹⁸ A joint F-test was conducted testing whether the difference between coefficients of Medical with Mission and Medical, No Mission was different from the difference between coefficients of Slider with Mission and Slider, No Mission.

results in Table 4 if more able individuals are also more motivated to persist in the ability-intensive task for longer periods. The possibility that more able individuals are more task motivated does not undermine the conclusions we draw from Table 4, since it is still the case that motivated individuals exert greater effort. Our data nevertheless allow us to reject the hypothesis that ability differences drive the results in Table 4.

We measure ability in each treatment by providing subjects with a small piece rate for the number of sliders placed correctly or the number of medical questions correctly answered, over four rounds (the blank screen task has no ability component). We then asked, in two separate estimations that included the controls in model IV in Table 4 plus the ability measure, what effect does mission have on time spent on the slider task, and what effect does it have on time spent on the medical task. Ability is insignificant, and the mission results are the same as in Table 4 (see Appendix Table 1).

Conclusion

This paper extends the literature on motivation and effort by offering new insights on the relative empirical importance of two nonpecuniary incentives, task and mission motivation, and by exploring the interactions among the two. Using a unique sample of students of the health professions in Burkina Faso, along with a medical effort task specifically designed to motivate them, we find that subjects exert significantly greater effort when task motivation is high, and this effect is large compared to the effects of mission motivation. Furthermore, we find evidence for diminishing returns to motivation: when task motivation is high, additional sources of nonpecuniary motivation (i.e. mission) do not increase effort. However, when task motivation is low, mission reinforcement significantly increases effort, as in previous research.

The principal threat to the external validity of these findings is the possibility that the experimental design widens the gap between mission and task motivation beyond what we would observe in real-world missions and tasks. This would lead us to spuriously generalize our conclusion that mission motivation yields less effort than task motivation. The lab-in-the-field design attenuates this concern by boosting mission motivation with a real mission – donations to a school – and by showing subjects pictures of children at the school. It is still possible that the children's welfare is less motivating for our subject pool than the welfare of patients personally known to them. However, even if this is the case, the experimental task that elicits the greatest motivation – watching videos and answering multiple choice questions regarding a fictitious patient – may also be somewhat less motivating than diagnosing and treating actual patients. The difference between the experimental mission and task motivations is not obviously larger than the real-world motivations of medical professionals, and might even be smaller, such that real-world effects might even be larger. This gives us greater confidence that the behavior we document in the lab is likely to mimic the behavior of medical staff in their everyday tasks.

Nevertheless, a goal of future research should be to better capture real-world motivations in experimental settings. Our experimental design comes closer than any other with which we are familiar in finding a task that closely maps the real-world work of our subject pool. Future research should try to do the same in the context of mission motivation – to give subjects a mission that replicates the mission of their real-world work. This is challenging. For example, a reasonable conjecture is that health workers are motivated to help the individuals who come to them for assistance. One could therefore imagine an experimental design that allows some subjects to undertake a medical task that helps a physically-present patient and one that allows them to

undertake the same task, but for a remote patient. However, apart from the noise introduced by a heterogeneous patient population (we used the same actress in all our video vignettes), this design raises difficult logistical and ethical challenges.

A second goal of future research should be to design low-motivation tasks for laboratory experiments that map more naturally into the real-world tasks of actual organizations. This is again challenging, though for different reasons: repetitive tasks that are reasonably considered to be the least motivating are also most likely to be automated in a world in which machine learning and robotics are rapidly advancing.

Our results have implications for organizations seeking to utilize mission reinforcement to increase effort and productivity. We find that campaigns reinforcing organization missions are likely to yield positive impacts on effort among low-motivation tasks. Importantly, however, matching workers to tasks that motivate them seems far more important for effort than mission matching. Future work can usefully focus on task motivation as a primary driver of intrinsic motivation, complementing the growing literature on mission and pro-social motivation.

References

- Abeler, J., Falk, A., Goette, L., & Huffman, D. (2011). Reference points and effort provision. *American Economic Review*, 101(2), 470-92.
- Ariely Dan, Anat Bracha and Stephan Meier (2009). "Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially." *American Economic Review* 99(1): 544-55
- Ashraf, N., Bandiera, O., & Jack, B. K. (2014). No margin, no mission? A field experiment on incentives for public service delivery. *Journal of Public Economics*, 120, 1-17.
- Bailey, Charles D. and Nicholas J. Fessler (2011). "The moderating effects of task complexity and task attractiveness on the impact of monetary incentives in repeated tasks." *Journal of Management Accounting Research Annual* 23: 189-210.
- Bandiera, Oriana, Iwan Barankay, and Imran Rasul. 2005. "Social Preferences and the Response to Incentives: Evidence From Personnel Data." *Quarterly Journal of Economics*, 120: 917-62.
- Bandiera, Oriana, Iwan Barankay, and Imran Rasul. 2007. "Incentives for Managers and Inequality Among Workers: Evidence from a Firm Level Experiment." *Quarterly Journal of Economics*, 122: 729-74.
- Bandiera, Oriana, Iwan Barankay, and Imran Rasul. 2009. "Social Connections and Incentives in the Workplace: Evidence from Personnel Data." *Econometrica* 77: 1047-94.
- Bandiera, Oriana, Iwan Barankay, and Imran Rasul. 2010. "Social Incentives in the Workplace." *Review of Economic Studies*, 77: 417-58
- Banuri, Sheheryar, Damien de Walque, Philip Keefer, Haidara Ousmane Diadie, Paul Jacob Robyn, Maurice Ye (2017) "The use of video vignettes to measure health worker knowledge. Evidence from Burkina Faso." Mimeo.
- Banuri, Sheheryar, and Philip Keefer (2016). "Pro-social motivation, effort and the call to public service." *European Economic Review* 83: 139-164.
- Bellemare, Charles and Bruce Shearer (2009). "Gift giving and worker productivity: Evidence from a firm-level experiment." *Games and Economic Behavior* 67: 23-244.
- Benabou, R., & Tirole, J. (2003). Intrinsic and extrinsic motivation. *The review of economic studies*, 70(3), 489-520.
- Bénabou, Roland and Jean Tirole (2006). "Incentives and prosocial behavior." *American Economic Review* 95(5):1652–1678.
- Besley, Timothy and Maitreesh Ghatak (2005). "Competition and incentives with motivated agents." *American Economic Review* 95(3):616–636.
- Breuer, Ludger (2013). "Tax Compliance and Whistleblowing—The Role of Incentives." *The Bonn Journal of Economics*, 2(2) 7-44.
- Carpenter, Jeffrey, Cristina Connolly and Caitlin Myers (2008). "Altruistic behavior in a representative dictator experiment." *Experimental Economics* 11:282–298.
- Carpenter, Jeffrey, and Erick Gong. "Motivating Agents: How much does the mission matter?" *Journal of Labor Economics* 34, no. 1 (2016): 211-236.
- Carpenter, Jeffrey, and Caitlin Knowles Myers (2010). "Why volunteer? Evidence on the role of altruism, image, and incentives." *Journal of Public Economics* 94.11: 911-920.
- D’Adda, Giovanna (2011). "Motivation crowding in environmental protection: Evidence from an artefactual field experiment." *Ecological Economics* 70(11): 2083-2097
- Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125, 627–668.
- Dellavigna, Stefano (2017). "Structural Behavioral Economics." Mimeo, University of California, Berkeley.

- Eckel, Catherine C. and Phillip J. Grossman (1996). "Altruism in anonymous dictator game." *Games and Economic Behavior* 16:181–191.
- Ellingsen, Tore and Magnus Johannesson (2008). "Pride and prejudice: The human side of incentive theory." *American Economic Review* 98(3):990–1008.
- Fehr, E., & Gächter, S. (2000). Fairness and retaliation: The economics of reciprocity. *The journal of economic perspectives*, 14(3), 159-181.
- Fehr, E., Gächter, S., & Kirchsteiger, G. (1997). Reciprocity as a contract enforcement device: Experimental evidence. *Econometrica: journal of the Econometric Society*, 833-860.
- Fessler, (2003). "Experimental evidence on the links among monetary incentives, task attractiveness and task performance." *Journal of Management Accounting Research* 15(1): 161-176, December.
- Forsythe, Robert, Joel L. Horowitz, N. E. Savin and Martin Sefton (1994). "Fairness in simple bargaining experiments." *Games and Economic Behavior* 6:347–369.
- Francois, Patrick (2000). "Public service motivation' as an argument for government provision." *Journal of Public Economics* 78:275–299.
- Frey, B. S., & Jegen, R. (2001). Motivation crowding theory. *Journal of economic surveys*, 15(5), 589-611.
- Frey, B. S., & Oberholzer-Gee, F. (1997). The cost of price incentives: An empirical analysis of motivation crowding-out. *The American economic review*, 87(4), 746-755.
- Friedrichsen, Jana and Dirk Engelmann (2017). "Who cares for social image?" Working Paper.
- Gagné, Marylene and Edward L. Deci (2005). "Self-determination theory and work motivation." *Journal of Organizational Behavior* 26(4): 331-362
- Georganas, S., Tonin, M., & Vlassopoulos, M. (2015). Peer pressure and productivity: The role of observing and being observed. *Journal of Economic Behavior & Organization*, 117, 223-232.
- Georgellis, Y., Iossa, E., & Tabvuma, V. (2010). Crowding out intrinsic motivation in the public sector. *Journal of Public Administration Research and Theory*, 21(3), 473-493.
- Gill, David, and Victoria Prowse (2012). "A structural analysis of disappointment aversion in a real effort competition." *The American Economic Review* 102.1: 469-503.
- Hoffman, Elizabeth, Kevin McCabe, Keith Shachat and Vernon Smith (1994). "Preferences, property rights, and anonymity in bargaining games." *Games and Economic Behavior* 7(3):346 – 380.
- Ibanez, Marcela, and Elke Schaffland (2013). "The Effect of Outside Leaders on the Performance of the Organization: An Experiment." No. 149. Courant Research Centre: Poverty, Equity and Growth-Discussion Papers.
- Judge, T. A., Thoresen, C. J., Bono, J. E., & Patton, G. K. (2001). The job satisfaction job performance relationship: a qualitative and quantitative review. *Psychological Bulletin*, 127, 376–407.
- Li, Sherry Xin, Catherine C. Eckel, Phillip J. Grossman and Tara Larson Brown (2011). "Giving to government: Voluntary taxation in the lab." *Journal of Public Economics* 95(9-10):1190–1201.
- Pokorny, Kathrin (2008). "Pay-but do not pay too much: An experimental study on the impact of incentives." *Journal of Economic Behavior & Organization*. 66(2): 251-64.
- Prendergast, Canice (2007). "The motivation and bias of bureaucrats." *American Economic Review* 97(1):180–196.
- Reeson, Andrew F. and John G. Tisdell (2008). "Institutions, motivations and public goods: An experimental test of motivational crowding out." *Journal of Economic Behavior & Organization* 68(1): 273-281.
- Seo, H., Kim, J.Y. and Yang, S.U., 2009. Global activism and new media: A study of transnational NGOs' online public relations. *Public Relations Review*, 35(2), pp.123-126.
- Whitt, Sam and Rick K. Wilson (2007). "The dictator game, fairness and ethnicity in postwar Bosnia." *American Journal of Political Science* 51(3):655–668.

Appendix

Photos shown to subjects of children at the poor primary school (Gampela 3)



Appendix Table 1: Task and Mission Motivation, Controlling for Observables

Dependent variable: Effort (Number of rounds)			
Task	I Blank	II Slider	III Medical
Treatment: with Mission	1.391** (0.66)	0.672* (0.38)	-0.408 (0.56)
Ability (piece rate task)	-- --	-0.023 (0.02)	-0.052 (0.15)
Training: Midwife	0.437 (0.91)	-1.087 (0.78)	0.709 (0.64)
Training: Doctor	-0.847 (1.46)	-1.092 (1.54)	-1.865 (1.32)
Education level (years)	1.076 (0.82)	0.653 (1.09)	-1.099 (0.77)
Female 1 = Female	-0.923** (0.45)	-1.135* (0.64)	-1.420 (0.90)
Age (years)	-0.031 (0.05)	0.007 (0.07)	-0.130** (0.05)
Current state of personal finances 4 = Excellent	-0.191 (0.53)	0.456 (0.44)	0.794 (0.75)
Risk preferences 5 = Risk seeking	-0.203 (0.25)	0.196 (0.18)	1.554*** (0.32)
Clarity of instructions 5 = Always clear	0.201 (0.31)	0.281** (0.14)	1.106*** (0.32)
Confidence in payment to schools 5 = Strongly Agree	-0.295 (0.27)	0.26 (0.36)	0.282 (0.55)
Motivation (dictator) CFA donated to school	0.001 (0.00)	-0.001 (0.00)	0.004 (0.00)
Constant	-0.231 (2.78)	-2.186 (3.06)	3.031 (4.33)
Sigma Constant	2.410*** (0.54)	2.110*** (0.32)	4.764*** (0.36)
Pseudo R2	0.055	0.040	0.037
Log Likelihood	-172.1	-175.4	-260.7
P	.	.	.
Observations	75	81	92
Right censored observations	1	0	8

Note: * p<0.1, ** p<0.05, *** p<0.01. Dependent variable is effort (number of rounds respondent chose to continue the task). Tobit specification with upper censors at 16 (the maximum rounds subjects continue prior to auto-exit), with clustered standard errors (by day) in parentheses. Table reports regression coefficients.