5-2023

Paying for what Kind of Performance? Performance Pay, Multitasking, and Sorting in Mission-Oriented Jobs

Daniel B. Jones
University of Pittsburgh

Mirco Tonin
Free University of Bozen-Bolzano

Michael Vlassopoulos
University of Southampton

K. Pun Winichakul
Smith College, kwinichakul@smith.edu

Paying for what kind of performance? Performance pay, multitasking, and sorting in mission-oriented jobs*

Daniel B. Jones
University of Pittsburgh
daniel.jones@pitt.edu

Mirco Tonin
Free University of Bozen-Bolzano; FBK_IRVAPP; IZA; CESifo; Carlo F. Dondena Centre for Research on Social Dynamics and Public Policies, Università Bocconi
mirco.tonin@unibz.it

Michael Vlassopoulos
University of Southampton, IZA
M.Vlassopoulos@soton.ac.uk

K. Pun Winichakul
Smith College
kwinichakul@smith.edu

May 2023

Abstract: How does pay-for-performance (P4P) impact productivity and the composition of workers in mission-oriented jobs when output has multiple dimensions? This is a central issue in the public sector, particularly in areas such as education and health care. We conduct an experiment, manipulating compensation and prosocial elements of the job, to answer these questions. We find that P4P has significantly smaller positive effects on productivity on the incentivized dimension in the prosocial setting relative to the non-prosocial setting. On the other hand, P4P generates no loss in performance on the non-incentivized dimension of effort in the prosocial setting, whereas it does so in the non-prosocial setting. In both settings, P4P attracts higher ability workers, but it does so at the expense of attracting prosocially-motivated workers in the prosocial setting.

Keywords: Prosocial motivation, Performance pay, Multitasking, Sorting
JEL Codes: C91, M52, J45

* We thank Robert Dur, Josse Delfgaauw, Michael Gibbs, Sera Linardi, Stephan Meier and seminar and workshop participants at Birmingham, Bocconi, Erasmus, Innsbruck, Cambridge Judge Business School, the Science of Philanthropy Initiative 2017 conference, and the IZA World Labor Conference 2018 for helpful comments. Financial support from the Riegel and Emory Human Resources Center at the University of South Carolina is gratefully acknowledged. We thank Matthew Harvey and Xiaocheng Hu for excellent research assistance.

1. Introduction

Performance-related compensation has become an important personnel tool in a large segment of the labor market across many countries.[1] This type of compensation has been recognized to offer two benefits for firms that make use of it: incentive effects on productivity that result from aligning the interest of workers with those of the firm, and, sorting effects that arise from attracting more productive workers to firms offering performance-related pay relative to fixed pay (Lazear, 1986). While bonuses, piece rates and other forms of pay-for-performance (P4P) are more widespread in the private for-profit sector, the mission-oriented sector (public/nonprofit firms and social enterprises) is often characterized by relatively rigid pay policies in which seniority-based rules prevail over schemes rewarding performance. In recent years, there has been lively debate on whether this should change, with disagreement over the desirability of the pay of workers in the mission sector becoming more conditional on performance (Burgess and Ratto, 2003; Finan et al., 2017). Those who advocate the adoption of pay-for-performance schemes extoll the potential efficiency gains,[2] while those who are more reluctant usually highlight two aspects of the mission sector as being particularly challenging regarding implementing performance-related compensation.

The first concern is that mission-oriented organizations are characterized by a multiplicity of objectives, for instance, advancing a prosocial goal while maintaining efficiency (Dixit, 2002; Besley and Ghatak 2017). This can potentially create a conflict between the productivity measure used to determine workers' payment and a productivity measure that determines the organization's broader mission. A second aspect making P4P potentially problematic in mission-oriented organizations is that prosocial motivation plays a particularly prominent role as helping beneficiaries is an important characteristic of the job (Francois and Vlassopoulos, 2008; Besley and Ghatak, 2018; Cassar and Meier, 2018). In such organizations, P4P may have adverse effects for two reasons. First, performance may decline due to crowding out of workers' prosocial motivation by higher-powered financial incentives (Gneezy et al., 2011). Second, P4P might lead to the attraction of less prosocially motivated workers and cause the dilution of prosocial motivation in the workforce (Jones, 2015; Finan et al., 2017).

[1] See, for instance, Lemieux et al. (2009) and Bloom and van Reenen (2011).
[2] See Neal (2011) and Imberman (2015) for surveys of the literature on the effects of performance pay in education.

This paper contributes to the debate about the desirability of pay-for-performance by studying a setting where prosocial motivation and multitasking are both present. We do so by providing experimental evidence, from both an in-person and a virtual laboratory setting, of incentive and sorting effects of performance-related compensation in a context characterized by a potential conflict between the productivity measure used to determine own payment and a productivity measure that determines a payment to a pro-social entity. We design a real-effort experiment with a novel task – a word formation task inspired by the board game Scrabble – that has two output dimensions: one captured by the number of words formed and another reflecting the complexity of each word formed, as measured by the sum of the letter point values in the word.[3] We first observe how performance in the two output dimensions differs between a treatment in which compensation is a fixed wage and one in which it is a piece rate on the first dimension. We also implement a treatment that induces prosocial motivation by linking the second dimension to a charitable donation and, again, observe performance under a fixed wage and a piece rate on the first dimension. To assess selection effects, in a second phase we allow subjects to choose whether to keep the compensation structure that was imposed upon them in the prior phase, or switch to a (higher) fixed wage without a prosocial dimension instead. This latter analysis sheds light into any potential dilution in the underlying prosocial motivation of the workforce on the extensive margin as a consequence of P4P, as noted above.

We find the following main results. In the standard setting without a prosocial element, our experiment reproduces the classical results of the literature on incentives, with the piece rate increasing performance in the incentivized dimension while reducing effort along the unincentivized one. We also find that higher ability workers self-select into the piece rate compensation scheme, therefore generating an additional positive effect of P4P on productivity due to sorting (Lazear, 2000). These results establish a baseline against which we compare results from the prosocial setting, which is the primary contribution of our paper. In the environment with a prosocial dimension, we instead find a much smaller incentive effect due to P4P while sorting of higher ability workers into P4P also takes place. Further, we find that people with high prosocial motivation are willing to give up financial gains to stay in the prosocial setting with fixed pay. Our results imply that P4P is less effective in increasing performance on the incentivized dimension when there is a prosocial component. On the other hand, the negative consequences on the unincentivized dimension are also not present. Our experimental design enables us to decompose

[3] For a review of alternative methods of measuring effort in economic experiments see Charness et al. (2018).

the overall effect of P4P on productivity into a sorting effect (the fact that individuals who select to work under P4P are more productive) and an incentive effect (the productivity change induced by P4P), and we find that, broadly speaking, they are of similar size.

This paper explores the interaction between financial incentives, prosocial motivation, and performance in an environment with multitasking. As such, it intersects with different strands of literature. We outline our contribution relative to existing work below.

First, we build on the literature on incentives that has long recognized the effect of incentive pay in multitasking environments (Holmstrom and Milgrom, 1991). The empirical evidence is mixed, with some studies reporting findings in line with the standard neoclassical theoretical prediction (e.g., Al-Ubaydli et al., 2015; Hong et al., 2018), while others finding no adverse effect due to piece rate compensation on the unincentivized dimension (Shearer, 2004; Copeland and Monnet, 2009). The issue has attracted considerable attention with regard to health care contracting, where multitask agent problems are indeed ubiquitous (Chalkley and Malcomson, 1998; see Scott et al., 2011, for a review). [4] This literature has also underlined the importance of sorting effects for incentive pay (Lazear, 1986), with empirical evidence supporting the theoretical findings (Lazear, 2000). Dohmen and Falk (2011) show in a lab experiment the importance of multidimensional sorting of workers along the risk preferences and self-assessment dimensions. We contribute to this literature by providing evidence on the impact of the introduction of pay-for-performance on the allocation of effort in an environment with *both* multitasking and a prosocial mission.

Second, we connect to the growing experimental literature that documents the importance of prosocial motivation and incentives for performance (Tonin and Vlassopoulos, 2010, 2015; Imas, 2014; Charness et al., 2016; Carpenter and Gong, 2016; DellaVigna and Pope, 2018; Kajackaite and Sliwka, 2017; Cassar, 2019; Briscese et al., 2021). For instance, Carpenter and Gong (2016) find that piece rate incentives raise workers' performance and are more effective when people are not working for a mission that they like, highlighting that mission-match and financial incentives can be substitutes. Imas (2014) and Charness et al. (2016) find that higher piece rates work better in motivating workers when they are paid to workers themselves than donated to a charity, whereas small piece rates have the opposite effect. We contribute to this literature by studying how prosocial

[4] For example, Douven et al. (2019) studies a situation of multitasking in the provision of mental health care and finds that more altruistic providers provide higher quality care.

motivation and pay-for-performance interact to affect performance in a multitasking environment where piece rate contracts provide incentives in one task but not in another.

Finally, a related strand has been concerned with the effect that financial incentives may have on the selection of motivated workers (Francois, 2007; Delfgaauw and Dur, 2007, 2010; Prendergast, 2007; Dal Bo et al., 2013; Banuri and Keefer, 2016; Ashraf et al., 2020; Barigozzi et al., 2018; Deseranno, 2019). A recent study in Uganda (Deseranno, 2019) finds that in a context where the job characteristics are not known to applicants, stronger financial incentives attract less prosocial applicants because they signal that the job does not have a strong prosocial orientation. Ashraf et al. (2020) find that highlighting the career benefits of a new civil service position in Zambia attracted applicants with lower prosocial motivation. We add to the limited evidence of sorting effects by ability and prosocial motivation in the mission sector. In particular, we are able to examine whether sorting in an environment with prosocial motivation is affected by the presence of performance pay, which to the best of our knowledge has not been empirically investigated before.

To summarize, relative to the existing literature our paper makes three main contributions: first, we provide evidence on the incentive and sorting effects of introducing pay-for-performance in an environment characterized by multitasking and a prosocial dimension. Second, we examine the incentive effect of performance pay on the incentivized and non-incentivized dimensions of *the same job*, both in a context where prosocial motivation may be active and when it is not, allowing us to isolate the differential effects of pay-for-performance in these two settings. Third, we add to the evidence on how performance pay influences who sorts into prosocial settings in terms of ability and prosocial motivation, contrasting the case in which there is pay for performance to the case when there is not. Taking this all together, our paper is the first with the ability to directly compare the impacts of performance pay across a prosocial setting and a standard setting, holding job/task characteristics constant – while also allowing for an analysis of sorting and multitasking. This setting yields the novel insights that performance pay does not pull prosocially motivated workers from an unincentivized task to the same degree that it does non-prosocial workers, but that performance pay may attract fewer prosocially motivated workers into the mission-oriented sector than would occur under flat pay.

The rest of this paper is organized as follows: the next section describes the experimental design, while section 3 lays out a simple theoretical framework and derives behavioral predictions. In

section 4 we present the experimental results. Section 5 offers a discussion of our results and section 6 some concluding remarks. The Appendix includes additional tables and the experimental instructions.

2. Experimental Design

The experiment involves a real effort task performed in three phases followed by a questionnaire administered at the end. Participants were informed about the nature of the task and the structure of the experiment at the beginning but received detailed instructions about each phase only at the start of the respective phase. The end-of-experiment questionnaire (in the Appendix) included incentivized questions eliciting risk preferences based on the instrument developed by Holt and Laury (2002), a dictator game with a chosen local charity as the recipient, demographic characteristics and a set of questions proposed by Perry (1996) measuring public service motivation (PSM). In what follows, we describe the novel word-formation task, the phases of the experiment, the treatments, and some procedural details.

2.1 Real Effort Task

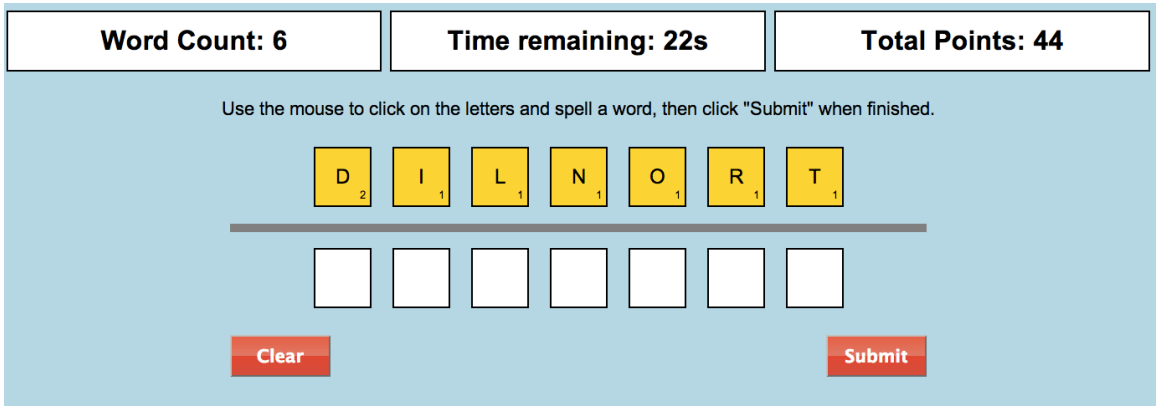
We employ a novel word-formation real effort task. Participants are presented with a set of seven letters, which they must use to spell a word using all or a subset of the letters (at least two). When a participant spells and submits a word a new set of letters appears on the screen. Letters have points attached to them that are indicated in the lower right-hand corner of each letter square, like – and using the same point values as – the real-life word game Scrabble (see Figure 1 for an example screen).[5] The points reflect the relative rarity of each letter and difficulty of constructing words using those letters. This opens the possibility to measure participants' performance on the task along two dimensions: (i) number of words completed, and (ii) points per word. Only words that are grammatically correct are permissible. If the submitted word is incorrectly spelled, the chosen letters are cleared, and the participant is asked to try again. Participants receive feedback about the number of words spelled, the total points earned and the time remaining at the top of the screen.

The main features of the task that are crucial for our purposes is that it allows us to measure multiple dimensions of production, number and complexity of words, and that the number of words that can

[5]An example of the Scrabble board game instructions can be found using the following link: https://service.mattel.com/instruction_sheets/53639-ENG.pdf

be formed in a given period of time is reduced by searching for words formed using rarer (and often more) letters that hence bear more points.[6] Depending on the compensation and motivational structure, workers may face a trade-off between these two dimensions. Note that to ensure that participants face a meaningful trade-off between spelling a large number of words and accumulating points, we place a restriction where the points associated with a certain word are credited only if the total points of the word are at least five.[7]

Figure 1: Screenshot of Word Formation Task



2.2 Phases

The experiment involves three phases aimed at measuring ability, the incentive effect, and the sorting effect of the incentive schemes.

- Phase 1 (Ability measurement): The purpose of this phase is to familiarize participants with the task and to elicit measures of ability on the task. We use a series of five steps to capture different aspects of ability on the task, described in more detail below.
- Phase 2 (Incentive effect measurement): Participants complete the word-spelling task for eight minutes. Payment differs by treatment (described below).

[6] This task allows for a much simpler and more precise measure of multiple dimensions of output compared to other tasks used in the experimental multitasking literature, such as packing envelopes (Al-Ubaydli et al., 2015), planting trees (Shearer, 2004), or various manufacturing tasks (Hong et al., 2018).

[7] Using the example in Figure 1, admissible words like "NO", "IN", "DO", "ROT", "TORN", "LOIN" and so on, all generate 0 counted points, while words that are more difficult to find like "LORD", "DIRT", "NITRO" or "RIND" generate 5 points. This introduces a trade-off between finding many words and generating points.

- Phase 3 (Sorting effect measurement): Participants spend another eight minutes completing the word task. Before starting, they can choose to continue with the treatment-assigned payment scheme (as they saw in Phase 2) or move to an outside option (common across treatments) in which they would receive a flat payment of $9. In this phase, we are primarily interested in the choice of payment scheme that participants make.

Final payment for participants consisted of a show-up fee, plus their payoff from the first phase and the payoff of a randomly chosen of the two other phases, plus any payoff arising from the additional decisions at the end (risk preference measurement and dictator game).[8]

2.3 Ability Measurement

The first phase of the experiment consisted of five short tasks aimed at measuring participants' task-relevant ability. The tasks faced in this phase – as well as in the other phases – were constant across treatments for the sessions conducted in the lab, while for the online sessions they were constant only for tasks 2 and 4 of Phase 1, due to technical issues. Treatment differences in the incentives participants would face in the following phases were not yet described. It is therefore a pure measure of ability, free of any potential treatment effects, which is directly comparable across all participants.

The five tasks were as follows: (i) complete five word formation jobs in two minutes (no compensation), (ii) complete five jobs within 60 seconds (participants start with an endowment of experimental currency units and are penalized the longer it takes to complete the five jobs), (iii) complete as many word formation jobs as possible in three minutes (receive piece rate for each word), (iv) complete five word formation jobs (compensation based on score accumulated for words worth five or more points), (v) complete as many word formation jobs as possible in three minutes (compensation based on score accumulated for words worth five or more points).

The first of the five tasks in Phase 1 was non-incentivized and was primarily meant to provide participants with practice using the interface. Tasks 2 and 3 incentivized participants to spell many

[8] As mentioned previously, sessions were run both in an in-person from 2014-2017 and a virtual laboratory setting (in 2021), drawing from two different subject pools. In in-person sessions participants received a $5 show-up fee, while in virtual sessions with the Pittsburgh Experimental Economics Lab participant pool, participants received a $6 show-up fee in accordance with the lab rules.

words in a short period of time, thereby measuring participants' abilities in the first dimension. The remaining two tasks incentivized participants to accumulate as many points as possible in a short period of time, thereby measuring participants' abilities in the second dimension. Based on earnings on tasks 2-5 we construct a composite ability measure. Specifically, we first standard-normalize participants' earnings in each task (as they are not otherwise measured in the same units), sum standard-normalized earnings across tasks 2-5, and standard-normalize again.[9]

2.4 Treatments

We implement a 2*2 between-subject design, depicted in Figure 2, in which we manipulate two aspects of Phase 2: the incentive scheme (flat or piece rate) and whether there is an opportunity to raise money for a charity (we refer to the presence of a charity as *prosocial* and absence as *non-prosocial*). For the treatments involving a charity, participants select a recipient from a list of local charities with diverse missions (see the Experimental Instructions in the Appendix for the list). In particular, the treatments are as follows:

- F-NP (flat-rate, non-prosocial): participants are paid $7 for working during Phase 2 regardless of the number of words spelled (or points accumulated). Together with the compensation associated with the other phases, this amount ensures that participants receive a fair compensation given the duration of the experiment. Given that they do not face incentives conditional on performance, participants should play in this treatment in the way they find most enjoyable.
- P-NP (piece-rate, non-prosocial): participants receive $0.10 per word spelled. We chose this piece rate to have, on average, a compensation similar to the one received in the flat-rate treatment. Indeed, in this treatment on average participants accumulated earnings of $5.84 (see Table 3).
- F-P (flat-rate, prosocial): participants are paid $7 for working during Phase 2 regardless of the number of words spelled (or points accumulated). In addition, participants generate $0.02 for a local charity of their choosing for each *point* that they accumulate, with points associated with each word credited only if they are five or greater. We chose this rate to have, on average, a donation that is substantial, but smaller than own compensation. Indeed, in this treatment the charitable donation is on average $4.15 (see Table 3). Notice

[9] To account for the difference in experimental environments, we standard-normalize *within* experiment type (in-person lab vs. virtual lab), though results are similar if we do not.

that to contribute to charity in a significant manner, a participant needs to produce many words of sufficiently high complexity. Excluding from the count words with less than five points strengthens the trade-off between forming a lot of words and forming words that yield many points.

- P-P (piece-rate, prosocial): participants receive \$0.10 per word spelled. In addition, participants generate \$0.02 for a local charity of their choosing for each point that they accumulate, with points associated with each word credited only if they are five or greater.[10] Considering that on average participants accumulate 5.62 points per word, the two incentives have roughly similar power.

Figure 2: Treatment Conditions

		Setting: Non-prosocial	Setting: Prosocial
Incentive Scheme	Flat rate	F-NP flat-rate, non-prosocial (N=61)	F-P flat-rate, prosocial (N=61)
	Piece-rate	P-NP piece-rate, non-prosocial (N=61)	P-P piece-rate, prosocial (N=61)

We summarize the timeline of the experiment across these four treatments in Figure 3.

In light of this parametrization, the \$9 flat payment we offer as an outside option in Phase 3 implies a concrete trade-off for F-P and also, on average, for P-NP and P-P, while it of course dominates F-NP. It is noteworthy that to reduce potential confusion about the purpose of having points

[10] It is possible that participants could maximize their monetary earnings in the experiment and donate to a charity outside the lab. This would reduce the difference between P-P and P-NP. Factors like transaction costs or mental accounting, however, would lead subjects to prefer donations made by the experimenter.

assigned to letters in the two non-prosocial treatments, we opted to tell participants in all treatments: "We are hoping to learn how many points people can accumulate in a fixed period of time".[11] This aligns expectations about the objective of the principal across treatments and aims to capture real workplace conditions involving performance trade-offs where, while one dimension of performance may be directly incentivized, it is usually understood by workers that the employer is also concerned about other dimensions.

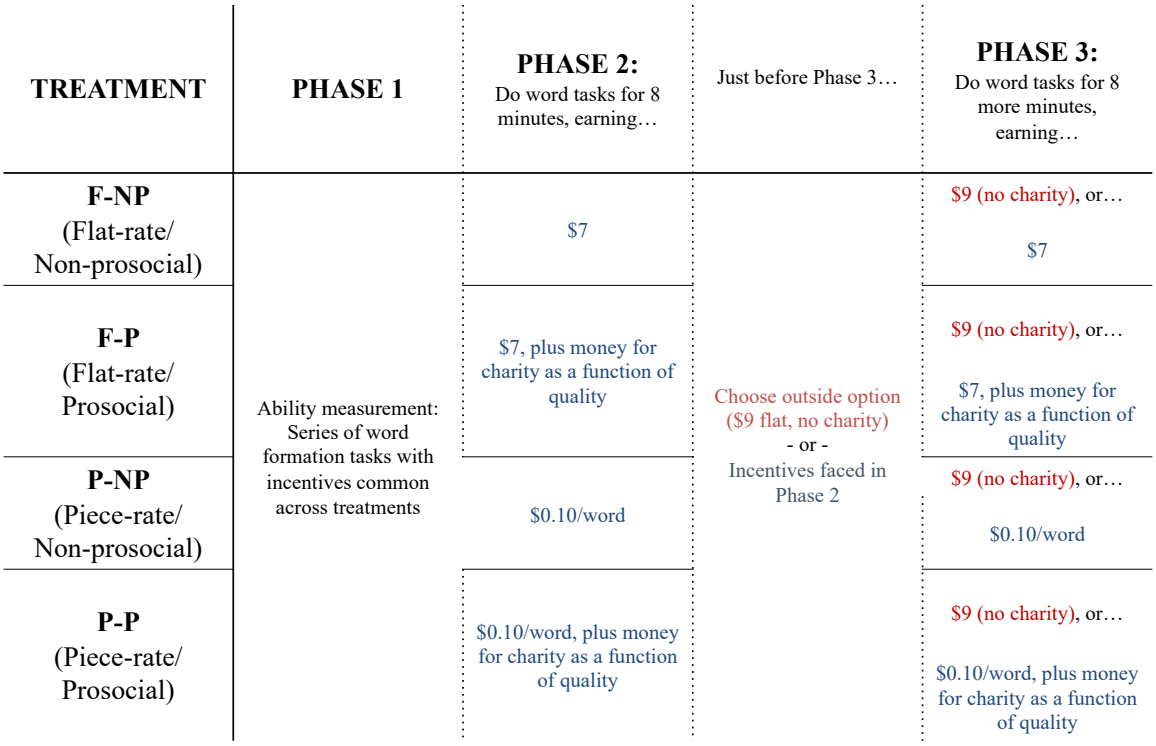
2.5 Procedures

Half of the sample was collected in-person in experimental sessions conducted in the Behavioral Lab at the University of South Carolina, while the other half was collected online through the Pittsburgh Experimental Economics Lab (PEEL) at the University of Pittsburgh. Payments were expressed in terms of experimental currency units (ECUs), which had an exchange rate of 10 ECUs to $1. We conducted 15 in-person sessions in total, with roughly 8 participants per session on average.[12]

The mean earnings across all participants were roughly $16. The typical session lasted approximately 75 minutes. In addition, we conducted 8 online sessions with the PEEL participant pool, with an average of 15 participants per session. Typical online sessions also lasted 75 minutes, and participants earned roughly $20. To best mirror the conditions of the in-person laboratory in the online sessions, we followed procedures outlined in Danz et al. (2021). In the in-person sessions, participants were recruited largely from introductory (first- and second-year undergraduate-level) economics courses from the business school at the University of South Carolina. Participation was voluntary and was not linked to grades in their introductory courses. For the online sessions, participants were recruited through the Pittsburgh Experimental Economics Laboratory recruiting database. Students in this participant pool represent a wider array of fields of study as they are recruited from introductory courses outside of economics, in addition to introductory economics courses.

[11] A full set of Experimental Instructions is reproduced in Appendix C.
[12] Because there was no strategic interaction between participants, the sessions did not require a particular number of participants to run. The number of participants varied across sessions, but the average number of participants per session is relatively constant across treatments.

Figure 3: Summary of the timeline of the experiment



3. Predictions

This section presents a simple theoretical framework to generate theoretical predictions. Consider an agent who exerts effort on two dimensions $e = (e_1, e_2)$. For simplicity, we assume that effort leads to outcomes linearly. Let us further assume that the cost of effort for the agent $c(e_1, e_2)$ is increasing and strictly convex with $\frac{\partial^2 c(e_1,e_2)}{\partial e_1 \partial e_2} > 0$, that is, the two efforts are substitutes. Furthermore, we assume that $c(e_1, e_2)$ attains a local minimum at an interior point $e^* = (e_1^*, e_2^*)$, so that, even without incentives the agent will exert some positive effort in the two tasks.

Agents are heterogeneous in two dimensions: ability on the task, denoted by $\gamma \geq 0$, and prosociality, denoted by $\theta \geq 0$. We assume that ability enters the cost function multiplicatively, $\frac{1}{\gamma} c(e_1, e_2)$.

Agents facing flat-rate compensation therefore maximize:

$$w + m\theta e_2 - \frac{1}{\gamma}c(e_1, e_2)$$

where *w* is the flat monetary rate received regardless of effort and *m* is an indicator variable equal to 1 if the worker is prosocial.

Similarly, agents facing piece-rate compensation maximize:

$$pe_1 + m\theta e_2 - \frac{1}{\gamma}c(e_1, e_2)$$

where *p* is the piece-rate that dictates how effort on the e_1 dimension is translated into payment.

We can use this simple framework to derive some predictions as to the relative levels of effort exerted in the various treatments, summarized in Table 2.

3.1 Phase 2

3.1.1 Flat-rate Non-Prosocial (F-NP)

In the absence of any pay-for-performance in the non-prosocial condition, the agent will choose $e^* = (e_1^*, e_2^*)$, given by $\frac{\partial c(e_1,e_2)}{\partial e_1} = \frac{\partial c(e_1,e_2)}{\partial e_2} = 0$. Effort on either dimension is not expected to vary with ability or prosociality. This treatment forms a baseline against which we can compare the other treatments below to form predictions.

3.1.2 Piece-rate Non-Prosocial (P-NP)

In this treatment, the agent is offered a piece rate p, in the e_1 dimension. The first-order conditions are given by $\frac{\partial c(e_1,e_2)}{\partial e_1} = \gamma p$ and $\frac{\partial c(e_1,e_2)}{\partial e_2} = 0$. Given that e_1 and e_2 are substitutes, it follows that the agent will increase effort in the e_1 dimension at the expense of e_2 and that effort in the e_1 dimension will be increasing in ability. We thus predict that (proofs in the appendix):

Prediction 1a: $e_1^{FNP} < e_1^{PNP}$ *and* $e_2^{FNP} > e_2^{PNP}$

Prediction 1b: $\frac{de_1^{PNP}}{d\gamma} > 0$

3.1.3 Flat-rate Prosocial (F-P)

In this treatment, we assume that a prosocial agent will derive an intrinsic benefit from producing e_2, which we denote by θe_2.[13] The agent thus maximizes $\theta e_2 - \frac{1}{\gamma}c(e_1, e_2)$, with first-order conditions given by $\frac{\partial c(e_1,e_2)}{\partial e_1} = 0$ and $\frac{\partial c(e_1,e_2)}{\partial e_2} = \gamma\theta$. Given the substitutability of effort we thus predict that the agent will increase effort in the $e_2$dimension at the expense of e_1:

Prediction 2a: $e_1^{FP} < e_1^{FNP}$ *and* $e_2^{FP} > e_2^{FNP}$

Prediction 2b: $\frac{de_2^{FP}}{d\gamma} > 0$

3.1.4 Piece-rate Prosocial (P-P)

In this treatment, the agent receives a piece-rate (p) for $e_1$and derives an intrinsic benefit θ from producing e_2. The agent thus maximizes $pe_1 + \theta e_2 - \frac{1}{\gamma}c(e_1, e_2)$, with first-order conditions given by $\frac{\partial c(e_1,e_2)}{\partial e_1} = \gamma p$ and $\frac{\partial c(e_1,e_2)}{\partial e_2} = \gamma\theta$. We thus predict that:

Prediction 3a: $e_1^{PP} > e_1^{FP}$ *and* $e_2^{PP} < e_2^{FP}$

Prediction 3b: $\frac{de_1^{PP}}{d\gamma} > 0$ *and* $\frac{de_2^{PP}}{d\gamma} > 0$

We can also show that in the two piece-rate treatments effort in the $e_1$dimension is higher in the non-prosocial condition than in the prosocial one, and vice-versa for the $e_2$dimension:

Prediction 4: $e_1^{PNP} > e_1^{PP}$ and $e_2^{PNP} < e_2^{PP}$

To summarize, we predict that effort in $e_1$will be highest under P-NP and lowest in F-P, while the reverse order will be true for effort in the $e_2$dimension.

[13] A similar result would arise if the agent cares about efficiency, with a weight θ.

Finally, one additional object of interest is the *difference* in the impacts of performance pay across the prosocial and non-prosocial conditions, or the direction of the inequality in $\left[e_d^{PNP} - e_d^{FNP}\right]$ $vs.$ $[e_d^{PP} - e_d^{FP}]$, in either efforts in e_1 or e_2 (i.e., $d \in \{1,2\}$). Our theoretical model does not provide a firm prediction on the direction of this inequality without additional assumptions and structure; this is therefore left as an empirical question, further motivating the need for an experiment.

3.2 Phase 3

We next generate predictions about choice of payment scheme in Phase 3 (all proofs are in the Appendix).

Flat-rate Prosocial (F-P): For a given level of ability γ sorting on prosociality type will take place, as agents with sufficiently high levels of prosocial motivation would choose F-M.

Piece-rate Non-Prosocial (P-NP): There will be sorting on ability, that is, the proposed scheme is more likely to be chosen by high ability types.

Piece-rate Prosocial (P-P): The proposed scheme is more likely to be chosen by high θ and/or γ agents. Note also that ability and prosociality are substitutes for sorting, that is, a higher ability individual requires less prosociality to sort and similarly a more prosocial individual requires less ability to find sorting in optimal.

4. Results

This section reports results from our experiment. We start by providing simple summary statistics. Table 1 provides a summary of participant characteristics (Table A1 in Appendix A presents a summary of these characteristics by treatment).[14] We also report other survey and decision task responses in the table. Females are slightly overrepresented in our sample and almost half of the sample declared a business field as their major.

[14] In Table A1, we note that 6 out of 90 pairwise tests (<7% of tests) for differences in participant characteristics across treatments are statistically significant. This share of statistically significant tests is what we might expect to randomly appear as statistically significant at the 5-10% level.

As we test for heterogeneity in participants' responses along both the ability dimension (using the composite ability measure described in the previous section) and along the prosociality dimension (using the dictator donation share - DDS - decision made at the end of the experimental session), it is worth documenting the relationship between these two measures in our setting. Figure 4 graphically depicts the relationship between the measures, with Ability on the x-axis and the DDS on the y-axis.

Table 1: Summary statistics – Participant characteristics

Female	0.57	Any monetary donations?	0.69
	(0.50)	(past year)	(0.46)
Age	19.67	Any goods donations?	0.77
	(1.76)	(past year)	(0.42)
Year in college	2.30	Any volunteering?	0.68
	(1.05)	(past year)	(0.47)
Business major	0.46	Any blood donations?	0.27
	(0.50)	(past year)	(0.45)
Native English?	0.92	Holt-Laury: Share of risky	0.50
	(0.27)	choices	(0.19)
Currently works?	0.65	Dictator: Share contributed	0.45
	(0.48)		(0.31)
Current wage (conditional	8.50	PSM Survey score	0.69
on current work)	(12.28)	(0-1, higher = more PSM)	(0.09)
Observations: 244			

Standard deviations in parentheses

The average dictator donation share, 0.45, is reported in Table 1; as usual for this type of decision, participants tend to choose round numbers like 20 or 50%. The average of the ability measure is, by construction, zero. Figure 4 shows that there is substantial variation in both the DDS and ability measures across participants. Importantly, there does not appear to be any clear relationship between DDS and ability in our setting (corr.=-0.02, p=0.75).[15] Any potential correlation between ability and prosociality obviously impacts the degree to which we may expect, and the degree to

[15] Dal Bo et al. (2013) find instead that ability and prosocial motivation are positively correlated in a sample of applicants to public sector jobs in Mexico.

which it is desirable to observe, high ability workers to sort into the prosocial payment scheme with performance pay. The fact that these measures are independent in our setting helps us separately identify sorting of workers along these two dimensions.

The remainder of the section assesses differences in participant behavior across treatments. First, we will present the basic averages and the distributions of outcomes. Second, we will use regression analyses to statistically test for treatment effects in Phase 2 of the experiment in response to randomly assigned payment schemes; that is, we will study the "incentive effects" of performance pay in tasks with some prosocial element vs. not. Finally, we will turn to exploring the types of participants that opt for the randomly assigned payment scheme over a common outside option, assessing "sorting effects" of performance pay.

Figure 4: Scatterplot depicting the relationship between the composite ability measure and the dictator donation share

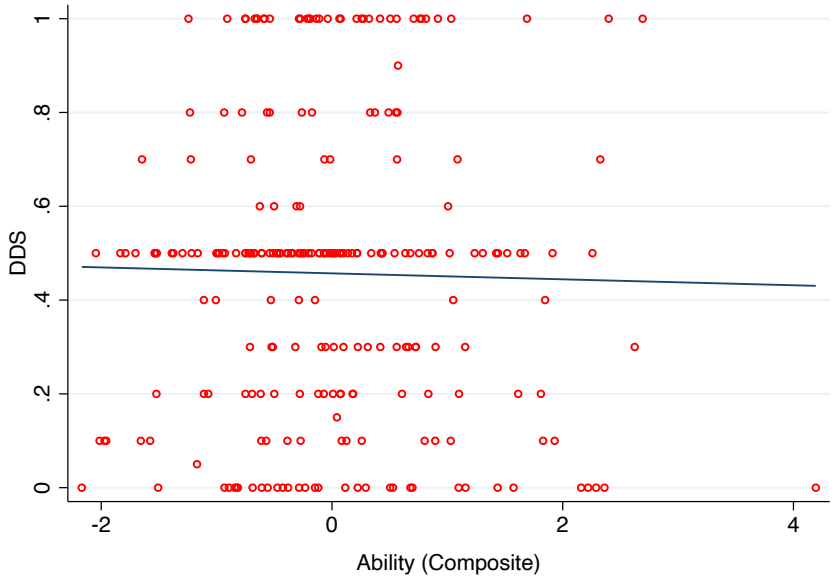


Figure note: The solid blue line is a linear fit.

4.1 Averages of outcomes across treatments

Before presenting the experimental results, for convenience, Table 2 below summarizes the main theoretical predictions concerning Phase 2.

Table 3 reports the averages of our two main outcome variables, total words spelled and points per word, across the four treatments, while Figure 5 depicts the full distribution.[16] As a point of reference, we note that the maximum number of words recorded is 136, the highest average points per word is 9.27, and the maximum total points is 517. Three broad findings are readily observable from Table 3 and Figure 5.

Table 2: Summary of effort predictions in Phase 2

Prediction 1a(#), Prediction 3a(##)	$e_1^{PNP} > e_1^{FNP}, e_1^{PP} > e_1^{FP}$
Prediction 2a($), Prediction 4($$)	$e_2^{FP} > e_2^{FNP}, e_2^{PP} > e_2^{PNP}$
Prediction 1b, Prediction 3b	$\frac{de_1^{PNP}}{d\gamma} > 0 \; and \; \frac{de_1^{PP}}{d\gamma} > 0$
Prediction 2b, Prediction 3b	$\frac{de_2^{FP}}{d\gamma} > 0 \; and \; \frac{de_2^{PP}}{d\gamma} > 0$
Unnumbered – immediate from framework	$\frac{de_2^{FP}}{d\theta} > 0 \; and \; \frac{de_2^{PP}}{d\theta} > 0$

The comma separating two quantities indicates that they cannot be unambiguously signed. The superscripts in the first column are included as references to the corresponding comparisons made in Table 3.

First, performance on the first dimension (number of words spelled) is much higher in the piece-rate non-prosocial treatment (P-NP) than any other treatment. The number of words spelled is somewhat higher in the piece-rate prosocial (P-P) treatment than the prosocial setting with flat-rate (F-P), but the difference is much smaller than the difference between the two non-prosocial treatments. We explore the statistical significance of these differences in the next subsection, but it is already clear that performance pay has a larger positive impact on the first dimension of performance in the non-prosocial treatments than in the prosocial treatments.

Second, performance on the second dimension (as measured by the average number of points associated with each word) is lower in the P-NP treatment than in any other treatment and is relatively similar across the remaining three treatments, as can be seen in the second row of Table 2. For our reported outcome of "Points per word," we divide total points accumulated (even for

[16] Table 3 reports outcomes from Phase 2, where participants are randomly assigned to a treatment. We report outcomes from Phase 3 in Appendix Table A2, reported separately for all participants and the subset who selected into the treatment-assigned payment scheme.

words worth less than five points, which are worth zero points from the participant's perspective) by total words spelled. As an alternative measure, we also report the share of words that are counted. Other alternative measures for the second dimension of performance that are similar to points per word are "Counted points per word" (number of points from words worth at least five points divided by total words) and "Counted points per counted word" (number of points from words worth at least five points divided by number of words worth at least five points). Because patterns of results are similar across these different measures, we report the simplest measure ("Points per word") in the main text moving forward and present the other measures in the Appendix (see Table A4).

Table 3: Averages of outcome variables by treatment

TREATMENT:	Piece-rate Non-Prosocial (P-NP)	Flat-rate Non-Prosocial (F-NP)	Piece-rate Prosocial (P-P)	Flat-rate Prosocial (F-P)
Phase 2: Words spelled	58.44$^{(\#)}$	36.23$^{(\#)}$	40.84$^{(\#\#)}$	33.89$^{(\#\#)}$
	(3.51)	(1.92)	(2.42)	(1.41)
Phase 2: Points per word	5.70$^{(\$\$)}$	6.42$^{(\$)}$	6.43$^{(\$\$)}$	6.77$^{(\$)}$
	(0.15)	(0.14)	(0.10)	(0.10)
Phase 2: Share of words that are counted (>4 pts.)	0.68	0.81	0.81	0.85
	(0.03)	(0.02)	(0.02)	(0.01)
Phase 2: Earnings	5.84	7.00	4.08	7.00
	(0.35)	(0.00)	(0.24)	(0.00)
Phase 2: Total Counted Points	236.98	210.23	224.28	207.67
	(8.92)	(10.82)	(10.71)	(7.91)
Phase 2: Charitable cont.	0.00	0.00	4.49	4.15
	(0.00)	(0.00)	(0.21)	(0.16)
Phase 3: Chose treatment-assigned payment?	0.38	0.02	0.49	0.54
	(0.06)	(0.02)	(0.06)	(0.06)
Observations	61	61	61	61

Standard deviations in parentheses. The superscripts are included as references to the corresponding predictions described in Table 2. $^{\#}p<0.01$; $^{\#\#}p=0.01$; $^{\$}p=0.04$; $^{\$\$}p<0.01$.

Recall that in the "prosocial" treatments, more money for charity is generated by spelling words worth more points, while individual payment is fixed or depends solely on the number of words. The points per word dimension of effort appears to be lowest in P-NP, where effort on the number of words dimension is highest, indicating that participants indeed faced a trade-off between number of words and points per word. They acted on this trade-off to a greater degree in the P-NP treatment than in the P-P treatment, where number of words is higher than in F-P, but with no large drop in points per word.

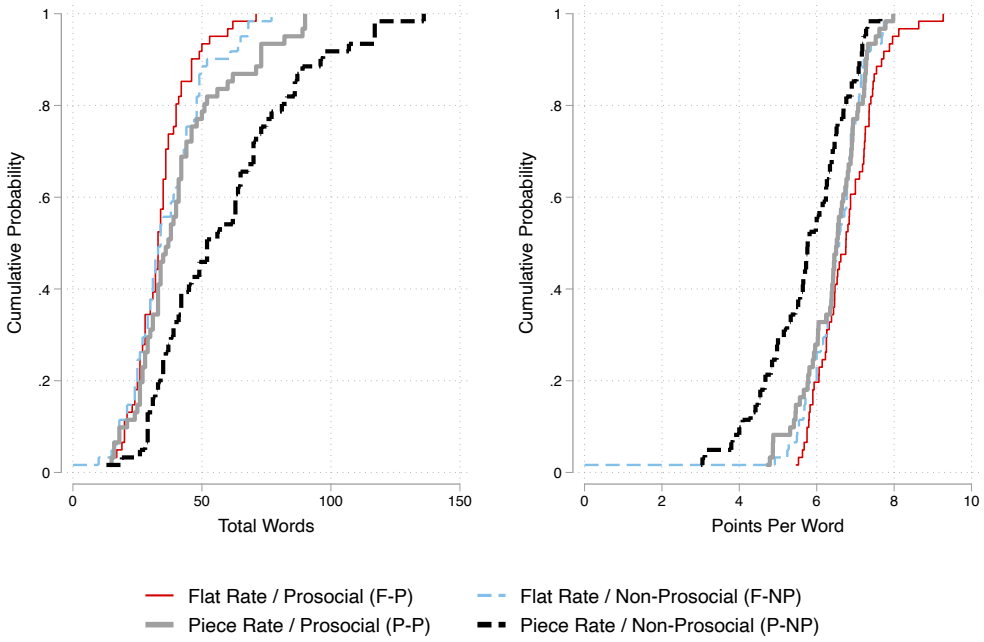
When considering the total counted points generated for a charity, we see that P-NP yielded slightly more points than P-P, on average, albeit the difference is not statistically significant. This suggests that for a principal who values both the number of words and the total points counted, there is no trade-off when choosing P-NP as an incentive scheme. However, if we focus on performance in phase 3 where participants self-select into the incentive scheme, we instead find a noticeable trade-off (see Table A2 in the Appendix) as P-P and F-P generate more total points counted than P-NP.[17]

Finally, the last row of Table 3 reports the choice of payment scheme that participants opted for in Phase 3. We report the share of participants who opt for the treatment-assigned payment scheme (e.g., piece-rate with money raised for charity) over the common outside option (flat-rate $9, with no opportunity to raise money for charity). Note that the $9 flat rate from the outside option is higher than the $7 flat rate from the F-P and F-NP treatment-assigned payment schemes. It is not surprising then to see that all but one of the participants in the F-NP treatment-assigned payment scheme preferred the outside option. More interesting is the fact that a large share of participants (54%) in the F-P treatment opted for the treatment-assigned payment scheme in Phase 3; that is, more than half of participants in that treatment were willing to accept a lower flat rate for the opportunity to raise money for charity. Lower rates of participants opted for the treatment-assigned payment schemes in the two piece-rate treatments. The rate of participants staying in the treatment-assigned scheme is relatively higher in P-P than in P-NP, which may not seem surprising given the differences in behavior, discussed above, in Phase 2 across the two treatments. In a later subsection, we explore *who* is opting for the treatment-assigned option and see that there are differences across the two treatments that run deeper than simply the number of participants choosing each option.

We further summarize our two key measures from Table 3 in distribution plots. In Figure 5 we provide cumulative distribution plots of words spelled and points per word, respectively. A central point we will return throughout the paper is evident in these figures: consistent with previous research, the distribution of effort and output in the non-prosocial treatments is substantially different when performance pay is in place (P-NP vs. F-NP). The same cannot be said of the difference between piece-rate and flat-rate prosocial treatments (P-P vs. F-P); differences are smaller and non-existent for a large range of the data. We formally test our theoretical predictions in the following section.

[17] The difference between P-P and P-NP is not statistically different in the full sample, but is in the subset of participants who self-selected into the treatment-assigned payment scheme.

Figure 5: Cumulative Distribution Function Plots of Phase 2 Outcomes



4.2 Assessing incentive effects in Phase 2

Table 4 presents regressions that allow us to test for statistical differences in effort on the first (number of words) and second (points per word) dimensions across treatments in Phase 2. In this table, we regress an outcome (e.g., in Column 1 total words spelled) on treatment indicator variables. F-NP is the omitted treatment, so all tests are relative to that treatment. The bottom of the table reports noteworthy tests of differences between other coefficients in the table (e.g., whether the P-P coefficient is significantly different than the F-P coefficient). To account for differences across the online and in-person experiment settings, we also include (but do not report) a dummy variable indicating whether the session was run online or not in all regressions in this table and those that follow.[18] In Appendix Table A3, we report similar regressions that include participant-level controls (gender, the Phase 1 ability measure, etc.). Appendix Table A5 reports results from nonparametric tests of differences across treatments for our two main outcome

[18] We pool observations across settings as participants are randomly allocated to treatments within each sample and differential behaviour across samples does not affect treatment effects.

variables (number of words spelled and points per word). Results in both tables are consistent with the results reported in Table 4.

We frame this discussion around the theoretical predictions. Theoretical Prediction 1a speaks to the comparison between piece-rate and flat-rate non-prosocial treatments -- the comparison between treatments P-NP and F-NP. The theoretical prediction suggests that number of words should be higher in P-NP than F-NP, while points per word should be lower as this dimension is sacrificed to increase number of words. We find clear evidence in line with both parts of this prediction. Column 1 shows that participants in the P-NP treatment spell an additional 22.32 words relative to the F-NP treatment. Recall that the mean of "words spelled" in the F-NP treatment was 36.23, so piece-rate payment substantially increases the number of words spelled. The simple comparison of averages suggested that this increase in number of words comes at the expense of points per word, and the statistical tests for differences in Table 4 bear that out: points per word is significantly lower in P-NP.

RESULT 1: In the non-prosocial treatments, piece rate increases the number of words solved compared to fixed pay. We also find that piece rate decreases points per word compared to fixed pay.

Theoretical Prediction 2a focuses on the difference between flat-rate payment in prosocial vs. non-prosocial treatment, or F-P vs. F-NP. We predicted a higher number of words in F-NP, but higher points per word in F-P. The basic intuition was that, in order to increase effort on points per word, workers in F-P would have to reduce effort on number of words. Directionally, the estimated differences are consistent with these predictions, but the differences are statistically significant only in the case of the measure of points per word.

RESULT 2: With flat-rate payments, participants generate higher points per word in the prosocial treatment compared to the non-prosocial treatment. We also find that participants solve a greater number of words in the non-prosocial treatment compared to the prosocial treatment.

Table 4: Testing differences in Phase 2 outcomes across treatments

	(1)	(2)
VARIABLES	Total words	Points per word

Treatment: F-NP	(omitted)	(omitted)
Treatment: F-P	-2.24	0.34**
	(2.35)	(0.17)
Treatment: P-NP	22.32***	-0.72***
	(4.00)	(0.20)
Treatment: P-P	4.71	-0.00
	(3.06)	(0.17)
Observations	244	244
R-squared	0.21	0.16
P-P vs. F-P p-val.	0.01	0.02
P-P vs. P-NP p-val.	0.00	0.00
DinD est. (PP-FP)-(PNP-FNP)	-15.37	0.39
DinD p-val.	0.00	0.11

All columns include an indicator variable for whether the data was collected online. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Finally, Theoretical Prediction 3a focuses on the impact of piece-rate pay in a prosocial setting. There, as in the non-prosocial setting, we predicted that piece-rate pay would positively impact number of words but negatively impacts points per word. Here, the relevant test is the difference between the P-P and F-P coefficients, the p-value of which is reported in the bottom portion of the table. We do find evidence that the number of words (Column 1) is higher in P-P than F-P, with P-P participants spelling 6.95 additional words on average (p-value = 0.01). While this is consistent with our theoretical prediction, it is worth noting that this response is substantially more muted than in the non-prosocial treatments. With non-prosocial treatments, as we noted above, performance pay led to an additional 22.32 words. We observe a similar phenomenon with respect to points per word: P-P participants spell words worth 0.34 fewer points than F-P participants. Although this difference is significant (p-value=0.02), the magnitude of the effect of performance pay is much smaller than the parallel comparison between P-NP and F-NP, where points per word were 0.72 points lower with performance pay. Thus, while there is evidence in favor of the theoretical prediction, the more notable finding is that the effects of piece-rate pay are much smaller in the prosocial treatments than in the non-prosocial treatments.

RESULT 3: In the prosocial treatments, piece-rate decreases points per word compared to fixed pay. We also find evidence that piece-rate increases the number of words solved compared to fixed pay. The effect magnitudes are smaller than those found in the non-prosocial treatments.

To put our main results differently: our experiment tests whether the impacts of performance pay are *different* in a setting where one dimension of effort is tied to a charitable donation versus another setting where it is not. Our results could therefore be presented as a difference-in-differences estimate, evaluating the difference between "PP – FP" and "PNP – FNP". We use our estimates in the main portion of Table 4 to report this difference-in-differences estimate in the bottom of Table 4. We have already seen that piece-rate has a smaller positive effect in prosocial treatments; we now see that this difference in effects is statistically significant (Column 1). Likewise, we have seen that the *negative* effect of piece rate on points per word is mitigated in prosocial treatments. That difference is captured by the positive difference-in-differences (p-val.=0.11) in Column 2.[19]

To assess predictions related to how ability and prosociality impact performance in the two dimensions, Table 5 reports heterogeneity in response to performance pay by worker ability and prosocial behavior. We draw on the ability measure we construct from participants' Phase 1 performances and on the share they donated to charity in the dictator game played toward the end of the experiment.[20] We modify the specification reported in Table 4, interacting treatment indicators with our constructed ability measure and the dictator donation share (DDS).

[19] To evaluate whether our results are affected by treatment differences in the proportion of participants who are non-Native English speakers (as noted in Appendix Table A1), we re-run the regressions reported in Table 4 after dropping non-Native English speakers. Results are very similar to those reported in Table 4.
[20] In Appendix Table A6 we present heterogeneity analysis by Ability and Prosociality, using PSM score as a measure of prosociality instead of Dictator Donation Share. We also note that we have tested for treatment effect heterogeneity by risk aversion and find no clear relationship.

Table 5: Heterogeneity by ability and prosocial motivation on two dimensions of productivity (Phase 2)

VARIABLES	(1) Total words	(2) Points per word
Treatment: F-NP	(omitted)	(omitted)
Treatment: F-P	-2.90	0.33*
	(1.80)	(0.17)
Treatment: P-NP	21.70***	-0.71***
	(3.45)	(0.20)
Treatment: P-P	5.42**	-0.00
	(2.24)	(0.17)
Dictator Donation Share (DDS)	-0.93	0.03
(mean 0, std. dev. 1)	(1.46)	(0.09)
Treatment: F-P X DDS	1.77	0.15
	(1.76)	(0.12)
Treatment: P-NP X DDS	0.35	0.04
	(3.34)	(0.19)
Treatment: P-P X DDS	0.40	-0.03
	(2.88)	(0.15)
Phase 1 Ability Measure	9.23***	0.14
(mean 0, std. dev. 1)	(1.40)	(0.09)
Treatment: F-P X Ability	-0.88	-0.16
	(1.94)	(0.14)
Treatment: P-NP X Ability	4.68	-0.36
	(3.47)	(0.22)
Treatment: P-P X Ability	3.96*	-0.21*
	(2.32)	(0.13)
Observations	244	244
R-squared	0.50	0.18

All columns include an indicator variable for whether the data was collected online. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Column 1 shows that higher ability workers spell more words: a one standard deviation increase in ability leads to 9.23 additional words in the F-NP treatment. The marginally significant interaction term on P-P X Ability also suggests the differential impact of ability in P-P. That performance

increases in ability in the piece rate treatments is consistent with Predictions 1b and 3b and echoes Lazear (1999) who notes that "a piece rate allows the more able to work harder and receive more from the job". That performance also increases in ability in the flat wage treatments can perhaps be attributed to the fact that our measure of ability also captures some intrinsic interest/joy in performing the task. With respect to prosociality, we also see in column 1 that it does not play a significant role in terms of number of words produced. Note that our model does not predict an influence of prosociality on the number of words produced.

Turning attention to predictions concerning performance on the second dimension (points per word) for prosocial treatments, column 2 shows no effect of ability in F-P (Prediction 2b), and contrary to Prediction 3b, a negative impact of ability in performance in P-P, with the interaction terms being significant only at the 10% level.[21]

To summarize, we have documented that – at least with respect to incentive effects – performance pay has a substantially smaller positive effect on the privately incentivized dimension of effort (number of words) when workers are in our experimental prosocial treatment compared to the non-prosocial treatment. We also document that the large positive effect of performance pay on the privately incentivized dimension in the non-prosocial treatments comes at the expense of the non-incentivized dimension of effort (points per word). While a similar pattern is found in the prosocial treatments, the negative effect on points per word is much smaller. Thus, performance pay is less successful in increasing effort on the privately incentivized dimension when workers can also direct effort towards generating money for charity, but also does less harm to other dimensions of effort.

4.3 Assessing sorting in Phase 3

We now turn to assessing the types of participants who opt for performance pay when given the choice and how these sorting decisions differ depending on whether the individuals are in prosocial or non-prosocial treatments. Recall that Phase 3 consists of an additional eight minutes of the real effort task, but unlike Phase 2 we give participants the option to either: continue to work under the treatment-assigned payment scheme, or switch to a common outside option ($9 flat-rate payment with no funds raised for charity).

[21] We also do not see differential effects by prosociality on points per word in the F-P and P-P treatments, as predicted by our model.

Table 6 reports results from regressions where we test for differences across treatments in the characteristics of workers who prefer the treatment-assigned payment scheme over the common outside option. The simplest regression (Column 1) regresses a dummy variable indicating that the participant chose the treatment-assigned payment scheme on a set of treatment indicator variables with F-NP (where only one participant opted for the clearly inferior treatment-assigned scheme) again serving as the omitted category and we also include our measures of ability and prosociality as controls. In the non-prosocial treatments (P-NP vs. F-NP), the regression reveals that the share of participants opting for the treatment-assigned payment scheme is significantly higher when the treatment-assigned scheme is a piece-rate. The opposite is true in prosocial treatments (P-P vs. F-P) where a relatively smaller share of participants opt for the treatment-assigned scheme when it pays a piece-rate, although this difference is insignificant. In other words, the results suggest that performance pay increases the attractiveness of jobs in non-prosocial settings but does not affect the attractiveness of jobs in prosocial settings, relative to an outside option that pays a high fixed wage with no opportunity to generate a donation to charity.

Because we are interested in the types of workers who sort into performance pay, our focus in the remainder of the table is in understanding how ability and prosocial motivation (as measured by the Phase 1 ability score and the share donated to charity in a dictator game, respectively, both standard-normalized) dictate the choice of the treatment-assigned option, especially in the P-P and P-NP treatments. In Column 2, both measures are fully interacted with treatment indicators to test how ability and prosociality differentially impact the decision to opt for the treatment-assigned payment scheme across treatments.

Notably, the decision to opt for the treatment-assigned scheme is only meaningfully impacted by ability when the treatment-assigned scheme is P-NP or P-P, that is, piece-rate with or without charity. In P-NP, workers who are one standard deviation higher in ability are 20 percentage points more likely to choose to work under the piece-rate scheme, while in P-P this figure is smaller, at 10 percentage points. Aside from the multitasking element, P-NP most closely mirrors the type of setting where performance pay has typically been studied, as in Lazear (2000) for instance. Lazear (2000) theoretically predicts and empirically documents that the most productive workers will be attracted to performance pay, which provides the beneficial "sorting effect" of performance pay. Our theoretical model makes the same prediction (predicting that higher ability workers are more likely to choose the performance pay scheme if given the choice) and the results reported in the

table support that prediction. There is no evidence of sorting on ability in F-P or F-NP, but there is also no reason to expect that there should be.

Table 6: Heterogeneity in ability and prosocial motivation of workers choosing treatment-assigned vs. outside option (Phase 3)

VARIABLES	(1) Inside option	(2) Inside option
Treatment: F-NP	(omitted)	(omitted)
Treatment: F-P	0.52***	0.51***
	(0.07)	(0.06)
Treatment: P-NP	0.37***	0.36***
	(0.06)	(0.06)
Treatment: P-P	0.49***	0.49***
	(0.07)	(0.06)
Dictator Donation Share (DDS)	0.06**	0.02
(mean 0, std. dev. 1)	(0.03)	(0.01)
Treatment: F-P X DDS		0.19***
		(0.06)
Treatment: P-NP X DDS		0.01
		(0.06)
Treatment: P-P X DDS		0.05
		(0.07)
Phase 1 Ability Measure	0.09***	0.01
(mean 0, std. dev. 1)	(0.03)	(0.01)
Treatment: F-P X Ability		0.04
		(0.07)
Treatment: P-NP X Ability		0.20***
		(0.06)
Treatment: P-P X Ability		0.10*
		(0.05)
Observations	244	244
R-squared	0.23	0.28
P-P vs. F-P p-val.	0.71	
P-P vs. P-NP p-val.	0.17	
DinD est. (PP-FP)-(PNP-FNP)	-0.40	
DinD p-val.	0.00	

All columns include an indicator variable for whether the data was collected online. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Turning attention to prosociality, dictator donation share is only a significant determinant of choosing the treatment-assigned option in the F-P treatment. Workers who are one standard deviation higher in their DDS are 21 percentage points ("DDS" + "F-P X DDS") more likely to opt for the $7 flat-rate with charity than the $9 flat-rate without. While the P-P X DDS coefficient is positive, it is not significant, thus, there is little evidence of sorting on the basis of DDS in this treatment. On the other hand, there is sorting based on DDS when the choice is between a fixed pay without a prosocial element and a (lower) fixed pay with the potential to generate a donation to charity.

To summarize, we find that high ability workers self-select into the piece rate compensation scheme in both prosocial and non-prosocial treatments, thereby generating an additional positive effect of P4P on productivity due to sorting (Lazear, 2000). Also, in the environment where individuals can generate a donation to charity, we find that with fixed pay there is sorting on prosociality, whereas with P4P this is not the case.

4.4 Sorting vs. Incentive Effects

Finally, we consider the relative roles of sorting effects (changes in aggregate productivity driven by characteristics of individuals who opt into a payment scheme) and incentive effects (changes in aggregate productivity driven by changes in behavior induced by the payment scheme) of piece rate compensation in our setting. We simply plot means of Phase 2 outcomes across treatments in Figure 6a and 6b, doing so separately for all participants and for the subset of participants who in Phase 3 go on to self-select into the treatment-assigned payment option. This facilitates a comparison between incentive effects (all participants) and the combination of sorting and incentive effects (self-selected participants).[22]

In both plots, we omit the self-selected data for the Flat Rate-Nonprosocial (F-NP) treatment, as only one participant selected into the treatment-assigned payment scheme in that treatment.[23] This fact makes it less straightforward to compare outcomes in the regression framework used elsewhere

[22] It may seem natural to instead compare Phase 2 means to Phase 3 means (amongst self-selected participants), because Phase 3 is the time period during which they have selected. However, comparing means within Phase 2 only yields a cleaner comparison – as Phase 3 outcomes may differ from Phase 2 for reasons other than selection; namely, participant experience with the task, different set of letters to draw from, participant fatigue, etc.

[23] This of course makes sense: the treatment assigned scheme was a $7 flat rate, with no contribution to charity; the outside option was a $9 flat rate, also with no contribution to charity.

– as the omitted category would represent the single participant opting for treatment-assigned payment in the F-NP treatment.

As we might expect, the total number of words spelled is substantially higher under piece-rate pay amongst self-selected participants, both for prosocial and non-prosocial treatments. In both cases, we have already documented that there are positive incentive and sorting effects – with higher ability participants more likely to opt for the treatment-assigned option.

For the prosocial treatments, we can roughly assess how much of the increased output is from incentive vs. sorting effects. If average output amongst self-selected participants represents the combination of incentive and sorting effects, and average output amongst all Phase 2 participants measures incentive effects alone, the *difference* in the effects of piece rate across Phases can isolate the sorting effect. To be more precise, let X_{all} represent average outcomes in treatment X for all participants in Phase 2 and let $X_{sel.}$ Represent average outcomes in treatment X for participants who go on to self-select into that treatment's payment scheme in Phase 3. Then, $Incentive\ Effect=PP_{all}-FP_{all}$ and $[Incentive+Sorting\ Effect]=PP_{sel.}-FP_{sel.}$. Thus, $Sorting\ Effect=(PP_{sel.}-FP_{sel.})-(PP_{all}-FP_{all})$.

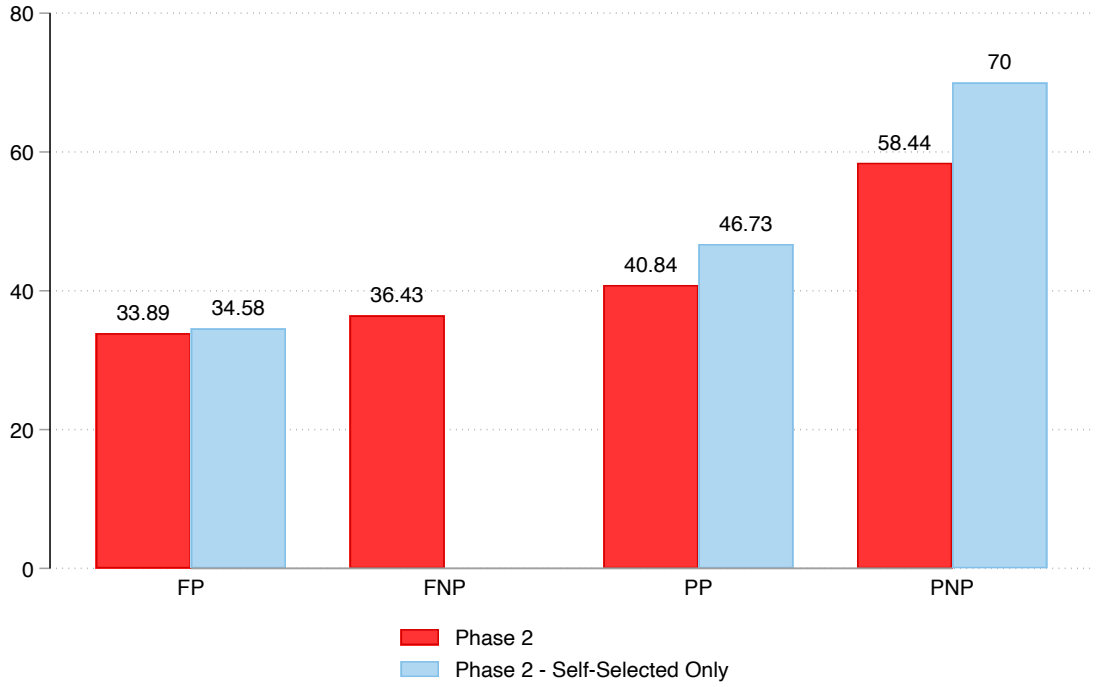
The incentive effect of piece rate on the number of words spelled in the prosocial setting is +6.95 (40.84-33.89). The combination of sorting and incentive effects is +12.15 words (46.73-34.58). Thus, we estimate that 57% of the increase in productivity with respect to number of words is from incentive effects (6.95/12.15) and 43% is from sorting effects.

Direct comparison of this sort for non-prosocial treatments is complicated by the fact that we do not have a reliable mean for $FNP_{sel.}$ Nonetheless, in that treatment, the treatment-assigned and outside option payment schemes both involve a flat rate with no contribution to charity. Thus, it is reasonable to adopt FNP_{all} as an approximation for $FNP_{sel.}$ With that, making the same comparison as above, incentive effects of piece rate in the absence of opportunity for charitable contributions are 22.01 (58.44-36.43); total effects of piece rate are 33.57 (70-36.43); so, sorting effects are 11.56 – or roughly 53% of the overall effect of piece rate. This magnitude is close to the classic finding in Lazear (2000) that roughly half of the impact of piece rate is driven by sorting.

The fact that sorting plays a larger role in total productivity in the non-prosocial treatments than in the prosocial treatments is consistent with our earlier results. Specifically, we found that while our

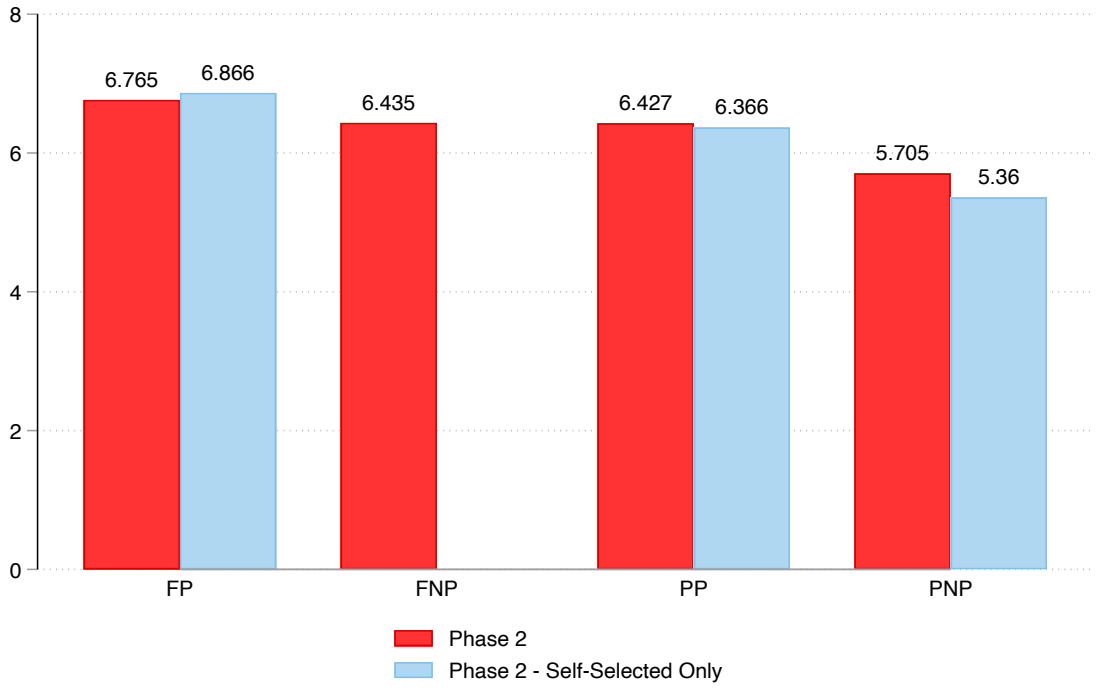
measure of ability was positively associated with productivity in both the PNP and PP treatments (Table 5), there was stronger evidence of sorting on ability in PNP than PP (Table 6).

Figure 6a: Total Words – Comparing Outcomes Under Treatment Payment Scheme in Phase 2 (conditional vs. unconditional on eventual selection into payment scheme)



While Figure 6a documents that impacts of piece rate on total words are quite large for both prosocial and non-prosocial treatments, the picture is slightly different in Figure 6b. There is a decline in points per word with piece-rate in prosocial and non-prosocial, but it is much smaller in the case of the prosocial treatment. Following the same decomposition of sorting and incentive effects, in prosocial treatments, the incentive effect of piece rate on points per word is -0.34. The combination of sorting and incentive effects is -0.5. Thus, we estimate that 68% of the negative effect in productivity with respect to points per word comes from incentives and 32% comes from sorting. In nonprosocial treatments, the incentive effect of piece rate on points per word is -0.73. The combination of sorting and incentive effects is -1.08. Thus, like in the prosocial treatments, we estimate that 68% of the negative effect in productivity with respect to points per word comes from incentives and 32% comes from sorting.

Figure 6b: Points per Word – Comparing Outcomes Under Treatment Assigned Payment Scheme in Phase 2 (conditional vs. unconditional on eventual selection into payment scheme)



5. Discussion

Our experiment is motivated by the question of whether performance pay operates differently in prosocial settings, both with regards to incentive effects and sorting effects. Many researchers have noted that performance pay may operate quite differently in the public and non-profit sectors due to the prevalence of multitasking amongst workers. Borrowing examples from Prentice et al. (2007), teachers in schools might aim to improve students' scores on standardized tests *and* improve cultural awareness; hospitals may aim to minimize re-admission of patients *and* improve the long-term health of patients. Importantly in each of these two examples, there is a trade-off between the two objectives that workers are expected to pursue, where one objective is harder to measure and therefore harder to incentivize. This raises questions around how incentives on an easily measurable dimension impacts effort on other dimensions. Our experiment mimics these elements by presenting participants with a task that involves a trade-off in terms of effort that can be directed towards different dimensions, captured by the number of words spelled and the points per word. Indeed, a feature of the task is that it is easier/faster to form, for instance, two-letter words

(out of the seven available letters) with more common letters than to form more difficult words that involve rarer letters and/or words that could be longer and would bear more points. The parallel here would be a short and superficial consultation with a doctor versus a longer and deeper consultation. Obviously, for the sake of providing evidence on these questions, we had to choose a task where the "quality" dimension *is* measurable.

In our prosocial treatments, to contribute to charity in a significant manner a participant needs to produce many words of sufficiently high complexity. This is related to the sense of mission of, for instance, a doctor who cares about the wellbeing of patients and, therefore, tries to see as many patients as possible, compatible with having enough time to provide a high-quality treatment to each of them. Linking effort to charitable donations is a well-established methodology to elicit prosocial motivation (e.g., Tonin and Vlassopoulos, 2010, 2015; Imas, 2014; Charness et al., 2016; DellaVigna and Pope, 2018; Cassar, 2019) and has the advantage of keeping the task itself constant between the different treatments. It is, however, an indirect way to do so and it may be less salient than having own effort directly contributing to a socially valuable cause.

Furthermore, pay-for-performance may have different implications in terms of the type of workers it attracts. We address this issue in Phase 3 of our experiment by offering participants a choice between the payment scheme they experienced in Phase 2 or an outside option (flat wage, non-prosocial), which we chose to have a common anchor to gauge the sorting patterns of the various treatments. Besides, the features of the chosen outside option arguably characterize most of the employment opportunities in the private sector. In practice, of course, workers choose among jobs that are prosocial or non-prosocial and offer fixed or variable payment schemes. However, we opted for a simpler pairwise choice to keep the experimental design manageable.

How does our experiment inform the debate on whether or not to introduce performance pay in sectors with a strong prosocial mission, like education or health care? On one side, our results suggest that it may be misleading to extrapolate the experience about high-powered incentives coming from standard for-profit sectors of the economy to sectors where prosocial motivation plays a more important role. For instance, those pushing for the adoption of P4P for teachers on the basis of its beneficial effects on productivity in the manufacturing sector may miss their target. On the other hand, our results also counter the argument that introducing pay-for-performance may have ruinous effects in a mission environment characterized by multitasking. Returning to the example of the educational sector, our findings do not support arguments that with bonuses based on test

scores teachers will be teaching solely to the test. We find that, when mission is present, the reallocation of effort away from the unincentivized dimension into the incentivized one is subdued. Regarding selection, by introducing pay-for-performance a mission sector may attract more skilled workers but give up the selection on motivation we observe with fixed pay. These results are of course related to the fact that, as shown in Figure 4, motivation and ability are uncorrelated in our sample and this may not be the case in all contexts.

6. Conclusions

We carry out a real effort experiment to address the following two questions: (i) How does performance pay impact productivity on incentivized and non-incentivized dimensions when workers are in a job with a prosocial dimension? (ii) How does performance pay impact the composition of workers in such jobs? We first show that, in absence of a prosocial dimension, performance pay has strong positive effects on productivity on the incentivized dimension and negative effects on the non-incentivized dimension. When the job has a prosocial element, however, the effect of performance pay is much more subdued, with much smaller changes on either dimension. Further, when workers can choose to remain in the experimentally-assigned payment scheme or opt for an outside option with flat payment and no prosocial element, we find that workers sort on ability, with lower ability workers opting out of the P4P scheme and this takes place in both sectors. In the prosocial setting, sorting also takes place along the prosocial dimension when the choice is between a fixed wage without a prosocial element and a (lower) fixed wage with a prosocial element. That is, people with high prosocial motivation are willing to give up financial gains to be able to contribute to the prosocial dimension of the job.

More generally, what we have shown is that the effects of pay-for-performance are not independent of mission. Therefore, whenever prosocial motivation matters, these two types of incentives should be studied together. While the lab environment allows for clean measurement and identification, it has clear limitations when addressing issues relevant to complex workplaces like schools or hospitals. Before drawing firm policy implications, it is therefore important to build a broader knowledge base, including field studies whose design can be informed by the results presented here.

REFERENCES

Al-Ubaydli, O., Andersen, S., Gneezy, U. and List, J.A., 2015. Carrots that look like sticks: Toward an understanding of multitasking incentive schemes. *Southern Economic Journal*, *81*(3), pp.538-561.

Ashraf, N., Bandiera, O., Davenport, E. and Lee, S.S., 2020. Losing prosociality in the quest for talent? Sorting, selection, and productivity in the delivery of public services. *American Economic Review*, *110*(5), pp.1355-94.

Banuri, S. and Keefer, P., 2016. Pro-social motivation, effort and the call to public service. *European Economic Review*, *83*, pp.139-164.

Barigozzi, F., Burani, N. and Raggi, D., 2018. Productivity crowding-out in labor markets with motivated workers. *Journal of Economic Behavior & Organization*, *151*, pp.199-218.

Besley, T. and Ghatak, M., 2017. Profit with purpose? A theory of social enterprise. *American Economic Journal: Economic Policy*, 9(3), pp.19-58.

Besley, T. and Ghatak, M., 2018. Prosocial Motivation and Incentives. *Annual Review of Economics*, *10*, pp.411-38.

Bloom, N. and Van Reenen, J., 2011. Human resource management and productivity. In *Handbook of Labor Economics* (Vol. 4, pp. 1697-1767). Elsevier.

Briscese, G., Feltovich, N., and Slonim, R. L., (2021) Who benefits from corporate social responsibility? Reciprocity in the presence of social incentives and self-selection. *Games and Economic Behavior, 126,* pp.288-304.

Burgess, S. and Ratto, M., 2003. The role of incentives in the public sector: Issues and evidence. *Oxford Review of Economic Policy*, *19*(2), pp.285-300.

Carpenter, J. and Gong, E., 2016. Motivating Agents: How much does the mission matter?. *Journal of Labor Economics*, *34*(1), pp. 211-236.

Cassar, L., 2019. Job Mission as a Substitute for Monetary Incentives: Benefits and Limits. *Management Science*, 65(2), pp. 896-912.

Cassar, L. and Meier, S. 2018. Nonmonetary Incentives and the Implications of Work as a Source of Meaning. *Journal of Economic Perspectives, 32(3)*, pp. 1–25.

Chalkley, M. and Malcomson, J.M., 1998. Contracting for health services with unmonitored quality. *The Economic Journal*, *108*(449), pp. 1093-1110.

Charness, G., Cobo-Reyes, R. and Sanchez, A., 2016. The effect of charitable giving on workers' performance: Experimental evidence. *Journal of Economic Behavior & Organization*, *131*, pp.61-74.

Charness, G., Gneezy, U. and Henderson, A., 2018. Experimental methods: Measuring effort in economics experiments. *Journal of Economic Behavior & Organization*, *149*, pp.74-87.

Copeland, A. and Monnet, C., 2009. The welfare effects of incentive schemes. *The Review of Economic Studies*, *76*(1), pp.93-113.

Danz, D., Gupta, N., Lepper, M., Vesterlund, L., and Winichakul, K.P., 2021. Going virtual: A step-by-step guide to taking the in-person experimental lab online. *Technical Report*.

Dal Bó, E., Finan, F. and Rossi, M.A., 2013. Strengthening state capabilities: The role of financial incentives in the call to public service. *The Quarterly Journal of Economics*, *128*(3), pp.1169-1218.

Delfgaauw, J. and Dur, R., 2007. Signaling and screening of workers' motivation. *Journal of Economic Behavior & Organization*, *62*(4), pp.605-624.

Delfgaauw, J. and Dur, R., 2010. Managerial talent, motivation, and self-selection into public management. *Journal of Public Economics*, *94*(9-10), pp.654-660.

DellaVigna, S. and Pope, D., 2018. What motivates effort? Evidence and expert forecasts. *The Review of Economic Studies*, *85*(2), pp.1029-1069.

Deserranno, E., 2019. Financial incentives as signals: experimental evidence from the recruitment of village promoters in Uganda. *American Economic Journal: Applied Economics*, *11*(1), pp.277-317.

Dixit, A., 2002. Incentives and organizations in the public sector: An interpretative review. *Journal of Human Resources*, pp.696-727.

Dohmen, T. and Falk, A., 2011. Performance pay and multidimensional sorting: Productivity, preferences, and gender. *American Economic Review*, *101*(2), pp.556-90.

Douven, R., Remmerswaal, M., and Zoutenbier, R. 2019. Do altruistic mental health care providers have better treatment outcomes? *Journal of Human Resources*, *54*(2), pp.310-341.

Finan, F., Olken, B.A. and Pande, R., 2017. The personnel economics of the developing state. In *Handbook of Economic Field Experiments* (Vol. 2, pp. 467-514). North-Holland.

Francois, P., 2007. Making a difference. *The RAND Journal of Economics*, 38, pp.714-732.

Francois, P. and Vlassopoulos, M., 2008. Pro-social motivation and the delivery of social services. *CESifo Economic Studies*, *54*(1), pp.22-54.

Gneezy, U., Meier, S. and Rey-Biel, P., 2011. When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives*, *25*(4), pp.191-210.

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

Holt, C. and Laury, S., 2002. Risk aversion and incentive effects. *American Economic Review*, 92, pp. 1644–1655.

Hong, F., Hossain, T., List, J.A. and Tanaka, M., 2018. Testing the theory of multitasking: Evidence from a natural field experiment in Chinese factories. *International Economic Review*, *59*(2), pp.511-536.

Imas, A., 2014. Working for the "warm glow": On the benefits and limits of prosocial incentives. *Journal of Public Economics*, *114*, pp.14-18.

Imberman, S.A., 2015. How effective are financial incentives for teachers? *IZA World of Labor*.

Jones, D.B., 2015. The supply and demand of motivated labor: When should we expect to see nonprofit wage gaps? *Labour Economics*, *32*, pp.1-14.

Kajackaite, A., and Sliwka, D. 2017. Social responsibility and incentives in the lab: Why do agents exert more effort when principals donate? *Journal of Economic Behavior & Organization*, 142, pp.482-493.

Lazear, E.P., 1986. Salaries and piece rates. *Journal of Business*, pp.405-431.

Lazear, E.P., 2000. Performance pay and productivity. *American Economic Review*, *90*(5), pp.1346-1361.

Lemieux, T., MacLeod, W.B. and Parent, D., 2009. Performance pay and wage inequality. *The Quarterly Journal of Economics*, *124*(1), pp.1-49.

Neal, D. 2011. The design of performance pay in education. In: Hanushek, E. A., S. J. Machin, and L. Woessmann (eds). *Handbook of the Economics of Education*: Vol. 4. Amsterdam: Elsevier, pp. 495–550.

Perry, J. L., 1996. Measuring public service motivation: An assessment of construct reliability and validity. *Journal of Public Administration Research and Theory*, *6*(1), pp.5-22.

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

Prentice, G., Burgess, S., and Propper, C. 2007. Performance pay in the public sector: A review of the issues and evidence. Office of Manpower Economics Report.

Scott, A., Sivey, P., Ouakrim, D.A., Willenberg, L., Naccarella, L., Furler, J. and Young, D., 2011. The effect of financial incentives on the quality of health care provided by primary care physicians. Cochrane Database of Systematic Reviews.

Shearer, B., 2004. Piece rates, fixed wages and incentives: Evidence from a field experiment. *The Review of Economic Studies*, *71*(2), pp.513-534.

Tonin, M. and Vlassopoulos, M., 2010. Disentangling the sources of pro-socially motivated effort: A field experiment. *Journal of Public Economics*, *94*(11-12), pp.1086-1092.

Tonin, M. and Vlassopoulos, M., 2015. Corporate Philanthropy and Productivity: Evidence from an Online Real Effort Experiment. *Management Science*, *61*(8), pp.1795-1811.

Appendix A: Additional Tables & Figures

Table A1: Summary statistics by treatment

TREATMENT:	(1) Flat-rate Prosocial (F-P)	(2) Flat-rate Non-Prosocial (F-NP)	(3) Piece-rate Prosocial (P-P)	(4) Piece-rate Non-Prosocial (P-NP)
Female	0.59	0.63	0.57	0.50
	(0.50)	(0.49)	(0.50)	(0.51)
Age	19.61	19.96	19.66	19.46
	(1.02)	(2.60)	(1.70)	(1.45)
Year in college	2.37	2.33	2.28	2.24
	(1.05)	(0.94)	(1.06)	(1.13)
Business major	0.53	0.48	0.38	0.46
	(0.50)	(0.51)	(0.49)	(0.50)
Native English?	0.94	0.98	0.87	0.90
	(0.24)	(0.15)	(0.34)	(0.30)
Currently works?	0.63	0.59	0.68	0.68
	(0.49)	(0.50)	(0.47)	(0.47)
Current wage (conditional	7.79	7.02	7.89	11.18
on current work)	(7.38)	(7.27)	(6.22)	(21.28)
Any monetary donations?	0.69	0.70	0.64	0.72
(past year)	(0.47)	(0.47)	(0.48)	(0.45)
Any goods donations?	0.88	0.72	0.70	0.78
(past year)	(0.33)	(0.46)	(0.46)	(0.42)
Any volunteering?	0.69	0.72	0.66	0.66
(past year)	(0.47)	(0.46)	(0.48)	(0.48)
Any blood donations?	0.33	0.33	0.21	0.24
(past year)	(0.47)	(0.47)	(0.41)	(0.43)
Holt-Laury: Share of risky	0.52	0.49	0.50	0.51
choices	(0.20)	(0.20)	(0.19)	(0.17)
Dictator: Share contributed	0.48	0.48	0.45	0.41
	(0.31)	(0.33)	(0.29)	(0.32)
PSM Survey score	0.68	0.68	0.70	0.70
(0-1, higher = more PSM)	(0.11)	(0.08)	(0.09)	(0.09)
Ability	0.01	0.08	-0.11	0.09
	(0.93)	(1.13)	(1.06)	(1.00)
Observations	61	61	61	61

We have conducted pairwise comparisons across all treatments for all outcomes. Six (of 90 pairwise comparisons) are statistically significantly different: in F-NP vs. P-P, "currently works" ($p<10\%$) and "native English" ($p<5\%$); in F-NP vs. P-NP, "native English" ($p<5\%$); in F-P vs. P-P, "any goods donation" ($p<5\%$) and "native English" ($p<5\%$); in F-P vs. P-NP, "any goods donation" ($p<10\%$). We note that 6 of 90 pairwise comparisons is less than 7%, a share we might expect to randomly appear as statistically significant at the 5-10% level. Standard deviations in parentheses.

Table A2: Averages of Phase 3 outcome variables by treatment

TREATMENT:	Piece-rate Non-Prosocial (P-NP)	Flat-rate Non-Prosocial (F-NP)	Piece-rate Prosocial (P-P)	Flat-rate Prosocial (F-P)
Panel A: All participants				
Phase 3: Words spelled	64.62	39.33	48.85	37.21
	(4.22)	(2.80)	(3.39)	(1.54)
Phase 3: Points per word	4.94	6.17	5.98	6.62
	(0.16)	(0.16)	(0.13)	(0.09)
Phase 3: Share of words that	0.55	0.76	0.71	0.82
are counted (>4 pts.)	(0.03)	(0.02)	(0.02)	(0.01)
Phase 3: Earnings	7.54	7.92	7.23	7.85
	(0.32)	(0.13)	(0.33)	(0.13)
Phase 3: Total Counted	207.16	203.79	221.72	220.56
Points	(11.16)	(11.00)	(11.88)	(8.79)
Phase 3: Charitable cont.	0.00	0.00	3.22	3.61
	(0.00)	(0.00)	(0.35)	(0.28)
Observations	61	61	61	61
Panel B: Self-selected only				
Phase 3: Words spelled	79.30	*	58.30	39.27
	(8.06)		(5.41)	(2.08)
Phase 3: Points per word	4.57	*	5.98	6.78
	(0.26)		(0.19)	(0.11)
Phase 3: Share of words that	0.47	*	0.71	0.85
are counted (>4 pts.)	(0.05)		(0.03)	(0.01)
Phase 3: Earnings	7.65	*	7.35	7.85
	(0.57)		(0.45)	(0.17)
Phase 3: Total Counted	206.17	*	264.57	241.48
Points	(23.88)		(18.43)	(11.70)
Phase 3: Charitable cont.	0.00	*	5.29	4.83
	(0.00)		(0.37)	(0.23)
Observations	23	1	30	33

Standard deviations in parentheses. (* Because only one participant selected into the treatment-assigned payment scheme in treatment F-NP, we do not report the averages for that treatment in Panel B, as they would be based on a single observation.)

Table A3: Testing differences in Phase 2 outcomes across treatments (with some additional controls)

VARIABLES	(1) Total words	(2) Points per word
Treatment: F-NP	(omitted)	(omitted)
Treatment: F-P	-1.70	0.27*
	(2.45)	(0.16)
Treatment: P-NP	23.22***	-0.76***
	(4.13)	(0.20)
Treatment: P-P	5.45	0.01
	(3.31)	(0.18)
Observations	244	244
R-squared	0.26	0.21
P-P vs. F-P p-val.	0.02	0.09
P-P vs. P-NP p-val.	0.00	0.00
DinD	-16.07	0.50
DinD p-val.	0.00	0.04

All specifications include the following additional controls: main Phase 1 ability measure, gender, years in college, business major indicator, native English speaker indicator, current work status. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A4 considers three alternative measures for the complexity dimension: "Counted points per word" divides total points accumulated (counting words worth less than 5 points as 0 points, as would be the experience of the participant) by total words spelled. This, then, is closer to the measure the participants would experience. Two further measures are aimed at assessing the margin along which participants changed their behavior to increase points and therefore money for charity, in the prosocial treatments. Participants might have focused on spelling more complex words (an intensive margin response), or they might have focused on making sure more of their words were at least five points (an extensive margin response). To get at this, we construct "Counted points per counted word", which divides total points accumulated from words worth at least five points by the number of words spelled that were worth at least five points, and "Share of words that are counted", which is the number of words worth at least five points divide by the total words spelled. The first of these would reveal the intensive margin response, the second would reveal the extensive margin response. In practice, we find that both follow similar patterns, so participants respond both by increasing the average complexity of words (even conditional on the word already achieving five points) and by increasing the fraction of words that count.

Table A4: Testing differences in Phase 2 points-related outcomes across treatments (using alternative measures of points per word)

VARIABLES	(1) Counted points per word	(2) Counted points per counted word	(3) Share of words that are counted
Treatment: F-P	0.43**	0.27*	0.04*
	(0.20)	(0.16)	(0.02)
Treatment: P-NP	-1.12***	-0.29*	-0.13***
	(0.26)	(0.16)	(0.03)
Treatment: P-P	-0.05	0.04	-0.00
	(0.21)	(0.15)	(0.02)
Observations	244	244	244
R-squared	0.19	0.07	0.21
P-P vs. F-P p-val.	0.01	0.03	0.04
P-P vs. P-NP p-val.	0.00	0.00	0.00
DinD est.	0.63	0.07	0.09
DinD p-val.	0.05	0.73	0.02

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A5: Nonparametric tests of differences in primary Phase 2 outcomes

	(1) Total words (Phase 2)	(2) Points per Word (Phase 2)
Treatment Comparison	*Diff. in means (in X vs. Y, reporting X minus Y) [p-value from Wilcoxon test]*	
F-P vs. F-NP	-2.34	0.34
	[0.44]	[0.11]
P-NP vs. F-NP	22.21	-0.71
	[0.001]***	[0.0001]***
P-P vs. F-NP	-4.60	-0.01
	[0.31]	[0.81]
P-NP vs. F-P	-24.56	-1.06
	[0.0001]***	[0.000]***
P-P vs. F-P	6.95*	-0.33
	[0.05]	[0.065]**
P-NP vs. P-P	17.60	0.72
	[0.0001]***	[0.0003]***

Each cell reports the difference in means between an outcome (noted in the column header) between two treatments. We conduct a Wilcoxon Rank-Sum Test for each treatment comparison and report the p-value in brackets beneath the differences in means. *** p<0.01, ** p<0.05, * p<0.1

Table A6: Phase 2 and Sorting Outcomes, Heterogeneity by Ability and Prosociality, taking Public Service Motivation score as measure of prosociality instead of Dictator Donation Share

VARIABLES	(1) Total words	(2) Points per word	(3) Sorting: Chooses Treatment-assigned payment
Treatment: F-NP	(omitted)	(omitted)	(omitted)
Treatment: F-P	-3.14*	0.33**	0.53***
	(1.78)	(0.16)	(0.07)
Treatment: P-NP	21.68***	-0.74***	0.36***
	(3.41)	(0.19)	(0.06)
Treatment: P-P	4.96**	-0.00	0.48***
	(2.21)	(0.15)	(0.07)
Pub. Srvc. Motivation (PSM)	2.79*	0.14	-0.00
(mean 0, std. dev. 1)	(1.56)	(0.13)	(0.01)
Treatment: F-P X PSM	-3.52*	-0.05	0.12***
	(1.81)	(0.14)	(0.05)
Treatment: P-NP X PSM	-6.57	0.01	-0.05
	(4.23)	(0.21)	(0.06)
Treatment: P-P X PSM	-1.36	-0.30*	0.03
	(2.46)	(0.16)	(0.07)
Phase 1 Ability Measure	9.56***	0.16	0.01
(mean 0, std. dev. 1)	(1.50)	(0.10)	(0.01)
Treatment: F-P X Ability	-1.52	-0.21	0.03
	(2.01)	(0.15)	(0.07)
Treatment: P-NP X Ability	4.65	-0.40*	0.19***
	(3.35)	(0.21)	(0.06)
Treatment: P-P X Ability	3.32	-0.21*	0.11**
	(2.11)	(0.13)	(0.05)
Observations	244	244	244
R-squared	0.51	0.19	0.27

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Appendix B: Additional details from the theory model

Prediction 1a.

Totally differentiating the F.O.C.s with respect to p yields: $\frac{\partial c(e_1,e_2)}{\partial e_1 \partial e_1}\frac{de_1}{dp}+\frac{\partial c(e_1,e_2)}{\partial e_1 \partial e_2}\frac{de_2}{dp}-\gamma=0$ and $\frac{\partial c(e_1,e_2)}{\partial e_2 \partial e_1}\frac{de_1}{dp}+\frac{\partial c(e_1,e_2)}{\partial e_2 \partial e_2}\frac{de_2}{dp}=0$. Solving the second equation for $\frac{de_2}{dp}$ and substituting into the first gives after rearrangement $\frac{de_1}{dp}=\frac{\gamma\frac{\partial c(e_1,e_2)}{\partial e_2 \partial e_2}}{\frac{\partial c(e_1,e_2)}{\partial e_1 \partial e_1}\frac{\partial c(e_1,e_2)}{\partial e_2 \partial e_2}-\left(\frac{\partial c(e_1,e_2)}{\partial e_1 \partial e_2}\right)^2}>0$, because of the convexity of the cost function. Also, $\frac{de_2}{dp}=-\frac{\frac{\partial c(e_1,e_2)}{\partial e_2 \partial e_1}\frac{de_1}{dp}}{\frac{\partial c(e_1,e_2)}{\partial e_2 \partial e_2}}<0$. Thus, relative to the case of a flat wage where the agent chooses $e^*=(e_1^*,e_2^*)$, given by $\frac{\partial c(e_1,e_2)}{\partial e_1}=\frac{\partial c(e_1,e_2)}{\partial e_2}=0$, introducing a piece rate p increases e_1 and decreases e_2.

Prediction 1b.

Implicit differentiation of $\frac{\partial c(e_1,e_2)}{\partial e_1}=\gamma p$ yields $\frac{de_1}{d\gamma}=\frac{p}{\frac{\partial c(e_1,e_2)}{\partial e_1 \partial e_1}}>0$

Prediction 2a.

Totally differentiating the F.O.C.s with respect to p yields: $\frac{\partial c(e_1,e_2)}{\partial e_1 \partial e_1}\frac{\partial e_1}{\partial \theta}+\frac{\partial c(e_1,e_2)}{\partial e_1 \partial e_2}\frac{\partial e_2}{\partial \theta}=0$ and $\frac{\partial c(e_1,e_2)}{\partial e_1 \partial e_2}\frac{\partial e_1}{\partial \theta}+\frac{\partial c(e_1,e_2)}{\partial e_2 \partial e_2}\frac{\partial e_2}{\partial \theta}-\gamma=0$. Solving the second equation for $\frac{\partial e_2}{\partial \theta}$ and substituting into the first gives after rearrangement $\frac{\partial e_2}{\partial \theta}=\frac{\gamma\frac{\partial c(e_1,e_2)}{\partial e_1 \partial e_1}}{\frac{\partial c(e_1,e_2)}{\partial e_1 \partial e_1}\frac{\partial c(e_1,e_2)}{\partial e_2 \partial e_2}-\left(\frac{\partial c(e_1,e_2)}{\partial e_1 \partial e_2}\right)^2}>0$ because of the convexity of the cost function. Similarly, it can be shown that $\frac{\partial e_1}{\partial \theta}=-\frac{\gamma\frac{\partial c(e_1,e_2)}{\partial e_1 \partial e_2}}{\frac{\partial c(e_1,e_2)}{\partial e_1 \partial e_1}\frac{\partial c(e_1,e_2)}{\partial e_2 \partial e_2}-\left(\frac{\partial c(e_1,e_2)}{\partial e_1 \partial e_2}\right)^2}<0$. Thus, relative to the case of a flat wage where the agent chooses $e^*=(e_1^*,e_2^*)$, given by $\frac{\partial c(e_1,e_2)}{\partial e_1}=\frac{\partial c(e_1,e_2)}{\partial e_2}=0$, when the agent derives an intrinsic benefit θ from producing quality e_1 is lower and e_2 is higher.

Prediction 2b.

Implicit differentiation of $\frac{\partial c(e_1,e_2)}{\partial e_2}=\gamma\theta.$ yields $\frac{de_2}{d\gamma}=\frac{\theta}{\frac{\partial c(e_1,e_2)}{\partial e_2 \partial e_2}}>0$.

Prediction 3a, 3b.

Similar to 1a, and 1b respectively, so omitted.

Prediction 4: Follows from the fact that $\frac{\partial e_2}{\partial \theta}>0$ and $\frac{\partial e_1}{\partial \theta}<0$.

Phase 3

Flat-rate Motivated (F-P)

The agent chooses F-**P** when the utility it furnishes is higher than or equal to the utility from the outside option:

$w + \theta e_2^{FM} - \frac{1}{\gamma} c(e_1^{FP}, e_2^{FP}) > w^O - \frac{1}{\gamma} c\left(e_1^O, e_2^O\right)$, *where* $w < w^O$.

$\theta > \frac{w^O - w + \frac{1}{\gamma}[c(e_1^{FP}, e_2^{FP}) - c(e_1^O, e_2^O)]}{e_2^{FP}}$.

Piece-rate Non-Motivated (P-NP)

The agent chooses the piece rate scheme if:

$pe_1^{PNP} - \frac{1}{\gamma} c(e_1^{PNP}, e_2^{PNP}) > w^O - \frac{1}{\gamma} c\left(e_1^O, e_2^O\right)$, *with* $c(e_1^{PNP}, e_2^{PNP}) > c(e_1^O, e_2^O)$

$$\gamma > \frac{c(e_1^{PNP}, e_2^{PNP}) - c(e_1^O, e_2^O)}{pe_1^{PNP} - w^O}$$

Piece-rate Motivated (P-P)

In the mission case, the agent chooses the piece rate scheme if:

$pe_1^{PP} + \theta e_2^{PP} - \frac{1}{\gamma} c(e_1^{PP}, e_2^{PP}) > w^O - \frac{1}{\gamma} c\left(e_1^O, e_2^O\right)$, *with* $c(e_1^{PP}, e_2^{PP}) > e(e_1^O, e_2^O)$

$$\theta > \frac{w^O - pe_1^{PP} + \frac{1}{\gamma}[c(e_1^{PP}, e_2^{PP}) - c\left(e_1^O, e_2^O\right)]}{e_2^{PP}} = \hat{\theta}$$

or

$$\gamma > \frac{c(e_1^{PP}, e_2^{PP}) - c(e_1^O, e_2^O)}{pe_1^{PP} - \theta e_2^{PP} - w^O} = \hat{\gamma}$$

The above inequalities suggest that the proposed scheme is more likely to be chosen by high θ and/or γ agents. Note also that $\frac{d\hat{\theta}}{d\gamma} < 0$ and $\frac{d\hat{\gamma}}{d\theta} < 0$, suggesting that ability and motivation are substitutes for sorting, that is, a higher ability individual requires less motivation to sort and similarly a higher motivated individual requires less ability to find sorting in optimal.

Appendix C: Experimental Instructions & Questionnaires

This appendix reproduces the instructions that were provided to participants during our experimental sessions. Sections of the instructions that are specific to the flat-rate treatment are preceded by "F-NP/{P}". Sections of the instructions that are specific to the piece-rate treatment are preceded by "P-NP/{P}". Within those sections, text within curly brackets ("{…}") are specific to mission-oriented treatments. All other sections of the instructions are common across treatments. Instructions prior to "Phase 1" were presented on participants' screens, but also read aloud by the experimenter. All instructions from "Phase 1" onward were presented on participants' screens as they preceded through the experiment at their own pace and were not read aloud.

Introduction

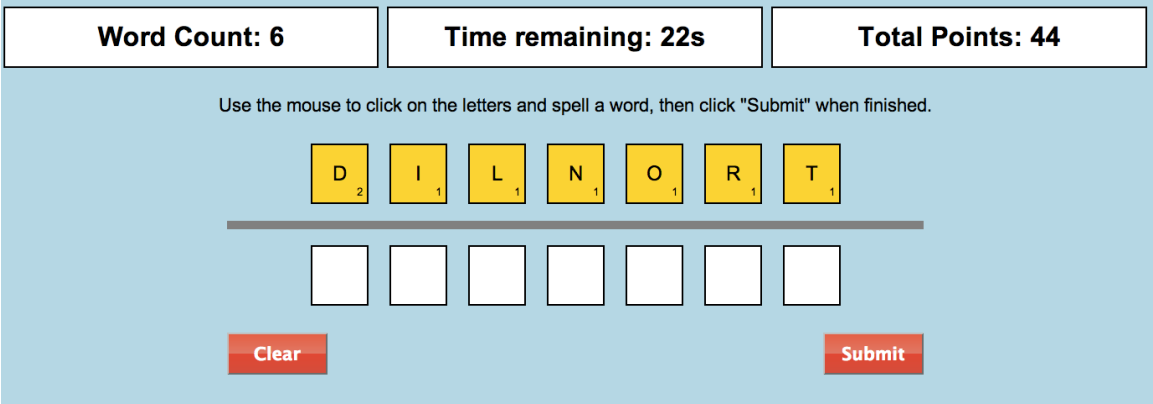
This experiment is a study of decision-making. Your earnings will depend on the actions that you take during the experiment. At the end of the experiment, you will be paid $5 for showing up plus whatever you earn during the course of the experiment. Throughout the experiment, your earnings will be reported in "Experimental Currency Units" or "ECUs". At the end of the experiment, we will convert however many ECUs you have earned into dollars at a rate of X ECU's = $1. Payments will be made privately and in cash. All decisions are made anonymously. Please do not talk to other participants during the experiment. If at any point you have a question, raise your hand and an experimenter will come to you to provide an answer.

The experiment will consist of three task phases and a questionnaire. In the three task phases, you will perform a *word formation task*. At the end of the experiment, we will pay you based on your performance in the first phase and your performance in *either* the second or third phase (plus the $5 show-up fee). We will randomly select whether you (and other participants in the room) are paid for the second or third phase, both of which are equally likely to be selected. We will discuss the specific procedures of each of these phases and how they may impact your earnings as they occur. First, we will describe the word formation task and software interface in detail.

*** screen break ***

The word formation task

During the experiment, you will be asked to complete a number of "word formation tasks." In the word formation task, you will be presented with a set of seven letters as in the figure below.



Your task is to form a word using at least two of the letters available to you. The word must be spelled as it is in the dictionary. For example, if you were provided the set of letters "D I L N O R T", you might spell the word "ON" or the word "LORD". To spell a word, click each letter in order and each letter's tile will move to the lower row. To clear the letters you have entered, click "Clear." Click "Submit" when you are done. If the word you have submitted is not an acceptable word or is incorrectly spelled, your chosen letters will be cleared and you will be asked to try again.

Each letter has a certain number of points associated with it, which is noted in the lower right hand corner of each letter square. When you submit a word, these points will be added up to determine your score for that word. If the total number of points associated with a word is less than five, your score for that word is zero. If the total number of points associated with a word is five or greater, your score for that word is given by the number of points. For example, the word "ON" is made up of two one-letter tiles and as such is worth zero points. The word "LORD" is a total of five points (1+1+1+2), and as such is worth five points.

So that you can practice working with the software, form a word with the seven letters on the next screen, then click "Submit".

*** screen break ***

Phase 1

We will now begin the first phase. This phase consists of five "sections", each of which has slightly different instructions. Follow the instructions on each screen.

*** screen break ***

Phase 1, Section 1

In this first section, you will be asked to complete 5 word formation tasks. Spell a word for each set of letters. You will have 2 minutes to do so. Remember, after spelling a word click "Submit" to move on to the next set of letters.

((task proceeds))

*** screen break ***

Phase 1, Section 2

In this section, the task is again to complete 5 word formation tasks. This time (and in all of the remaining sections of this phase), you have the opportunity to earn ECUs.

In this section, you will start with 20 ECUs. Once you click "Start" the first task will appear. You then have 60 seconds to complete all 5 tasks. You lose 1/3 ECU for each second that it takes you to complete and submit all five tasks. For example, if it takes you 6 seconds to complete *each* of the five tasks, you will have used a *total* of 30 seconds. As a result, your earnings would be 20 ECUs – (1/3)30, or 10 ECUs.

((task proceeds))

*** screen break ***

Phase 1, Section 3

This section is slightly different. You will again have the opportunity to earn money by completing word formation tasks. This time, you will have 3 minutes to complete as many word formation tasks as you can. You will earn 1 ECU for each word formation task you complete.

((task proceeds))

*** screen break ***

Phase 1, Section 4

In this section, you will complete five word formation tasks. You have 25 seconds to complete each one. In this section, the amount of money you earn will depend on the score you accumulate rather than the number of tasks you complete. For each point, you will earn 0.2 ECUs. Remember, words with point totals less than five receive a score of zero points. There is no penalty for taking the full 25 seconds for each task, and no advantage from not doing so, so take the time to spell the best word you can.

((task proceeds))

*** screen break ***

Phase 1, Section 5

In this section, you will have 3 minutes to complete word formation tasks. There is no limit to the number of words you can form during the 3 minute period. This time, your payment again depends on the number of points you accumulate. Every time you spell a word worth five or more points, you will accumulate points from that word. You will be paid 0.2 ECUs for the total number of points you have accumulated at the end of the three minutes. For example, if you spell 9 words that

are worth 4 points and 10 words that are worth 8 points, you will be paid for the points accumulated on the 10 words worth at least 5 points. That is, you would receive 8 pts * 10 words * 0.20 ECUs, or 16 ECUs.

((task proceeds))

*** screen break ***

Phase 2

We will now begin the second phase of the experiment. Unlike the previous phase, there is only one section. We are hoping to learn how many points people can accumulate in a fixed period of time.

The rules are as follows:

In this phase, you will have 8 minutes to complete word formation tasks. There is no limit to the number of words you can form during the 8 minute period. Again, you can earn money. {In this phase, your efforts can also benefit a charity of your choosing. You will be allowed to choose a charity in a moment.} You {and a charity} will be paid for the outcome of this phase only if this phase is randomly chosen instead of the third phase. You will not know which phase is being randomly selected until the end of the experiment, so you should proceed as though you may receive payment. There is a 50% chance this Phase will be selected, and a 50% chance Phase 3 will be selected.

If this phase is selected for payment:

F-NP/{P}: [Regardless of how many word formation tasks you complete in the next eight minutes, you will receive 70 ECUs. {Additionally, when you score 5 or more points while spelling a word, you will generate some money for charity. For every point that you score on that word, you will generate 0.2 ECUs for charity. For example, a 5 point word will generate the monetary equivalent of 1 ECU for charity. A 6 point word will generate 1.2 ECUs for charity. A 4 point word will generate 0 ECUs for charity (because a word must be worth at least 5 points to generate money for charity). }

For example, if you complete 40 word formation tasks {and accumulate 150 points on words worth at least 5 points}, you will receive 70 ECUs {and your charity will receive the equivalent of 30 ECUs}. If you complete 100 word formation tasks {and accumulate 250 points on words worth at least 5 points}, you will receive 70 ECUs {and your charity will receive 50 ECUs}.]

P-NP/{P}: For each word formation task that you complete, you will receive 1 ECU. {Additionally, when you score 5 or more points while spelling a word, you will generate some money for charity. For every point that you score on that word, you will generate 0.2 ECUs for charity. For example, a 5 point word will generate the monetary equivalent of 1 ECU for charity. A 6 point word will generate 1.2 ECUs for charity. A 4 point word will generate 0 ECUs for charity (because a word must be worth at least 5 points to generate money for charity). }

For example, if you complete 40 word formation tasks {and accumulate 150 points on words worth at least 5 points}, you will receive 40 ECUs {and your charity will receive the equivalent of 30

ECUs }. If you complete 100 word formation tasks {and accumulate 250 points}, you will receive 100 ECUs {and your charity will receive 50 ECUs}.]

{We will donate money generated for charities after the experiment has ended. So that you can verify that this has happened, we will post receipts from the charities on the experimenter's website. Before we begin this Phase, please indicate which of the following charities you would like to benefit.}

((Participants presented with list of charities with radio buttons.))

We will now ask you a few questions to make sure the instructions are clear. After you have answered the questions, raise your hand and the experimenter will come check your answers.

((Comprehension Check (below) passed out on paper))

1. This phase will end after:
    a. 50 words are spelled
    b. 8 minutes
    c. 20 minutes
    d. The end is randomly determined.
2. Suppose: At the end of this phase you have spelled a total of 25 words. For the sake of this example, suppose that 10 of the words were worth 2 points apiece and the remaining 15 words were all worth 10 points apiece. If this phase is randomly selected for payment, your earnings from this phase are _____ {and charity receives _____}. (Fill in the blanks.)
3. Suppose: At the end of this phase you have spelled a total of 50 words. For the sake of this example, suppose that 49 of the words were worth 4 points apiece and the remaining word was worth 15 points. If this phase is randomly selected for payment, your earnings from this phase are _____ {and charity receives _____}. (Fill in the blanks.)
4. Which of the following is accurate:
    a. You will be paid for Phase 1 OR Phase 2 OR Phase 3.
    b. You will be paid for Phase 1 AND, either Phase 2 OR Phase 3.
    c. You will be paid for Phase 1 AND Phase 2 AND Phase 3.
    d. You will be paid for Phase 1 OR Phase 2, AND Phase 3.

Phase 3

We will now begin the third phase of the experiment. This is the final phase that will require you to complete word formation tasks. After this phase, you will complete a questionnaire and then the experiment will be over.

This phase is similar to Phase 2. Once again you will have 8 minutes to complete as many word formation tasks as you want. The main difference is that this time you can choose how you are paid. You have two options:

(Previous Option) [Repeat treatment specific payment information.] {The money generated for charity will go to the charity that you chose in the previous phase.}
(New Option) Regardless of how many word formation tasks you complete in the next eight minutes, you will receive 90 ECUs.

Recall that you will receive payment {and benefit charity} as according to your choice only if this phase is the one that is randomly selected for payment at the end of the experiment.

Please respond to two quiz questions to ensure that these instructions are clear. After you have answered the questions, raise your hand and the experimenter will come check your answers.

((Comprehension Check (below) passed out on paper))

1. Suppose you have selected the "Previous Option". At the end of this phase you have spelled a total of 25 words. {For the sake of this example, suppose that 10 of the words were worth 2 points apiece and the remaining 15 words were all worth 10 points apiece.} If this phase is randomly selected for payment, your earnings from this phase are _____ {and charity receives _____}. (Fill in the blanks.)
2. Suppose you have selected "New Option". Again suppose that at the end of this phase you have spelled a total of 25 words. {For the sake of this example, suppose that 10 of the words were worth 2 points apiece and the remaining 15 words were all worth 10 points apiece.} If this phase is randomly selected for payment, your earnings from this phase are _____ {and charity receives _____}. (Fill in the blanks.)

Now, please select the option you prefer. After you have made your selection, this phase will begin.

Additional decisions and questionnaire

You have now completed the part of the experiment that involves word formation tasks. We will now ask you to make a few more decisions and answer a brief questionnaire. After the questionnaire is completed, your payment for the experiment will be determined.

Decision task 1: Lottery vs. sure payment decisions

This part concerns the choice between a **lottery** and **a sure payment**.

On the following screen, 14 situations will be displayed. The lottery is the same in each situation, but the sure payment varies.

In the lottery you get 30 ECUs with 50 percent probability and 0 points with 50 percent probability (determined by a random draw of the computer).

The following screen will present the 14 situations. Please decide in each situation whether you opt for the lottery or the sure payment.

Once you have made your choice in each situation, the computer will randomly select one situation. If you selected the sure payment in the randomly selected situation, you will receive the sure payment associated with that situation. If you selected the lottery in the selected situation, you will face a 50 percent probability of receiving 30 ECUs and a 50 percent probability of receiving 0 points.

Situation	Lottery	Sure payment	Choice
1	50% chance of 30 ECUs and 50% chance of 0 ECUs	2	(radio buttons)
2	" "	4	
3		6	
4		8	
5		10	
6		12	
7		14	
8		16	
9		18	
10		20	
11		22	
12		24	
13		26	
14		28	

Decision task 2: Opportunity to make a donation

In this part you can choose how to allocate **$10** between yourself and a charity of your choosing. The 10 dollars you are allocating in this section are new and do not come out of your earnings from the previous phases. You will decide how much of the **$10** to keep for yourself and how much to pass to your selected charity. You may elect to keep it all for yourself and pass nothing to the charity, keep nothing for yourself and pass it all to the charity, or keep some for yourself and pass the remainder to the charity.

The choice you make in this part will be implemented at the end of the experiment with a 10% probability. That is, on average, 1 out of every 10 participants will receive some money and/or generate money for a charity based on his/her decision in this part.

To make your choice you will need to select ONE charity from the list provided and the amount, if any, you wish to pass to your charity of choice.

Please click "Continue" to make your decisions.

Procedures Questionnaire

Please respond to the following items by clicking the number on the rating scale that best represents your opinion about that item.

	Strongly Disagree				Strongly Agree
1. The money you passed to the Charity will be sent to the charity.	1	2	3	4	5
2. The instructions about the task were clear and easy to follow	1	2	3	4	5
3. The instructions about my compensation were clear and easy to follow	1	2	3	4	5
4. The recipients of donations to the Charity are deserving of support	1	2	3	4	5

General Questionnaire

Please, answer the following questions. There are no correct or wrong answers. Your responses are completely confidential.

1. Gender: Male [ ] Female [ ]

2. What is your ethnic group? Tick the appropriate box to indicate your ethnic group.

[ ] White [ ] Asian [ ] Black [ ] Hispanic or Latino [ ] None of these [ ] Would rather not say

3. What is your age?: _____

4. Is English your native language?: [ ] Yes [ ] No

5. : What is your area of study:

[ ] Arts and Sciences [ ] Business [ ] Education [ ] Engineering and Computing [ ] Law [ ] Medicine [ ] Other

6. Year of study at University:

[ ] Freshman [ ] Sophomore [ ] Junior [ ] Senior [ ] Graduated

7. What is the highest level of education you expect to complete?
[ ] Bachelor's degree [ ] Master's degree [ ] Doctoral degree [ ] Other professional degree [ ] None of these

8. After graduation, what type of jobs will you seek out?
[ ] For-profit sector [ ] Non-profit sector [ ] Government [ ] Any of these sector, will seek best fit [ ] Any of these, will seek highest payment

9. Do you currently work part-time, full-time, or neither? [ ] Part-time [ ] Full-time [ ] Neither

10. If yes, how much is your hourly wage? _____

11. Do you regularly attend religious services? Yes [ ] No [ ]

12. In the last 12 months, did you make a monetary donation to a charitable organization? Yes [ ] No [ ]

13. In the last 12 months, did you donate goods or clothes to a charity? Yes [ ] No [ ]

14. In the last 12 months, did you do volunteer work for a charity? Yes [ ] No [ ]

15. In the last 12 months, have you donated blood? Yes [ ] No [ ]

On a scale of 1 to 5, please indicate the degree to which you agree or disagree with the following statements. “1” indicates *strongly disagree*, “5” indicates that you *strongly agree*, and “3” indicates that you neither agree nor disagree.

- Politics is a dirty word.
- I respect public officials who can turn a good idea into law.
- Ethical behavior of public officials is as important as competence.
- The give and take of public policy making doesn't appeal to me.
- I don't care much for politicians.

- People may talk about the public interest, but they are really concerned only about their self-interest.
- It is hard for me to get intensely interested in what is going on in my community.
- I unselfishly contribute to my community.
- Meaningful public service is very important to me.
- I would prefer seeing public officials do what is best for the whole community even if it harmed my interests.
- An official's obligation to the public should always come before loyalty to superiors.
- I consider public service my civic duty.

- I believe that there are many public causes worth championing.
- I do not believe that government can do much to make society fairer.
- If any group does not share in the prosperity of our society, then we are all worse off.
- I am willing to use every ounce of my energy to make the world a more just place.
- I am not afraid to go to bat for the rights of others even if it means I will be ridiculed.

- When public officials take an oath of office, I believe they accept obligations not expected of other citizens.
- I am willing to go great lengths to fulfill my obligations to my country.
- Public service is one of the highest forms of citizenship.
- I believe everyone has a moral commitment to civic affairs no matter how busy they are.
- I have an obligation to look after those less well off.
- To me, the phrase "duty, honor, and country" stirs deeply felt emotions.
- It is my responsibility to help solve problems arising from interdependencies among people.

- I am rarely moved by the plight of the underprivileged.
- Most social programs are too vital to do without.
- It is difficult for me to contain my feelings when I see people in distress.
- To me, patriotism includes seeing to the welfare of others.
- I seldom think about the welfare of people whom I don't know personally.
- I am often reminded by daily events about how dependent we are on one another.
- I have little compassion for people in need who are unwilling to take the first step to help themselves.
- There are few public programs that I wholeheartedly support.

- Making a difference in society means more to me than personal achievements.
- I believe in putting duty before self.
- Doing well financially is definitely more important to me than doing good deeds.
- Much of what I do is for a cause bigger than myself.
- Serving citizens would give me a good feeling even if no one paid me for it.
- I feel people should give back to society more than they get from it.
- I am one of those rare people who would risk personal loss to help someone else.
- I am prepared to make enormous sacrifices for the good of society.

List of Charities

South Carolina (in-presence sessions):	Pennsylvania (online session):
Harvest Hope Food Bank	Greater Pittsburgh Community Food Bank
The South Carolina Historical Society	Heinz History Center
Make-a-wish Foundation of South Carolina	Make-a-wish Foundation of Greater PA and WV
Coastal Conservation League	Western Pennsylvania Conservancy
ETV Endowment of South Carolina	WQED Multimedia
United Way of the Midlands, South Carolina	United Way of Allegheny County
Humane Society Columbia	Western Pennsylvania Humane Society
South Carolina Cancer Alliance	Mario Lemieux Foundation