The London School of Economics and Political Science

Essays on Cooperation

Roberto Claudio Sormani

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made. This thesis may not be reproduced without my prior written consent.

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of about 30,000 words, and less than 35,000.

Statement of Conjoint work

I confirm that Chapter 2 was jointly co-authored with Marco Hernandez, Julian Jamison, Ewa Korczyc, and Nina Mazar. I contributed 45% of the work in this chapter. In particular I contributed to the definition of the treatments and proposed the *Omission Taxpayer* intervention; I lead the treatment assignment, randomisation, and design procedure; I oversaw the econometric analysis presented in the chapter; I personally organised the structure and the conclusions of most of this chapter, based on the conjoint work with my coauthors. .

Statement of inclusion of previous work

I can confirm that chapter 1 was the result of previous study for an MRes degree I undertook at the London School of Economics and Political Science.

Acknowledgments

These essays are the outcome of years of work, thoughts, and effort to address a question that has always been at the centre of my interests, namely how can we use economics to affect people's contribution to the improvement of the world we temporarily share. They are also the outcome of an immense struggle with myself and with the unknown that is inherent in any research endeavour. If I started this journey at all, and if it is now coming to a close, I owe it to the support, the encouragement, the challenging stimulus given to me by so many people. A non-exhaustive list follows.

I want to thank Muizz Aziz, Marcella Cartledge and Yari Napolitano provided excellent research assistance and actively contributed to the realisation of this project.

Thanks to Tamara Anson and Kruthika Kamath from the LSE Behavioural Research Lab (BRL) were extremely helpful in making the best use of the facilities that were provided by the BRL and sharing their insight. Thanks to Rose Harris from ESRC, for her incredible help in the whole PhD journey and helped me administer as best I could the research grants from ESRC. Thanks to Jane Dickson and Angela Swain from STICERD, who helped me with the logistics and the administration of the research grant that was provided by STICERD. Thanks to Mark Wilbor for his kind and helping work in the Department of Economics.

Thanks to ESRC, LSE Department of Economics, and STICERD for financial support to this research. Thanks to *Avvocatura per i Diritti LGBTI - Rete Lenford* and its President, Maria Grazia Sangalli, who do every day an amazing job and agreed to share a platform for the realisation of this project. Thanks to Sir Thomas Hughes-Hallett, Marie Curie Cancer Trust, The Good Company, who agreed to share their time to exchange precious information and gain a better understanding of the third sector.

Thanks to Oriana Bandiera for her insightful supervision but most of all for being a huge source of inspiration and encouragement in these years. Thanks to Maitreesh Ghatak for his supervision and coaching, pushing me to new directions and stimulating curiosity. Thanks to Nava Ashraf for her excellent supervision and devoting 100% of her attention every time one needed her insight, and for her help in a better design of the experimental part of this research.

Thanks to my PhD examiners Simon Gächter and Matteo Galizzi for their precious comments and inisghts. Thanks to Tim Besley, Gharad Bryan, Robin Burgess, Erik Eyster, Gregory Fisher, Matthew Levis, Drazen Prelec, Guido Tabellini for their helpful comments. Thanks to participants in LSE work in progress seminars and EDP conference in Brussels 2013 for helpful comments.

Thanks to Pedro Alves, Michel Azulai, Florian Blum, Elia Boe, Shuai Chen, Laura Derksen, Erika Deserranno, Andrea Di Miceli, Dana Kassem, Nicola Limodio, Stephan Maurer, Panos Mavrokonstantis, Francesco Sannino, Arthur Seibold, Manuel Staab, Giulia Tagliaferri, Lisa Windsteiger, Giulia Zane for helpful comments and friendly presence in my life.

Abstract

These essays provide arguments and evidence as to circumstances that encourage or hinder cooperation.

Chapter 1 theoretically argues that incentives to participation in a cooperative environment can increase participation, but disrupt cooperation itself. Following Bénabou and Tirole's (2003) crowding-out theory, I propose a model where cooperation is strategically complementary and principals face a trade off between pay and cooperativeness of agents. If such trade off is anticipated by agents, this can lead to multiple equilibria where the informative power of incentives disrupts cooperation.

Chapter 2 presents results from a randomised controlled trial in Poland which used reminder letters to promote voluntary compliance among 150,122 taxpayers who declared their Personal Income Tax but had failed to pay by the deadline. Taxpayers were randomly allocated to receive the letter originally used by the authorities or one of nine letters adapted using behavioural design. Among other results, we find that, relative to a control "behavioural" letter, there is a significant negative effect of a "social norm" message informing of the high frequency of compliant taxpayers. There is also a significant negative effect of a "public goods" messages that appeals to cooperation on the taxpayers' side by reminding them of the role of taxes to support the services provided by the government. In this context, therefore, we do not find evidence of reciprocity or conditional cooperation.

Chapter 3 presents results from a lab-in-the-field experiment conducted at the LSE in 2017. Experimental subjects are asked to take part in a two-person team task for an LGBTI charity. The production function exhibits perfect complementarities in individual levels of effort. Subjects are asked to decide whether they want to volunteer their compensation, are then randomly assigned a partner and informed whether their partner volunteereed. I find that being matched with a volunteer increases effort among those attending, in particular those who have reported more interest in working for the cause. Charities are thus particularly affected by the composition of their teams in terms of career choice and motivation.

Contents

Chapter 1

Few But Good. Financial Incentives to Participate and Cooperation

1.1 Introduction

Does pay decrease cooperation? Sizeable and numerous strands of literature address this questions from a variety of angles, yet little attention has been devoted to the role of expectations and beliefs in this correlation. Yet this channel is a very relevant one as it is pervasive in individuals' decision making, it can spill down to a wide variety of different contexts and it can be directly fine-tuned as a relatively low-cost policy instrument.

To study this correlation and the role of expectations I consider a model whereby a principal offers a bonus to recruit agents who take part in a project. There are two agents and, if both participate, they may decide to cooperate or not. Cooperation is strategically complementary as it takes unilateral action and is rewarded only when jointly undertaken. As agents differ in their cost of cooperation (motivation or cooperativeness), so does the reservation wage they require to take part in the project. The principal, who cares both participation, monetary costs, and cooperation in the team, faces a trade off between higher pay and higher likelihood of cooperation arising.

The role of information is separately understood by introducing different states of the world: a good one and a bad one with higher or lower frequency of motivated agents in the population. I separately study the correlation between pay and cooperation arising when the principal cannot observe the state of the world, and the causal effect that pay can have on cooperation when the principal can access more information.

I find that the trade off depends on primitives such as the principal's returns to participation and

cooperation, agents' value from being assured that the environment is cooperative, and the distribution of motivated (cooperative) types. I show that, when the principal can observe the state of the world, the negative correlation between pay and cooperation becomes stronger and, under some circumstances. Under some circumstances I find that information gives rise to multiple equilibria: one with cooperation and full participation, and one with a trade off between participation and cooperation. In the equilibrium with full participation and cooperation, agents do not interpret (correctly) higher pay as a signal of a bad state of the world. I thus conclude that the availability of information on the principal's side exacerbates the negative correlation between pay and cooperation and can actively disrupt cooperation if agents use pay as a signal.

To appreciate the relation between pay and cooperation, consider the following stylised scenario: a local government is supporting a training course for teachers in the community and has to decide how much to compensate for participation. As many educational experiences, the training course might involve formal or informal cooperative learning (Johnson and Johnson 1989), whose success and effectiveness depend on every participant's effort and contribution. Yet there is a strategic complementarity, as making a contribution might involve the sharing of personal experiences, the effort of thinking harder into the problem at hand and focusing better. More motivated participants will find it easier and to make this cooperative effort, but be happy to take it if they are assured that their course mates are equally motivated. If they can be assured of this, they will need little compensation to participate as they will enjoy the outcome of cooperation that they expect to arise during the activity. Less motivated agents need higher compensation, as they do not expect high rewards from engaging in cooperative effort that they might not even be interested to undertake. There are then two possible scenarios: a high bonus scenario, with more and less motivated agents, and a low bonus scenario, limited to the less motivated who, assured by the bonus that only motivated agents are participating, will be happy to cooperate. From the provider's point of view this is a trade off between participation on the one hand, and cooperation (together with the ability to extract the "cooperative surplus" from motivated agents) on the other hand. Which option gets chosen will most likely depend on how much the local administration cares about participation vis a vis cooperation.

In this framework, what happens if participants have little experience with this training course (for instance because it involves the discussion of novel teaching techniques) and little idea about how interested their peers are in taking part in this course? If they anticipate the principal's trade off, higher pay will not only discourage cooperation by bringing in less cooperative people, but it will additionally reveal that the administration does not expect the general population to be motivated enough to cooperate. How strong is this effect, what does it imply for the likelihood of observing cooperation and for agents' welfare? The model is aimed at addressing such questions.

The issue is relevant in a number of circumstances in which a group activity gives scope for cooperation

among agents who do not know well about each other and with heterogeneous motivation for cooperation. It is mostly suited to the study of mission oriented organisations, where cooperation is very important and social interaction an important driver of participation, especially when it comes to volunteering. In some Jewish communities, for instance, rabbis manage courses for teenagers to learn rhetoric techniques against antisemitism. Cooperation among participants makes these courses more effective as participants can share their opinions, question each other, manage mock-debates autonomously, make the classes more memorable and make the activity so enjoyable that they will want to repeat a similar experience.

The importance of cooperative interactions within an organisation is however more widespread. Many private companies pride themselves for their cooperative environment and make cooperation a basic element of their organisational identity. Offering a more cooperative environment, however, might be an intangible benefit offered by the company to their employees, potentially creating a wage differential with similar, less cooperative, companies. In this paper we explore the possibility that this wage differential is self-sustaining and that offering higher wage might bring into the workers' pool a number of people who care less about the intangible benefit of cooperation. If the share of these new people is too high, higher wages might discourage everybody from cooperating and create a whole different work environment.

Some companies or organisations might, however, have a strong reputation for their cooperative work environment and their propensity to team work. Some others, especially start-ups, are not known by the wider public. The informative power of pay can be very relevant in how these organisations position themselves in the quest for motivated contributors, be they workers or volunteers.

This paper inserts itself in the broader economic literature of incentives and the possibility of "dysfunctional behavioural responses" (Prendergast 1999). Arguments for the possible counter-effectiveness of monetary incentives have been moved based on their distortionary effects on the allocation of agents' effort (Holmstrom and Milgrom, 1991).

Research in industrial psychology (Deci 1971, Cameron and Pierce 1994, Eisenberger and Cameron 1996) has additionally considered the problem of crowding out of motivation and how high incentives may disrupt "work morale" (Frey 1994). Economics has addressed crowding out with additional empirical research (ranging from confirmation of the prediction, as in Gneezy and Rustichini 2000, to disconfirmation as in Ashraf et al 2014) and explaining it by accounting for the possibility that incentives provide information about the task (Bénabou and Tirole 2003) or that they inform of the principal's expectations.

Other related economic literatures are those that interpret differences in pay as compensating wage differentials for how suitable a job is to a worker, such as Besley and Ghatak (2005) who find a rationale for matching on the dimension of mission orientation and Akerlof and Kranton (2005) who show that organisational identity can be a substitute for financial rewards.

A related literature is that on the selective effect of incentives, in particular when it comes to organisations that are driven by a mission. Dal Bó et al (2013) discuss evidence from recruitment of public workers in Mexico. Lazear et al (2012) instead show in the lab that incentives to participate to an environment in which sharing is possible attract individuals who are less willing to share. Field experiments provide mixed evidence on whether high pay attracts less motivated agents in the provision of public services is discussed in Deserranno (2019) and Ashraf et al (2018).

The consideration for the role of social interactions within the organisation owes to the research on reciprocity and conditional cooperation (e.g. Fehr and Gächter 2000, 2001, Fehr and List 2004) and how these affect the role of incentives in an organisation (Fehr et al 2000).

In this context, pay structure is found to have strong effects on the cooperativeness of the work environment, as noted by Lazear (1989) and further studied empirically by Bandiera et al (2005, 2010).

In the context of these literatures, this paper argues for an additional reason why incentives might be counter-effective and discourage cooperation, by jointly accounting for what type of agents are selected in by incentives and how agents decision to cooperate depend on the social environment they find in the organisation.

The paper is structured as follows. Section 2 shows a simple model with only one state of the world and no other informational asymmetries over the fact that each agent's motivation is private. This section shows the relation between the distribution of the population and motivated agents' decision to cooperate. In section 3 I extend the model to account for the informational effect of pay, drawing on Bénabou and Tirole (2003) who posit that incentives to participate in a task signal its quality. In section 4 I consider some welfare implications. Section 5 discusses the findings and their relevance. Section 6 concludes.

1.2 The Basic Model

The first basic model in this section highlights the importance of selection on agents' decision to cooperate in the presence of strategic complementarities. The driving mechanism is heterogeneity in relative preferences for quality of participation (which depends on cooperation, which in turn is strategically complementary) and the ensuing negative correlation between agents' reservation wage and the degree of motivation. As cooperation is strategically complementary, then, there is a trade off between participation and cooperation whose size depends on the share of cooperative agents in the population.

The model has three players: one principal indexed by 0 and two agents indexed by $i \in I = \{1, 2\}$. The principal has mass 1 while each agent has mass $\frac{1}{2}$. The game unfolds in three stages $t = 0, 1, 2$. At time $t = 0$ each agent observes a cost of effort $h_i \in H_i \equiv \{h_L, h_H\}$, with $0 < h_L < h_H$, $h_1 \perp h_2$ and probability

distribution $\theta : H_i \to (0,1)$ given by

$$\theta(h) = \begin{cases} \theta & \text{if } h = h_L; \\ 1 - \theta & \text{if } h = h_H. \end{cases}$$

Cost of effort is here interpreted as motivation or cooperativeness equivalently, as the project yields individual returns that depend on the positive interaction with the partner. Interaction is however more or less costly for agents with different degrees of motivation for the project (e.g. with different attachment to the cause) or who find it more or less difficult to make their own contribution (e.g. asking questions or sharing personal experiences to contribute to a group discussion). Note that h_i encompasses impure altruism (Andreoni 1990) as well as intrinsic motivation from cooperating in the cause. Returns to cooperation are assumed to be homogeneous across the population, but this does not significantly affect the predictions of the model.

Each agent's cost h_i is private information to them and is unknown to the other agent and the principal. At time $t = 0$ the principal sets a bonus $b \in B \equiv \mathbb{R}$ to be offered to each agent and to be paid to each agent if they participate at time $t = 1$. At time $t = 1$ agents 1 and 2 simultaneously decide whether they want to participate in the project or not. If they participate they earn bonus b but incur a cost of participation $f > 0$. If both participate the game moves to stage $t = 2$; if any agent does not participate, the game ends at $t = 1$. At time 2 the subjects decide simultaneously whether they want to cooperate. Cooperative effort costs h_i to each agent but yields benefit $\gamma > 0$ which is realised if and only if both agents cooperate. At the end of time 2 payoffs are realised. The principal obtains gross payoff equal to $\frac{\alpha}{2}$ for each agent participating in the project, and $\beta \geq 0$ if both agents cooperate.

Assumption 1. *Unmotivated agents find cooperation always too costly; motivated agents find it worthy to cooperate if they are certain that their partner cooperates too. That is,*

$$0 < h_L < \gamma < h_H \quad .$$

Assumption 2. *Participation of agents is always worthy to the principal if she can compensate their participation (Internal Rationality):*

$$\alpha > f \quad .$$

To summarise, the principal's strategy is $b \in B \equiv \mathbb{R}$. Each agent i's action space is $A_i = \{(p_i, c_i)\} = \{0,1\} \times \{0,1\}$, where p_i is a binary variable equal to 1 if i participates, and c_i is a binary variable taking on value 1 if i cooperates.

The final payoffs can thus be written as $\pi(b, a_1, a_2)$ and $u_i(b, a_1, a_2; h_i) = u(b, a_1, a_2; h_i)$ with

$$
\begin{aligned}
\pi(b, p_1, c_1, p_2, c_2) &= (p_1 + p_2)\frac{\alpha - b}{2} + p_1 p_2 c_1 c_2 \beta \quad , \\
u_i(b, p_i, p_j, c_i, c_j; h_i) &= u^1(b, p_i) + p_i p_j u^2(c_i, c_j; h_i) \quad , \\
u^1(b, p_i) &= p_i(b - f) \quad , \\
u^2(c_i, c_j; h_i) &= c_i(\gamma c_j - h_i) \quad .
\end{aligned}
$$

A strategy for agent i is a pair of functions $s_i = (p_i : B \times H_i \to [0,1], c_i : B \times H_i \to [0,1],)$. For the principal, a strategy is simply the bonus b offered to all agents.

Now, a conjecture on agent $j \in \{1, 2\}$ is a function $\boldsymbol{\mu^j} : H_j \times B \to [0, 1]$ representing the probability that agent $i \neq j$ assigns to j being of type $h \in H_j$, conditional on having observed bonus b and agent j having decided to participate at stage 1. If stage 2 occurs, at the beginning of stage 2 i thus uses $\boldsymbol{\mu^j}$ to assess the probability that agent j is of type h, with $\mu^j(h; b) \geq 0 \quad \forall h$, and $\mu^j(h_L; b) + \mu^j(h_H; b) = 1$. Note that $\mu^j(\cdot; b)$ is conditional on agent j having participated at stage 1.

We can now define the equilibrium of the game.

Definition 1. *An equilibrium of the game is a profile of strategies and conjectures*

$$
\left(b^*, (s_i^* = (p_i^*, c_i^*))_{i=1,2}, \boldsymbol{\mu^1}, \boldsymbol{\mu^2}\right) \tag{1.1}
$$

such that

$$
\begin{aligned}
c_i^*(b; h_i) &\in \arg\max_{c_i} \sum_{h_j \in H_j} u^2(c_i, c_j^*(b; h_j); h_i)\mu^j(h_j; b) && \forall i, j \in \{1, 2\}, i \neq j, \forall b \in B, \forall h_i \in H_i \quad ; \\
p_i^*(b; h_i) &\in \arg\max_{p_i} \sum_{h_j \in H_j} u\left(b, p_i, p_j^*(b; h_j), c_i^*(b; h_i), c_j^*(b; h_j); h_i\right)\theta(h_j) && \forall i, j \in \{1, 2\}, i \neq j, \forall b \in B, \forall h_i \in H \quad ; \\
b^* &\in \arg\max_b \sum_{(h_1, h_2) \in H_1 \times H_2} \pi(b, s_1^*(b; h_1), a_2^*(b; h_2))\theta(h_1)\theta(h_2) \quad ; \\
\mu^j(h; b) &= \frac{\theta(h)p_j^*(b, h)}{\sum_{h' \in H} \theta(h')p_j^*(b, h')} && \forall b : \sum_{h' \in H_j} p_j^*(b, h') > 0, \quad \forall h \in H_j, \forall j \in \{1, 2\} \quad .
\end{aligned}
$$

1.2.1 Equilibrium

Let us now show what equilibria can arise from the game.

Proposition 1. *Consider equilibria where agents participate and cooperate whenever it is optimal, and*

both agents play the same strategy. In any such equilibrium, unmotivated agents do not cooperate and they participate whenever the bonus offered covers the participation costs. For motivated agents, the decision to cooperate is non-increasing in the bonus level over the region where their equilibrium strategy is to participate. That is, $c_i^*(b, h_L)$ is non-increasing in b over $\{b \in B : p_i^*(b, h_L) = 1\}$.

If the share of motivated agents in the population is not large enough ($\theta < \frac{h_L}{\gamma}$), motivated agents cooperate only when the bonus offered is low enough to discourage unmotivated agents from participating, and the principal offers a low bonus iff

$$\theta^2(\alpha + \beta + \gamma - h_L) - \theta f \geq \alpha - f \quad . \tag{1.2}$$

Condition (??) is satisfied for all θ larger than a threshold value $\underline{\theta} \in (0, 1)$.

If the share of motivated agents in the population is large enough ($\theta \geq \frac{h_L}{\gamma}$), motivated agents always cooperate in the second stage, and the principal offers a low bonus iff

$$\theta^2(\alpha + \gamma - h_L) - \theta f \geq \alpha - f \quad . \tag{1.3}$$

Condition (??) is satisfied for all θ larger than threshold value $\underline{\theta} \in (\underline{\theta}, 1)$.

Proof. First, let us show that unmotivated agents must not cooperate in any such equilibrium: $c_i^*(b, h_H) = 0$.

By Assumption **??**, unmotivated agents always find it convenient not to cooperate if they are at stage 2 if there is any type that cooperates at stage 2, as payoff from the second stage is maximised at

$$u^2(b, 0, c_j^*; h_H) = 0 \geq \sum_{\{h_j : c_j^*(b; h_j) = 1\}} (\gamma - h_H)\mu^j(h_j; b) = u^2(b, 1, c_j^*; h_H) \quad .$$

For the limit case where the two are equal, note that this could only be sustained if $c_j^*(b; h_j) = 0 \quad \forall h_j$, but in that case the symmetry condition requires that player i still does not cooperate either, when facing $h_i = h_H$.

Unmotivated agents thus never cooperate and prefer to participate if and only if the bonus is greater than their participation cost: $p_i^*(b; h_H) = \mathbf{1}(b \geq f)$.

As motivated agents' payoff from participation is not less than $b - f$, whenever $b \geq f$ they will participate too, therefore by bayesian updating it must be $\mu^j(h_L; b) = \theta \quad \forall b \geq f$. Motivated agents cooperate only if $\mu^j(b, h_L)\gamma \geq h_L$, so $c_i^*(b, h_L) = 1 \quad \forall b \geq f$ if and only if $\theta \geq \frac{h_L}{\gamma}$.

At lower levels of b, if there is any participation, it then must be that $\mu^j(h_L; b) = 1$ and therefore motivated agents always cooperate if the bonus is less than f and they participate. Given this, motivated

agents participate whenever $b - f + \theta(\gamma - h_L) \geq 0$, so

$$p_i^*(b; h_L) = \mathbf{1}(b \geq f - \theta(\gamma - h_L)) \quad .$$

Thus, for bonus levels that make motivated agents participate, the equilibrium decision to cooperate is

$$c_i^*(b; h_L) = \begin{cases} \mathbf{1}\left(f - \theta(\gamma - h_L) \leq b < f\right) & \text{if } \theta < \frac{h_L}{\gamma} \\ \mathbf{1}\left(b \geq f - \theta(\gamma - h_L)\right) & \text{if } \theta \geq \frac{h_L}{\gamma} \end{cases},$$

which is non-increasing in the bonus level.

The decision of which bonus is offered on equilibrium thus depends on the principal's incentive compatibility constraint

$$\max_b \sum_{h_1, h_2} \left[(p_1^*(b; h_1) + p_2^*(b; h_2)) \frac{\alpha - b}{2} + p_1^*(b, h_1) p_2^*(b, h_2) c_1^*(b, h_1) c_2^*(b, h_2) \beta \right] \theta(h_1) \theta(h_2)$$

Case 1: $\theta < \frac{h_L}{\gamma}$.

The principal's problem thus becomes

$$\max_b \begin{cases} 0 & \text{if } b < f - \theta(\gamma - h_L), \\ \theta(\alpha - b) + \theta^2 \beta & \text{if } f - \theta(\gamma - h_L) \leq b < f, \\ \alpha - b & \text{if } b \geq f. \end{cases}$$

The principal will thus choose to offer $b = f - \theta(\gamma - h_L)$ if the incentive compatibility constraint is satisfied:

$$\theta\left((\alpha - f + \theta(\gamma - h_L)) + \theta^2 \beta \geq \alpha - f \quad . \right. \tag{1.4}$$

We can see that there is a unique value $\underline{\theta}$, strictly less than 1, such that condition (**??**) holds for all $\theta \geq \underline{\theta}$. Indeed, rearranging (**??**) we get

$$\theta^2(\beta + \gamma - h_L) \geq (\alpha - f)(1 - \theta) \tag{1.5}$$

whose LHS is strictly increasing over $\theta \in [0, 1]$, while the RHS is strictly decreasing, with $LHS < RHS$ at 0 and $LHS > RHS$ at 1.

Case 2: $\theta \geq \frac{h_L}{\gamma}$.

In this case, the principal's problem is

$$\max_b \begin{cases} 0 & \text{if } b < f - \theta(\gamma - h_L), \\ \theta(\alpha - b) + \theta^2\beta & \text{if } f - \theta(\gamma - h_L) \le b < f, \\ \alpha - b + \theta^2\beta & \text{if } b \ge f. \end{cases}$$

The principal's incentive compatibility constraint for offering a low bonus is

$$\theta\left(\alpha - f + \theta(\gamma - h_L)\right) + \theta^2\beta \ge \alpha - f + \theta^2\beta \tag{1.6}$$

Note that this ICC is a stronger condition than the ICC for the case where $\theta < \frac{h_L}{\gamma}$, the reason being that, for the principal, the expected returns from cooperation can now also arise when offering the higher bonus and selecting a larger pool of participants.

Rearranging (**??**) we obtain

$$\theta^2(\gamma - h_L) \ge (\alpha - f)(1 - \theta) \quad , \tag{1.7}$$

whose LHS is strictly increasing over $\theta \in [0, 1]$ while the RHS is strictly decreasing over the same interval, with $0 = LHS < RHS$ at $\theta = 0$ and $LHS > 0 = RHS$ at $\theta = 1$. There must then be a value $\underline{\theta} \in (0, 1)$ such that (**??**) holds for all $\theta \ge \underline{\theta}$.

Furthermore, $\underline{\underline{\theta}} > \underline{\theta}$ as the RHS is the same in both (**??**) and (**??**), but the LHS at (**??**) is strictly greater than that at (**??**).

There is then a unique value $\underline{\underline{\theta}} \in (\underline{\theta}, 1)$, such that condition (**??**) holds for all $\theta \ge \underline{\underline{\theta}}$.

Semi-separating equilibria with unmotivated agents participate with positive probability, strictly less than 1, at $b = f$, and this probability is low enough for motivated agents to still cooperate, are ruled out by the requirement that on equilibrium all agents participate whenever it is rational to do so.

This thus concludes the proof.

□

The equilibria imply three different equilibrium outcomes.

If $\theta < \min\{\frac{h_L}{\gamma}, \underline{\theta}\}$ motivated agents only cooperate if the principal offers a low bonus, but the principal's ICC prescribes them to offer a high bonus $b = f$. All agents participate and no one makes cooperative effort. Expected welfare in this case is $\alpha - f$.

If $\frac{h_L}{\gamma} \le \theta < \underline{\underline{\theta}}$ motivated agents cooperate in any case and the principal offers a high bonus. Expected welfare is thus $\alpha - f + \theta^2(\beta + \gamma - h_L) - \theta(1 - \theta)h_L$.

In all other cases, the principal offers a low bonus and motivated agents participate and cooperate.

Expected welfare is then $\theta^2(\alpha - f + \beta + \gamma - h_L)$.

1.3 Imperfect information on own returns

This section shows how the presence of uncertainty about agents' motivation affects the negative correlation between pay and cooperation. I model uncertainty about the agents' distribution as the presence of two states of the world that differ on how frequent motivated types are in the population. If the principal cannot observe the state of the world, the principal's choice between a higher and a lower pay resembles that of the previous section. If the principal can observe the state of the world, the bonus offered might reveal the observed state and thus increase the negative correlation between pay and cooperation.

There are now two equally likely states of the world $x \in X = \{0, 1\}$. When $x = 1$ it is more likely that agents are cooperative. When $x = 0$ this is less likely. Agents' types are $z_i \in Z_i = \{0, 1\}$, with $z_i = 0$ for an uncooperative type and $z_i = 1$ for a cooperative type, that is $h_i = h_L z_i + h_H(1 - z_i)$. In each state of the world agents' types are independent of one another: $z_1 \perp z_2 | x$. Let $f^x(z)$ be the distribution of z_i when the state of the world is x. That x is the "good" state is reflected in the assumption that $f^1(1) > f^0(1)$. Then, before the game starts, the unconditional distribution of z_i is $f(z_i) \equiv \frac{1}{2}f^0(z_i) + \frac{1}{2}f^1(z_i)$.

After observing z_i and without any information on the state of the world, the probability that the partner is cooperative is positively affected by being cooperative, as each agent observes her own signal to update her beliefs about the state of the world. This is not the main point of the model, but it does not limit its implications. If anything, it allows to allow for projection bias and it adds a specific channel explaining it, that is Bayesian updating. So, for each agent,

$$Pr(x|z_i) = \frac{f^x(z_i)}{\sum_{x'} f^{x'}(z_i)} \quad,$$

and

$$Pr(z_j|z_i) = Pr(x=0|z_i)f^0(z_j) + Pr(x=1|z_i)f^1(z_j)$$

$$= \frac{\sum_{x} f^x(z_i)f^x(z_j)}{\sum_{x'} f^{x'}(z_i)} \quad.$$

Let us define

$$\lambda \equiv Pr(z_j = 1|z_i = 1) = \frac{[f^1(1)]^2 + [f^0(1)]^2}{f^1(1) + f^0(1)}$$

.

It can be shown that $f^1(1) > \lambda > f(1)$, as observing one's own signal is informative of the state of the world and, through that, of the chances the opponent is cooperative.

Let us also define the probability that both agents are motivated as $\varphi \equiv \frac{[f^1(1)]^2 + [f^0(1)]^2}{2}$. Note that $\varphi = \lambda f(1)$ and so $\varphi < [f^1(1)]^2$.

The principal's strategy is a function $b : X \to B$ that assigns a bonus level to each state of the world. We are going to study two cases. In the baseline case, the principal cannot observe the state of the world and therefore $b(x) \equiv b \in B$ in all states of the world; in the case with observed states of the world, the principal observes x and so her strategies are unconstrained functions in B^X.

1.3.1   State of the world unobserved by everyone

At the beginning of time 1 each agent thus forms an **expectation on the population**, i.e. a conjecture on the distribution of types in the population, which is a probability distribution over Z_j, with

$$Pr(z_j|z_i) = \frac{\sum_x f^x(z_i)f^x(z_j)}{\sum_{x'} f^{x'}(z_i)}$$

as shown above. They use this population expectation to evaluate their expected utility in the first stage.

At stage 2, however, they will form an **expectation on participants**, i.e. a conjecture on the teammate's type updated conditional on the teammate having participated at time 1. Such expectation is also a function of own type z_i and of bonus level b, as own type is informative of the state of the world, as seen above, and the bonus level is informative of what type of agents will participate. This is $\boldsymbol{\mu^j}(b, z_i) \in \Delta^{Z_j} = \{t : Z_j \to \mathbb{R}^+, t(z) \geq 0 \quad \forall z \in Z_j, \sum_z t(z) = 1\}$ and it is defined by

$$\mu^j(z_j; b, z_i) = Pr(z_j|p_j = 1; b, z_i) \quad .$$

Expectations on participants will have to satisfy Bayesian updating, based on the strategies played on equilibrium. That is, if strategy p_j^* is played on equilibrium,

$$\mu^j(z_j; b, z_i) = \frac{p_j^*(b, z_j)Pr(z_j|z_i)}{\sum_{z_j'} p_j^*(b, z_j')Pr(z_j'|z_i)} = \frac{p_j^*(b, z_j)\sum_x f^x(z_i)f^x(z_j)}{\sum_{z_j'} p_j^*(b, z_j')\sum_x f^x(z_i)f^x(z_j')} \quad .$$

Definition 2. *An equilibrium of the game is a profile of strategies and conjectures*

$$\left(b^*, \left(\boldsymbol{s_i^*} = (p_i^*, c_i^*), \boldsymbol{\mu^j}\right)_{i,j\in\{1,2\}}\right) \tag{1.8}$$

such that

$$c_i^*(b, z_i) \in \arg\max_{c_i} \sum_{z_j \in Z_j} u^2(b, c_i, c_j^*(b, z_j); z_i)\mu^j(z_j; b, h_i) \quad \forall i \neq j, \forall b \in B, \forall z_i \in Z_i;$$

$$p_i^*(b, z_i) \in \arg\max_{p_i} \sum_{z_j} u_i\left(b, p_i, c_i^*(b, z_i), p_j^*(b, z_j), c_j^*(b, z_j); z_i\right) Pr(z_j|z_i) \quad \forall i \neq j, \forall b \in B, \forall z_i \in Z_i;$$

$$b^* \in \arg\max_{b \in B} \sum_{z_1, z_2} \pi(b, \boldsymbol{s_1^*}(b, z_1), \boldsymbol{s_2^*}(b, z_2))f(z_1)f(z_2) \quad ;$$

$$\mu^j(z_j; b, z_i) = \frac{p_j^*(b, z_j)\sum_x f^x(z_i)f^x(z_j)}{\sum_{z_j'} p_j^*(b, z_j')\sum_x f^x(z_i)f^x(z_j')} \quad \forall b : \sum_{z_j'} p_j^*(b, z_j') > 0 \quad ;$$

with

$$Pr(z_j|z_i) = \frac{\sum_x f^x(z_i)f^x(z_j)}{\sum_{x'} f^{x'}(z_i)} \quad .$$

Equilibria where cooperation is not sustained and the principal offers bonus f are always sustained by "pessimistic" off-equilibrium expectations on participants. What we want to know, however, is the circumstances under which an equilibrium with cooperation can be sustained. Let us characterise equilibria where cooperation is sustained.

Proposition 2. *Consider equilibria where all agents participate and cooperate whenever a set of beliefs or opponents' strategies justifies that, and agents play symmetrical strategies.*

In any such equilibrium, unmotivated agents never cooperate. Motivated agents cooperate for bonus levels lower than f and can cooperate for bonus levels greater than or equal to f if their signal is sufficiently precise on the population's distribution, i.e. if $\lambda\gamma \geq h_L$.

The principal offers a low bonus $b^* = f - \lambda(\gamma - h_L)$ if

$$\alpha - f \leq \frac{f(1)}{f(0)}\lambda(\gamma - h_L) + \frac{\varphi}{f(0)}\beta\mathbf{1}(\lambda\gamma < h_L) \quad . \tag{1.9}$$

The principal offers $b^* = f$ otherwise, or if the above condition holds with equality.

Proof. First note that, for unmotivated agents, at stage two cooperation is never chosen. Participation in

the game, at stage 1 and with backward induction, will thus yield utility $b - f$ to unmotivated agents, so they will participate if and only if $b \geq f$.

For motivated agents, at time 1, $b - f$ is the lower bound on the expected utility from participation, so they always participate when $b \geq f$. By Bayesian updating, then, for such large values of b the expectation on participants should be $\mu^j(z_j; b, z_i) = Pr(z_j|z_i)$ and motivated agents will cooperate if $Pr(z_j = 1|z_i = 1)\gamma \geq h_L$.

We are now looking for equilibria where cooperation is sustained also at lower levels of b. Fix any $b' < f$. For cooperation to be sustained, it must be the case that agents participate, and the participating agents can only be the motivated ones. If that happens, by bayesian updating, $\mu^j(1; b', z_i) = 1$. If so, cooperation arises by Assumption **??** as it yields $\gamma - h_L > 0$. At stage 1, the motivated agent will participate if there is a positive expected benefit, i.e.

$$b' - f + \lambda(\gamma - h_L) \geq 0$$

. With this, we have shown that in any equilibrium where cooperation and participation arise whenever they can be rationalised, motivated agents participate if $b \geq f - \lambda(\gamma - h_L)$ and cooperate whenever $b < f$ or for $b \geq f$ when $\lambda\gamma \geq h_L$.

The principal's problem is thus

$$\max_b \begin{cases} 0 & \text{if } b < f - \lambda(\gamma - h_L) \\ f(1)(\alpha - b) + \varphi\beta & \text{if } f - \lambda(\gamma - h_L) \leq b < f \\ \alpha - b + \varphi\mathbf{1}\left(\lambda\gamma \geq h_L\right)\beta & \text{if } b \geq f. \end{cases}$$

The function has two local maxima: a low bonus of $f - \lambda(\gamma - h_L)$, and a high bonus at f. The choice between the two yields condition (**??**) above.

$\square$

1.3.2 State of the world observed by the principal

In this section we are going to see what happens when the principal can observe, at the beginning of time 0, the state of the world, $x \in X$. In this case, the principal's strategy will be a function $b : X \to B$. Agents' strategies are still $\boldsymbol{s_i} : B \times Z_i \to A_i$. Expectations on the population and on participants now additionally depend on b, as the principal's bonus might provide a signal on the state of the world. We will then define $\boldsymbol{\theta}^j(b, z_i) \in \Delta^{Z_j}$ as the conditional probability assigned to value z_j after observing z_i and principal's action b: $\theta^j(z_j; b, z_i) \equiv Pr(z_j|b, z_i)$.

$$\theta^j(z_j; b, z_i) = \sum_x Pr(z_j|x, b, z_j) Pr(x|b, z_i) = \sum_x f^x(z_j) Pr(x|b, z_i) \quad .$$

On equilibrium, given equilibrium strategy b^*, we will want expectations on the population to be updated bayesianly, therefore

$$Pr(x|b, z_i) = Pr(x|b^*(x) = b, z_i) = \frac{Pr(b^*(x) = b|x, z_i) Pr(x|z_i)}{\sum_{x'} Pr(b^*(x') = b|x', z_i) Pr(x'|z_i)}$$

$$= \frac{\mathbf{1}(b^*(x) = b) f^x(z_i)}{\sum_{x'} \mathbf{1}(b^*(x') = b) f^{x'}(z_i)} \qquad \forall b : \sum_{x'} \mathbf{1}(b^*(x') = b) > 0 \quad ,$$

so the Bayesian updating requirement becomes

$$\theta^j(z_j; b, z_i) = \frac{\sum_x f^x(z_j) \mathbf{1}(b^*(x) = b) f^x(z_i)}{\sum_{x'} \mathbf{1}(b^*(x') = b) f^{x'}(z_i)} \qquad \forall b : \sum_{x'} \mathbf{1}(b^*(x') = b) > 0 \quad .$$

It is easy to see that, if the equilibrium is a separating equilibrium, $\theta^j(z_j; b^*(x), z_i) = f^x(z_j)$, while if it is a pooling equilibrium, $\theta^j(z_j; b^*(x), z_i) = Pr(z_j|z_i)$ exactly like in the scenario where the principal was not observing the state of the world.

Expectations on participants will have to satisfy Bayesian updating, based on the strategies played on equilibrium. That is, if strategy p_j^* is played on equilibrium,

$$\mu^j(z_j; b, z_i) = \frac{p_j^*(b, z_j) \theta^j(z_j; b, z_i)}{\sum_{z_j'} p_j^*(b, z_j') \theta^j(z_j'; b, z_i)} \qquad \forall b : \sum_{z_j'} p_j^*(b, z_j') > 0 \quad .$$

This means that, when only one motivated agents participate (as could be the case if there is cooperation and the bonus is below f), the partner is correctly recognised as motivated; while, when both types of agents participate, there is no updating and expectations on participants remain the same as expectations on the population.

Definition 3. *An equilibrium of the game with asymmetric information is a profile of strategies and conjectures*

$$\left(b^*, \left(\boldsymbol{s_i^*} = (p_i^*, c_i^*), \boldsymbol{\theta^j}, \boldsymbol{\mu^j} \right)_{i,j \in \{1,2\}} \right) \tag{1.10}$$

such that

$$c_i^*(b, z_i) \in \arg\max_{c_i} \sum_{z_j \in Z_j} u^2(b, c_i, c_j^*(b, z_j); z_i) \mu^j(z_j; b, h_i) \quad \forall i \neq j, \forall b \in B, \forall z_i \in Z_i;$$

$$p_i^*(b, z_i) \in \arg\max_{p_i} \sum_{z_j} u_i\left(b, p_i, c_i^*(b, z_i), p_j^*(b, z_j), c_j^*(b, z_j); z_i\right) \theta^j(z_j; b, h_i) \quad \forall i \neq j, \forall b \in B, \forall z_i \in Z_i;$$

$$b^*(x) \in \arg\max_{b \in B} \sum_{z_1, z_2} \pi(b, \boldsymbol{s}_1^*(b, z_1), \boldsymbol{s}_2^*(b, z_2)) f^x(z_1) f^x(z_2) \quad \forall x \in X;$$

$$\theta^j(z_j; b, z_i) = \frac{\sum_x f^x(z_j) \mathbf{1}(b^*(x) = b) f^x(z_i)}{\sum_{x'} \mathbf{1}(b^*(x') = b) f^{x'}(z_i)} \quad \forall b : \sum_{x'} \mathbf{1}(b^*(x') = b) > 0 \quad ;$$

$$\mu^j(z_j; b, z_i) = \frac{p_j^*(b, z_j) \sum_x f^x(z_i) f^x(z_j)}{\sum_{z_j'} p_j^*(b, z_j') \sum_x f^x(z_i) f^x(z_j')} \quad \forall b : \sum_{z_j'} p_j^*(b, z_j') > 0 \quad .$$

Assumption 3. *Cooperation is never worthwhile for unmotivated agents. For motivated agents who are partnered up with a random agent from the population, cooperation is worthwhile in the good state and not worthwhile in the bad state.*

$$h_H > f^1(1)\gamma > h_L > f^0(1)\gamma$$

Let us now characterise equilibria in this framework.

Proposition 3. *Consider equilibria where all agents participate and cooperate whenever a set of beliefs or opponents' strategies justifies that, and agents play symmetrical strategies.*

In any such equilibrium with $b(1) \neq b(0) < b(1) < b(0) = f$ and cooperation arises only if $b(1)$. Such equilibria can occur is the net value that the principal can extract from offering cooperation to the motivated agents $(\gamma - h_L)$ is large enough to be worthy the expected reduction in participation by offering a lower bonus in the good state.

In any equilibrium with $b(1) = b(0)$ it must be that $b(0) = b(1) = f$ and cooperation might or might not arise among motivated agents, depending on the parameters of the problem. Such equilibria can occur for sufficiently large values of $(\gamma - h_L)$ to discourage the principal from deviating and offering a lower bonus.

Proof. Let us start from separating equilibria, i.e. those with $b(0) \neq b(1)$.

Note, first, that in any equilibrium agents participate whenever bonus $b \geq f$ and unmotivated agents do not participate for any lower b, while motivated agents participate at $b < f$ only if cooperation will subsequently arise with sizeable probability. As we are only considering equilibria where all agents participate

whenever possible, we conclude that whenever unmotivated agents participate also the motivated ones must.

Secondly, unmotivated agents never want to cooperate and motivated agents are willing to cooperate for any $b < b(0)$

Secondly, it cannot be the case that everyone participates under both $b(0)$ and $b(1)$. Suppose that were the case, therefore $\min\{b(0), b(1)\} \geq f$. As the population participating would be the same, the decision to cooperate for motivated agents must be the same across the two bonus levels. Then, however, the principal would always have an incentive to the lowest of the two, which is incompatible with equilibrium.

It must then be that $\max\{b(0), b(1)\} = f$, as any higher bonus level would be dominated by a deviation to a lower b' which would preserve full participation and not affect cooperation.

Then if must be $b(0) = f > b(1)$. If the opposite were true ($b(1) = f > b(0)$), then motivated agents would cooperate when offered $b(0)$ (else they would not accept anything less than f) but also under $b(1)$, as $f^1(1)\gamma > h_L$. The principal's incentive compatibility constraints, however, would not be satisfied, as they would imply

$$\alpha - f \geq f^1(1)(\alpha - b(0)) > f^0(1)(\alpha - b(0)) \geq \alpha - f \quad .$$

In any separating equilibrium we thus have $b(0) = f > b(1)$, and cooperation arising under $b(1)$. As we are considering equilibria where agents cooperate whenever their beliefs allow them to do so, they also participate in a separating equilibrium whenever $b \geq f - f^1(1)(\gamma - h_L)$.

The principal's incentive compatibility constraints, however, require the following:

$$\begin{cases} f^1(1)(\alpha - b(1)) + [f^1(1)]^2\beta \geq \alpha - f \\ \alpha - f \geq f^0(1)(\alpha - b(1)) + [f^0(1)]^2\beta \quad , \end{cases}$$

implying

$$b(1) \in \left[\alpha - \frac{\alpha - f}{f^0(1)} + f^0(1)\beta, \alpha - \frac{\alpha - f}{f^1(1)} + f^1(1)\beta\right] \quad .$$

We thus have

$$b^*(1) = \max\left\{f - f^1(1)(\gamma - h_L), \alpha - \frac{\alpha - f}{f^0(1)} + f^0(1)\beta\right\}$$

.

This equilibrium exists as long as

$$\alpha - \frac{\alpha - f}{f^1(1)} + f^1(1)\beta > f - f^1(1)(\gamma - h_L) \quad ,$$

i.e. the following condition holds:

$$\alpha - f \leq \frac{[f^1(1)]^2}{f^1(0)} (\gamma - h_L + \beta) \quad . \tag{1.11}$$

Let us look now at pooling equilibria, that is equilibria with $b(0) = b(1) = \tilde{b}$.

First, note that it must be $\tilde{b} = f$. If $\tilde{b} > f$ there would be a profitable deviation for the principal to some $b' < \tilde{b}$. If, instead, $\tilde{b} < f$, then only motivated agents must participate and cooperate, which in a pooling equilibrium can only happen if $\tilde{b} \geq f - f(1)(\gamma - h_L)$. However, as we are imposing cooperation to arise for any belief that can sustain it, any deviation of the principal to $b' < f - f(1)(\gamma - h_L)$ with $b' \geq f - f^1(1)(\gamma - h_L)$ would induce cooperation as motivated agents would believe the status to be the good one.

Therefore, on any pooling equilibrium there is $\tilde{b} = f$, all agents participating, motivated agents cooperating if and only if $\lambda\gamma - h_L \geq 0$. The principal's best deviation would be to $\hat{b} = f - f^1(1)(\gamma - h_L)$, and the incentive compatibility constraint for the principal (upon observing the good state, as that is the binding constraint) is then

$$\alpha - f + [f^1(1)]^2\beta\mathbf{1}(\lambda\gamma \geq h_L) \geq (f^1(1))(\alpha - f + f^1(1)(\gamma - h_L)) + [f^1(1)]^2\beta \quad , \text{ that is}$$

$$\alpha - f \geq \frac{[f^1(1)]^2}{f^1(0)} (\gamma - h_L + \beta\mathbf{1}(\lambda\gamma < h_L)) \quad . \tag{1.12}$$

To sum up, if $\lambda\gamma < h_L$ then there is a separating equilibrium when

$$\alpha - f \leq \frac{[f^1(1)]^2}{f^1(0)} (\gamma - h_L + \beta)$$

and a pooling equilibrium when

$$\alpha - f \geq \frac{[f^1(1)]^2}{f^1(0)} (\gamma - h_L + \beta) \quad .$$

If $\lambda\gamma \geq h_L$, then there is a separating equilibrium for

$$\alpha - f \leq \frac{[f^1(1)]^2}{f^1(0)} (\gamma - h_L + \beta)$$

and a pooling equilibrium for

$$\alpha - f \geq \frac{[f^1(1)]^2}{f^1(0)} (\gamma - h_L) \quad .$$

$\square$

1.4   Welfare Analysis

1.4.1   Cooperation requires restricted participation

For this part of the welfare analysis we are going to focus on the case with $\lambda\gamma < h_L$, i.e. the case where motivated agents do not cooperate unless they are substantially confident that their partner cooperates.

Let us compare social welfare in the two scenarios where the principal has information on the state of the world and where the principal does not.

Expected welfare in the information scenario is, as shown above,

$$EW^{\text{info}} = \begin{cases} \alpha - f & \text{if } \alpha - f > \bar{\bar{k}} \equiv \frac{[f^1(1)]^2}{f^1(0)}\left(\gamma - h_L + \beta\right), \\[2ex] \frac{1}{2}(\alpha - f) + \frac{1}{2}\left[f^1(1)(\alpha - f) + [f^1(1)]^2(\beta + \gamma - h_L)\right] & \text{if } \alpha - f \leq \bar{\bar{k}}. \end{cases}$$

In the case where the principal cannot observe the state of the world, expected welfare is

$$EW^{\text{no info}} = \begin{cases} \alpha - f & \text{if } \alpha - f > \bar{k} \equiv \frac{f(1)}{f(0)}\lambda(\gamma - h_L) + \frac{\varphi}{f(0)}\beta, \\[2ex] f(1)[\alpha - f] + \varphi(\beta + \gamma - h_L) & \text{if } \alpha - f \leq \bar{k}. \end{cases}$$

Note that $\bar{\bar{k}} > \bar{k}$.

For $\alpha - f > \bar{\bar{k}}$ the principal would act the same way regardless of whether they have information. For $\bar{k} < \alpha - f \leq \bar{\bar{k}}$, the principal has an incentive to increase social welfare when they observe a good state, as then she is confident enough to decrease participation while improving the chances of cooperation by offering a lower bonus. This happens both because the principal directly cares about cooperation, and because cooperation allows her to extract surplus from the agents. Social surplus can be shown to be higher in the information scenario than in the no-info scenario. However, in this region of the parameter space, the surplus created by participation $(\alpha - f)$ is still large enough for the principal not to want to undertake the risk when she does not have sufficient information. The availability of information in the hands of the principal, therefore, makes it easier to bring about cooperation among participants, but creates a trade-off between monetary compensation and cooperation.

For $\alpha - f < \bar{k}$, instead, the principal cares about participation relatively less, to the point of choosing to offer a lower bonus even when she does not know the state of the world. The bonus offered, however, is larger in the no information scenario than in the good state of the information scenario, as the agents need

higher compensation for their lack of information.

1.4.2 Cooperation is compatible with full participation

Let us now see what happens when $\lambda\gamma \geq h_L$, i.e. when cooperation of motivated agents is sustained on equilibrium when everybody participates and under the population beliefs $Pr(z_j = 1|z_i = 1) = \lambda$.

In such case, there is a substantial difference between a pooling equilibrium, where cooperation is sustained under a high bonus, and a separating equilibrium, where higher bonus signals a bad state of the world and thus discourages cooperation. While the conditions for a separating equilibrium remain unchanged, those for a pooling equilibrium become weaker and there is now a range of parameters for which a separating and a pooling equilibrium are both possible. In particular,

- $\alpha - f < \frac{[f^1(1)]^2}{f^1(0)}(\gamma - h_L)$: a separating equilibrium exists, with the principal optimising social welfare conditional on the agents maximising agent's welfare in the second stage and conditional on the principal's information.

- $\frac{[f^1(1)]^2}{f^1(0)}(\gamma - h_L) \leq \alpha - f < \frac{[f^1(1)]^2}{f^1(0)}(\gamma - h_L + \beta)$ a separating equilibrium and a pooling equilibrium exist. In the pooling equilibrium, high bonus $b = f$ is offered, there is full participation and motivated agents cooperate. In the separating equilibrium, there is less-than-full participation and cooperation under the lower bonus that is offered in the good state of the world, or complete participation with no cooperation under the high bonus $b = f$, which is offered in the bad state of the world.

- $\alpha > \frac{[f^1(1)]^2}{f^1(0)}(\gamma - h_L + \beta)$: there is a pooling equilibrium with full participation and cooperative effort undertaken by motivated agents.

The case where the separating and the pooling equilibrium coexist is highly interesting, as it shows that if agents interpret bonus as a signal, a high bonus may discourage cooperation. Taking into account the informative power of bonuses thus exacerbates the negative correlation between bonus offered and cooperation, in ways that go beyond the simple composition of the pool but that crucially depend on the inference made by agents.

Furthermore, it is ambiguous whether expected social welfare would be higher under the pooling or the separating equilibrium. The pooling equilibrium yields higher ex-ante expected social welfare if

$$\alpha - f \geq \frac{[f^1(1)]^2 - 2\varphi}{f^1(0)}(\beta + \gamma) + \frac{2f(1) - [f^1(1)]^2}{f^1(0)}h_L.$$

The reason for this loss in expected social welfare is that, while in the bad state it is unambiguously better to prevent motivated agents from cooperating, in the good state it is not clear whether it is better

to have only motivated agents participate or bring in all agents, with higher participation but some risk (proportional to $f^1(0)$ that motivated agents might needlessly exert cooperative effort without finding a match.

From the point of view of motivated agents, however, the pooling equilibrium always yields higher expected utility than the separating one, where the principal is able to use all information in order to extract the surplus from cooperation, and the principal's choice of a high bonus in the bad state signals that there is no expected benefit from cooperating left.

1.5 Discussion

I have shown that incentives to participate may harm cooperation by worsening the quality of the pool of participants and by signalling that cooperative participants are rare in the population.

A key parameter of interest is $\alpha - f$, the net surplus that can be derived from participation of each agent. In circumstances where the general population is relatively not cooperative, a principal who cares relatively more about participation than about cooperation in the project will have a higher incentive to offer high bonuses, and the high bonus will discourage cooperation by selecting into the pool less cooperative types. This is a quantity-quality type of trade off, with participation raising returns to the principal linearly according to quantity, but quality of the project being a convex function of agents' effort. The reason why the principal might be more interested in the participation rate than in the quality of the project can be due to the principal's career incentives, verifiability of quality, or the principal's time preferences.

Back to the motivating example of the training course sponsored by a local administration, the principal might have political incentives to show that the project was large and involved many participants, either for reelection purposes or for the ability to extract higher rents from donors' contribution to the project. In such case the principal might start a larger scale project with lower educational value for the teachers taking part in the training. Verifiability might affect this choice in that the principal might not be able to show the exact quality of the project and so might be unable to rip the benefits from higher cooperation in the environment. The local authorities, for instance, might not be able to show the quality of the project to their constituents, and will therefore have more of an incentive to focus on its scale.

Time preferences could affect the principal's choice: while participation in the course is immediate, the quality of the learning and how interactive it is might have a small but more long-lasting effect. This could be a problem if the principal discounts future returns relatively highly, for instance because of term limits that prevent the administration from ripping the full stream of benefits from improved teachers' skills.

Additionally, a factor limiting the principal's willingness to restrict the pool of participants is the size of

the compensating differential that they can obtain, by giving agents the assurance that their partners are motivated and likely to cooperate. This boils down to how much heterogeneity there is in the population in terms of motivation and how quickly the motivation of the marginal participant declines as the bonus offered increases. If there is a sizeable difference between more and less motivated teachers, for instance, and if there is a small group of teachers who are willing to take part in the training without any compensation as long as the training is productive and there is scope for cooperative learning, then the principal will have more of an incentive to restrict the pool of participants as that would save more money.

A reason why the principal in this model has interest themselves in keeping the pool restricted is that restricting the pool of agents is the only way they have to ensure cooperation, if they have any at all. The principal is otherwise unable to contract on cooperation with the agents and to reward them from that - something that for high values of β they may well want to do. This impossibility to contract on cooperative effort might be explained by difficult measurability of effort. Also, it might be that cooperation requires more than one dimension of effort and contracting on them might distort multitasking incentives (Holmstrom and Milgrom 1991).

Even if the principal were able to contract on cooperative effort, however, there would need to be sufficiently high returns from effort on the principal's side, or the principal's ability to extract additional surplus obtained by motivated agents in a cooperative environment.

All this can happen when the principal does not have more information on the general population than the individual agents do. As we show in our model, however, the availability of information for the principal has two effects. On the one side, it allows the principal, under a range of parameters, to adapt to the information she receives and share it with the agents, who might benefit from it. Teachers in the training course would benefit from realising that other participants are less likely to be cooperative, and then refrain from exerting cooperative effort which would not be worthwhile. If the principal's interest in cooperation and in participation are somewhat balanced, there is an incentive for the principal to share this information, which can be of benefit to the agents if it refrains them from wasting effort. When agents need reassurance that their partners are highly motivated ($\lambda\gamma < h_L$) the availability of information allows the principal to tell the agents what they need to know.

If, however, the general population is relatively cooperative ($\lambda\gamma \geq h_L$) and the case where the principal raises the bonus is a signal of a very uncooperative population ($f^0(1)\gamma < h_L$), there are multiple equilibria that are themselves sustained by different beliefs on the agents' side. If agents "over-interpret" the principal's bonus, the signalling power of a high bonus will discourage cooperation beyond the mere effect of broadening the pool of participants, which would have not been a problem per se. This creates a strong negative correlation between pay and cooperation, that would be avoidable if agents did not attach much informative

power to the incentives, as in the pooling equilibrium. From the point of view of ex-ante social welfare, it is dubious which of the two scenarios is better.

In the case of the training group highlighted above, the difference would be important depending on how sophisticated participants are, how they believe the administration to behave, and how sophisticated does the administration expect the teachers to be. Depending on how people make their inference we might or might not have cooperation. This particular case raises the importance of assessing agents' beliefs and their response to prices not only when selling products (Bagwell and Riordan 1991, Cohen and Dupas 2010), but also when "selling" participation in an activity which might yield intrinsic benefits to the participant.

When does, however, the power of information held by the principal matter the most? When the principal does not have strong reputation; when the project is relatively original and not generally known of; when the reference community is not small enough and when there is little scope for communication between participants prior to the meeting; when there is little screening in the recruitment process. Back to the example of the training course, the effect of high pay will be more detrimental in larger cities than in small villages where people know each other. It will be less detrimental in organisations that have strong selection practices and a strong identity, thus being able to combine higher pay with a cooperative environment.

1.6 Conclusions

This paper has shown that there are reasons to find a negative correlation between pay and cooperation in a work environment, and that this correlation becomes stronger if incentives are thought to carry some informational power.

The contribution to the literature is to take into account the social dimension of the workplace (Bandiera et al, 2005) and agents' strategic response to that (Fehr et al 2000), while considering pay as a powerful selection device (Besley and Ghatak 2005, Deserranno 2019).

The results complement existing results in the literature on when incentives might give rise to "dysfunctional responses" (Prendergast 1999). The definition of motivation here is deliberately vague and can capture different characteristics, from outright cooperativeness of a person to the person's specific willingness to cooperate within the specific organisation, either because she is aligned to the organisation's mission (Besley and Ghatak 2005) or identity (Akerlof and Kranton 2005). What this paper adds to the existing literature on the matter is that the social environment is considered as a feature of the job in its own right, over which workers have their own heterogeneous preferences and which companies can choose and adapt to the type of workers that they are seeking.

With the modelling choice of distinguishing between an extensive margin and an intensive one, the former

which is contractible upon and the latter which is not, the theory also brings a modest contribution to the literature on the detrimental effect of incentives, which Eisenberger and Cameron (1996) find to crowd out intrinsic motivation to perform only when they are made conditional on the extensive margin. When incentives are conditional on the intensive margin, as would happen in this paper's model, performance is not impacted. The paper thus provides an explanation to the phenomenon of crowding out that is compatible with an established pattern.

As the theory's results are crucially driven by agents' expectations and their ability to anticipate the principal's trade off and information upon observing the pay choice, the model suggests a few empirical and experimental tests about this theory.

At an individual level, the model posits that there is an inverse relationship between reservation bonus or wage and motivation to participate in team work. This relation might be more or less present in different contexts and requires an empirical assessment. In a professional setting, for instance (see Ashraf et al 2018), reservation wage will reflect the level of ability acquired by the worker, which in turn could be positively correlated to the motivation that an agent has for performing in the job. As long as the job requires cooperation and teamwork, then, the negative correlation is less likely to be found. This assumption can be tested empirically case by case provided that good specific measures of cooperation can be found. In an experimental setting, we would need to assess the relationship between reservation wage and willingness to cooperate in a specified setting, such as a public good game. Lazear et al (2012) use a design where they incentivise participation in a dictator game vis a vis an experimental environment with a fixed payment. Raising the pay of the dictator game attracts individuals who are less willing to share. The negative correlation that is found there between reservation bonus and willingness to share should be tested in an environment where effort is exerted, as reservation wage might then be confounded by a large amount of other factors.

Independently of the empirical correlation that could be found between motivation (cooperativeness) and reservation wage, however, what matters to this theory is that agents expect such negative correlation to be there. Empirical tests of this mechanism would rely on elicited beliefs in a cooperative environment where subjects' partners were invited under different reservation bonuses. A design in two stages, for instance, could give half of the subjects the option to choose whether they want to participate in a cooperative game (consider one with the same payoff structure as the model's, with individual monetary reward fixed at $b + x$ and some social return equal to $\beta \min\{e_1, e_2\}$ where e_i is effort in the task. A charity could be compensated, for instance, increasingly in the minimum effort exerted by the two agents in a simple task, such as placing sliding cursors at the centre of a screen) or get b for sure. Incentive x should be varied in such a way to infer the reservation level $\tilde{x}$ that makes them willing to participate in the cooperative task. The true value

of x offered in the first stage will be randomised and participation will depend on whether $x > \tilde{x}$. In the first stage, the other half of the subjects will not have participated.

In the second stage, however, pairs are formed between subjects who played in the first and subjects who did not play in the first stage and each pair has to play the same cooperative game with reward b and the same structure of social returns. Those who did not play in the first stage would be informed of their partner's reservation value $\tilde{x}$ and will not be told whether their partner participated, but will be asked to make a guess about the partner's effort in case the partner participated. Expected effort should be increasing in $\tilde{x}$. Secondly, once they play the game in the second stage, effort should be decreasing in the reservation wage of the partner if agents respond to their expectations.

Empirically, this theory would predict a negative correlation between pay and cooperation, more strongly so when the information about the environment is scarce. To test this one would need to measure the level of cooperation observed in similar organisations (perhaps through questionnaires with an incentive compatible design).

It is harder to identify this theory in opposition to others, as it complements rather than excluding them. Hopefully this model will have convinced the reader of the importance of environmental factors as a job feature that agents might care about and that can be affected by seemingly small policy changes.

Chapter 2

Using Behavioural Insights to Improve Tax Collection: Evidence from Poland

Introduction

Mobilising tax collection efficiently and improving tax compliance has long been an objective of tax authorities and economists alike. For the former, the issue was one of effectiveness and minimising enforcement costs while maximising expected revenues. For economists, the question is a reflection of a deep theoretical debate, on whether taxpayers are strategically driven by expected utility maximisation or by culture and beliefs (tax morale). The convergence of interests in the question has brought to a rich literature on the application of behavioural insights to tax authorities' communication with the taxpayers, and how compliance was improved by messages focused on deterrence as opposed to messages based on other, behaviourally

informed, interventions: in particular, igniting a sense of reciprocity, appealing to the desire to conform to social norms, or framing non-compliance as an active choice as opposed to an omission.

The evidence so far is substantially unanimous on the effectiveness of deterrence-based messages and mixed on messages that highlighted moral suasion (e.g. by highlighting the relationship between compliance and the delivery of public goods) or descriptive social norms. The inconclusiveness of this evidence pushes towards a better understanding of what typologies of countries experience different response to different types of messages. In particular, little research has been made on post-socialist countries. This is surprising as the unique experience of such countries suggests a specific relationship of the citizens with publicly provided goods. Exposure to socialism might uniquely affect the way citizens relate to the government and the way they interpret government's messages and respond to them.

It does so by reporting results from a policy experiment that randomly allocated a variety of letters to be sent to Polish taxpayers to remind them to pay their taxes. The experimental subjects had declared their personal income tax (PIT) for the 2015 fiscal year but had failed to pay what they owed by the deadline. The trial took place between May and June 2016 and covered a total of 150,122 individuals. The control intervention was an informal reminder sent by the authorities via ordinary mail and with the use of simple language and clear information provision. The main interventions, analysed against this control one, mainly consisted of a message linking tax revenues to the delivery of local public goods (in two variants); a message based on social norms and reporting that a certain percentage of taxpayers had paid in the same region as the experimental subject; a deterrence message highlighting the negative consequences from non-compliance (coming in two variants); an *omission-commission* message reporting that not paying taxes would henceforth be considered not a delay but an active choice on the part of the taxpayer (itself coming in two variations).

More than simply replicating existing interventions in the Polish context, however, the large sample size and the pre-existing context allowed to introduce tests that would improve our understanding of compliance behaviour among late taxpayers.

First of all, we could fill the lack of evidence on the effect of simplicity of the tax authorities' communication with respect to content and delivery, separately. Indeed, while it is shown that reminders have a positive effect (Hallsworth et al. 2014) and that simplification works in a context where reminders are not traditionally used (Dwenger et al. 2017), the Polish tax system already had a system of reminders in place that were delivered through formal (so called dunning) letters sent via registered post at a substantial cost for the authorities. The delivery system imposes a cost on taxpayers who, if absent at the time of delivery, had to personally go to the local post office to collect and sign for the dunning letter. A simplified delivery would consist, instead, of informal reminders sent directly to the taxpayer via ordinary mail. Additionally, the traditional reminders are formal and written in complex language. While the formality adds to their

credibility, the complexity of the language made the behaviour requested of the taxpayer much harder to understand. Simplification might then have ambiguous effects when it is introduced in a system tradition-ally blighted with technical jargon. This paper adds to the existing literature by studying in a randomised setting, and separately, the effect of sending a formal letter via ordinary vs. registered mail, and the effect of writing a formal letter vs. an informal simplified message, via ordinary mail.

Secondly, we could address some sources of interaction when it comes to the application of behavioural insights. The use of behavioural insights relies by definition on subtle cues or framing: even small changes in language are thus found to have profound effects on taxpayers' behaviour. While this makes behavioural interventions very easy and cheap to implement for the policy maker, it also calls for extreme cautions with the interaction effects between different types of messages or means of communication. Messages that make deterrence more salient, for instance, might become more or less effective depending on the underlying credibility of enforcement. If the context makes enforcement look less likely, for instance by making the tax authorities look softer and potentially understanding of minor delays with the payment, making it more salient might backfire. If the context instead clarifies that there is no tolerance for delayed payments, messages highlighting enforcement will have a stronger effect. We thus address the issue of how deterrence depends on credibility in two different ways. First, we test the addition to the deterrence message of a visual cue that shows how the authorities usually act upon missed payments - a copy of the executive order that is sent to non-compliant taxpayers when the official procedures are started. If the deterrence message is not found, in and of itself, fully credible, adding this visual cue will increase compliance. Secondly, we test whether the deterrence message has different effects when interacted with the omission-commission message, that specifies that late payments are considered as an active choice that is therefore sanctioned by the law. If the effect of deterrence depends on the credibility of enforcement, the interaction between the two treatments will be positive.

Finally, we contribute to improving the understanding of the effect of omission-commission messages and their interpretation. Traditionally, messages clarifying that omission is considered an active choice by the tax authorities have been found to be effective, and their success was imputed to some form of behavioural bias. More recently, Hallsworth et al. (2015) have argued that the effect of omission-commission messages is rather based on a rational response to expectations of enforcement. If that is the primary channel, however, slight changes in the language would matter. This would call for caution in the framing of such messages, and lead us to expect attenuated effects among messages that do not clarify that the perspective adopted is that of the tax enforcement agency. If, on the other hand, the primary driver were solely behavioural and internal, clarifying that the perspective adopted is the tax authority would have no sizeable effects. Such subtle differences would allow to shed light not only on effectiveness of tax policies, but also on the

understanding of what drives non-compliant taxpayers to comply.

To achieve the outlined goals, we allocated taxpayers from the sample randomly to one of eleven different treatment conditions:

- **Standard behavioural**: a simple reminder to pay taxes, with clear simple language and delivered via ordinary mail

- **Dunning registered**: a formal "dunning" letter reminding the taxpayer to pay their taxes, written in lengthy and bureaucratic language and delivered via <u>registered</u> mail.

- **Dunning regular**: a formal "dunning" letter reminding the taxpayer to pay their taxes, written in lengthy and bureaucratic language and delivered via <u>regular</u> mail

- **Social norms**: like standard behavioural, with a message reporting percentage of compliant taxpayers in the region and inviting taxpayer not to be part of a minority.

- **Public good positive**: like standard behavioural, with a message reminding that paying taxes supports the provision of local public goods and services.

- **Public good negative**: like standard behavioural, with a message reminding that not paying taxes harms the provision of local public goods and services.

- **Deterrence**: like standard behavioural, with a message reminding the possible actions taken against non-compliers.

- **Deterrence + Execution**: like Deterrence, with a copy of the execution order sent to non-compliers.

- **Omission Administration**: like standard behavioural, with a message informing that non-compliance will henceforth be regarded as an active choice.

- **Omission Administration + Deterrence**: like standard behavioural, with the message used in Omission Administration and the message used in Deterrence.

- **Omission Taxpayer**: like standard behavioural, with a message telling the taxpayers that they should henceforth regard non-compliance as an active choice.

Here we follow Hallsworth (2014)'s approach to tax compliance problems, interpreted as

'the unintentional failure of tax- payers to pay their taxes correctly' (Webley et al., 1991).

Compliance includes three main obligations, not all of them applicable to all actors: (i) filing tax

returns on time; (ii) making accurate reports on these returns; (iii) paying any tax owed on time (US Treasury, 2009).

In the context of our intervention, this refers to paying - in full or in part - the amount owed to the tax authorities as a result of personal income tax declarations and not yet paid by the deadline.

The results from the experiment highlight several patterns.

First, simplification improved compliance. Among letters sent via ordinary mail, however, we find that simplified language drove the payment rate up by 6 to 6.2 percentage points from 40.4%. Simplification of the delivery method, instead, did not significantly affect tax payment, but brought significant savings on the delivery cost of registered mail.

Second, behaviourally inspired messages based on public goods and social norms significantly backfired when added to the standard letter. This outcome is on the opposite end of the spectrum of what is found by Hallsworth et al. (2017) in the United Kingdom and stronger than the mixed results found in other countries such as Germany (Dwenger et al., 2017) and Argentina (Castro and Scartascini, 2015). We found that taxpayers receiving a public good message had significantly lower payment rates, by 1.55 percentage points and 3.01 percentage points depending on whether the message framed non-compliance as the reference point (thus framing the effect of compliance on public goods as a gain) or compliance as a reference point (framing the effect of non-compliance as a loss on public goods), respectively. Surprisingly, not only are both effects negative but the latter message performs significantly worse than the previous. The social norms message was similarly detrimental to compliance, with a 3.01 percentage points reduction in the payment rate.

Third, messages that increase the salience of deterrence significantly lower repayment rate by 1.5 percentage points when added to the standard letter, while they increase repayment by 1.3 to 1.5 percentage points if added to the omission-commission message. This result shows the importance of interaction effects between different treatments that, in and of themselves, can improve compliance. It can be argued that, as the omission-commission message clarifies that the tax authority sees non-payment as punishable, the effectiveness of the deterrence message is higher due to the credibility of the threat. That the deterrence message is, instead, counter-effective when added to the standard letter, especially as the standard letter is simpler and lighter in tone than the traditionally used dunning letter, might be due to the fact that the kind tone set in the standard letter does not make deterrence very credible.

That credibility of punishment is crucial to the effectiveness of deterrence messages is suggested by the fact that adding a copy of the execution order sent to non-compliant taxpayers undoes the negative effect that the deterrence message has. This is particularly true as the inclusion of the execution order does not

add any information to what is stated in the deterrence message, but is a display of intention on the tax authority's side.

This paper relates to a broader literature on tax compliance that has its roots in Allingham and Sandmo (1972) and Yitzhaki (1974), who classically framed tax compliance as a rational choice between payment and the expectation of punishment. Numerous departures from this seminal work have been followed to take into account the expenditure side of tax collection (public goods provision, analysed by Cowell and Gordon 1988), tax morale (Dell'Anno 2009), conformity (Myles and Naylor 1996).

Empirical research on the topic, in particular, has been carried to investigate the effectiveness of behavioural messages on income declaration, on the one hand, (see, among many others, Kettle et al. 2017, Slemrod et al. 2001, Blumenthal et al. 2001, Wenzel and Taylor 2004, Hasseldine et al. 2007, Kleven et al. 2011, Ariel 2012, Torgler 2013) and of payment of a due, known amount on the other (Torgler 2004, Fellner et al. 2013, Castro and Scartascini 2015, Del Carpio 2013, Hallsworth et al. 2015 and 2017, Dwenger et al. 2016, Chirico et al. 2017). Most of this experimental evidence is reviewed in Hallsworth (2014), who summarises the main findings.

The rest of the paper is organised as follows: section ?? introduces the background of this policy experiment. Section ?? explains the methodology; section ?? shows the treatments and their rationale. Section ?? reports and discusses the results. Section ?? concludes.

2.1 Background

The intervention came in a context whereby the government of Poland's priority was the efficient mobilisation of domestic revenues. In the area of tax policy, in 2016 the government introduced a new tax on financial institutions and increased the progressiveness of the Personal Income Tax (PIT). In March 2017 it launched a comprehensive reform of the tax administration in order to integrate tax and customs offices. These traditional measures are often politically challenging to design and negotiate and can take time to bring tangible results. As they pursued tax policy and tax administration reforms, the Polish authorities decided to see whether applying insights from behavioural economics to their communications with taxpayers - making small changes to regular processes - could promote tax compliance quickly and at low cost.

In recent years, tax authorities in different countries have begun to experiment with different types of communications (letters, emails, SMSs, websites) using insights from the behavioural science literature to persuade taxpayers to pay what they owe in taxes. International evidence suggests that behaviourally-informed reminders can affect tax behaviour, at least in some contexts. Results from a number of trials have shown that applying behavioural insights to tax communications can help to promote tax compliance and

both raise tax revenues and reduce administrative costs. In these trials, tax authorities have tested a variety of reminder messages highlighting patriotic motives for paying taxes, social norms, possible sanctions, or information from third parties.

The World Bank has also been exploring the applications of behaviourally-informed policies. Its 2015 World Development Report, "Mind, Society, and behaviour," noted that applying behavioural insights had been found to enhance the effectiveness of public policy because people think automatically and socially and often use mental models that are unconscious - that is, they use heuristics and shortcuts that do not always apply in a given context. They also tend to think in terms of stories or narratives rather than data points. These observations can help policymakers better align their communication strategies with the behaviour of citizens. [2]

Behavioural interventions to promote tax compliance have proven to be cost-effective since they generally involve modifications to existing systems and processes. In recent years the impact of behavioural interventions has been measured using randomised controlled trials (RCTs), which are now widely used in medicine, business, and international development. For example, the effectiveness of tax reminders has been tested using RCTs in a number of countries including Australia, Argentina, Austria, Chile, Costa Rica, Denmark, Germany, Guatemala, Israel, Peru, Switzerland, the United States, the United Kingdom, and Venezuela.[3] A common feature of these RCTs is that they are based on data already collected by the tax authorities, and as such they were implemented quickly and at low cost. Hence, recent experience suggests that behavioural interventions can be easily replicated and scaled-up, stimulating a process of adaptive learning.

Recent research has concentrated on using different behavioural messages to incentivise higher tax compliance. Seminal research by the UK behavioural Insights Team (BIT) and the tax authority in the UK had originally shown the efficacy of sending taxpayers letters with behavioural messages. Some of the behavioural messages tested invoked social norms, such as how many people pay on time, or associated taxes with gaining or losing public goods. Relative to the control group, the former increased payment of declared tax liabilities by up to 5.1 pp within 23 days of delivery, and the latter (whether framing loss or gain) increased payments by 1.6 pp (Hallsworth et al. 2017). More recently, the World Bank worked closely with tax authorities in Guatemala to design better communication strategies using these and other behavioural insights. Letters sent to Guatemalans who had failed to declare their income taxes in 2014 showed that, in 11 weeks, the letter highlighting a social norm increased the average amount paid per taxpayer by 13.97 (210 percent), and the deliberate choice letter (omission vs. commission) brought in 17.95 more (269 percent) than not sending

---

[2]The report can be found at: http://www.worldbank.org/en/publication/wdr2015.

[3]See Blumenthal et al. 2001; Wenzel 2005; Torgler 2007; Kleven et al. 2011; Ariel 2012; Castro and Scartascini, 2015; Del Carpio 2014; Ortega and Sanguinetti 2013; Pomeranz 2013; Dwenger et al., 2017; Hallsworth et al., 2017; Kettle et al. 2017, and Brockmeyer et al. 2018.

a reminder letter. Remarkably, compliance by those who received these two letters was also high in the next fiscal year (Kettle et al., 2017). Brockmeyer et al. (2018) document similar findings using email reminders in Costa Rica. Another field experiment in Argentina sent behavioural letters to payers of property tax with messages related either to levels of enforcement (deterrence), social norms, or provision of public goods, and found deterrence to be the most effective. While the latter two had no effects, the deterrence message increased compliance by nearly 5 pp relative to the control group (Castro and Scartascini 2013). Clearly, the effectiveness of behavioural letters can vary. One important reminder from the behavioural science literature is that context matters greatly. Interventions that previously worked in one setting may fail in others. Thus, while evidence from other countries can provide insights to inform the design of behavioural interventions, there was still a need to experiment to see whether behavioural insights can help promote tax compliance in Poland.

In this context, the Polish tax authorities decided to start testing if behavioural insights can strengthen tax collection and promote higher tax compliance in Poland. Due to methodological considerations the authorities decided to focus its experiment on the personal income tax (PIT) and implemented a pilot experiment in 2015 to test the impact of behavioural insights in promoting tax compliance. PIT revenues constitute around 17.3 percent of all tax revenues and correspond to around 2.5 percent of GDP (equivalent to around PLN45 billion). The majority of PIT is paid in the form of monthly advances by the employers (around 85 percent of total PIT, i.e. PLN 39 billion). Towards the end of the year, the employee is obliged to submit a PIT statement to the tax office for the final PIT resettlement. At this stage the outstanding liability is paid or the tax office returns the paid tax due to the use of tax breaks or tax credits. In 2014, the total tax liability to be paid by tax payers in end April (for the fiscal year 2014) was around PLN 6 billion, out of which around 30 percent is not collected by the deadline. The analysis of historical taxpayer compliance rates found that while the tax base has expanded since 2011, more and more taxpayers do not pay their taxes until after the statutory deadline, and the share of those who do not pay at all (10 months after the deadline) has also been heading up. At the same time, the vast majority of taxpayers in arrears have a very small tax liability (Figure **??** and **??**). Such conditions seemed conducive to behavioural interventions in the form of letters to remind taxpayers to pay. In this context, in 2015, with the support of the World Bank and the UK Behavioural Insights Team, the Polish Tax Administration carried out its first RCT trial. The experiment targeted PIT taxpayers in arrears in two regions of the country and tested the impact of two reminder letters on tax compliance outcomes. The results of this pilot trial are available upon request.
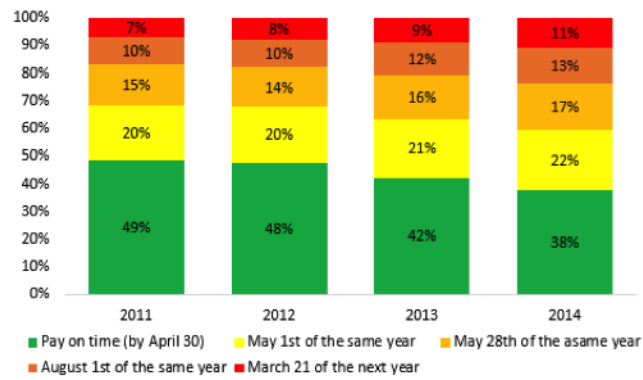
Figure 2.1: Source: Polish Ministry of Finance.
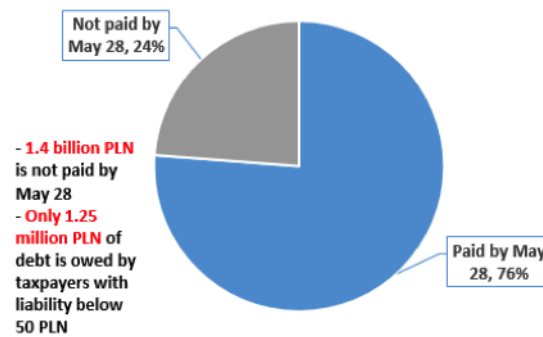
Figure 2.2: Source: Polish Ministry of Finance.

2.1.1 Experimental Design

This RCT trial had two objectives: (1) to increase PIT payments and identify key features of an effective notification strategy; and (2) to test whether the delivery method (registered vs. regular mail) had an impact on tax compliance.

Taxpayers in arrears were defined as those who had filed their PIT declarations on time (by April 30, 2016) but had failed to pay their tax liability since. The trial involved all taxpayers that had non-negligible net tax liability (PLN 50 and above) and had not paid their income tax for 2015 in full by May 13, 2016. These taxpayers were sent a letter by the regional tax office in the days immediately following 23 May 2016. The type of letter was experimentally varied and assignment of each taxpayer to a different letter was randomly selected.

Two types of impact were tested:

- Impact of the content of the letters

- Impact of the delivery method (registered vs. regular mail)

Outcome indicators were payment of any positive amount and amount paid, in the period between 23

Table 2.1: Key Dates for the Experiment

Date	Step
May 13, 2016	Cut-off date for identifying the sample
May 13-22, 2016	Randomization and preparation of letters
May 23, 2016	Roll-out: all letters are sent to taxpayers
June 13, 2016	End of the 1st monitoring round: until this date, there were no other attempts to contact the taxpayers covered by the trial (i.e., the tax authority did not follow up with taxpayers regarding their liability); under regular proceedings, tax offices undertake "soft execution" measures, which implies that they would contact taxpayers in arrears by phone, e-mail, text message, among other forms of communication.
July 4, 2016	End of 3rd monitoring round

May and 13 June 2016. After this date, indeed, the authorities started the ordinary procedures to recover the outstanding credit with the taxpayers. These procedures will have attenuated treatment effects and might potentially have interacted with the pre-existing treatments.

2.1.2 Methodology

The test measured three outcomes using anonymised tax records for the 150,122 participants in the trial, updated as of June 13, 2016 (four weeks after the Tax Authority sent the letters). This sample covered all taxpayers in arrears in Poland with liability above 50 PLN; excluded were taxpayers with no liability, those for whom data were missing for the covariates values, and outliers.[4]

This monitoring date was chosen because until June 13, 2016, the tax administration did not intend to take any other enforcement activities. After that date it is harder to interpret the findings of the experiment because other interventions occurred, such as tax office enforcement activities. For example, a number of taxpayers from all treatment arms that failed to pay by June 13 were sent registered dunning letters. Meanwhile executive proceedings were initiated for taxpayers who had been sent the registered dunning letter when the trial began. These factors could have created confounding effects that might have affected the validity of the estimates. By setting June 13 as the monitoring date, the estimates of the treatment effects are not contaminated by other interventions by the tax authority.

The Polish Tax Office collected outcome variables regularly and automatically as administrative data. The outcomes of interest are these:

- Payment (binary): dummy variable equal to one if the taxpayer paid a non-zero amount of tax by the given date, and zero otherwise. This is our preferred behavioural outcome variable due to its easy interpretability and the comparability with alternative literature.

---

[4]Outliers were identified as observations that, in the OLS regression of payment amount on treatments and all covariates, had a residual of more than 2.5 standard errors.

- Log payment amount (PLN): the (log) amount paid unconditional on payment. This outcome variable is preferred over actual payment as it is scale-free and less dependent on outliers, which highly affect differences. The logarithm was winsorised and equated to log(50) for any payment amount below 50 PLN. The threshold of 50 PLN was chosen as it was the lowest amount at which tax authorities would pursue payment, thus effectively treating it negligibly different from 0.
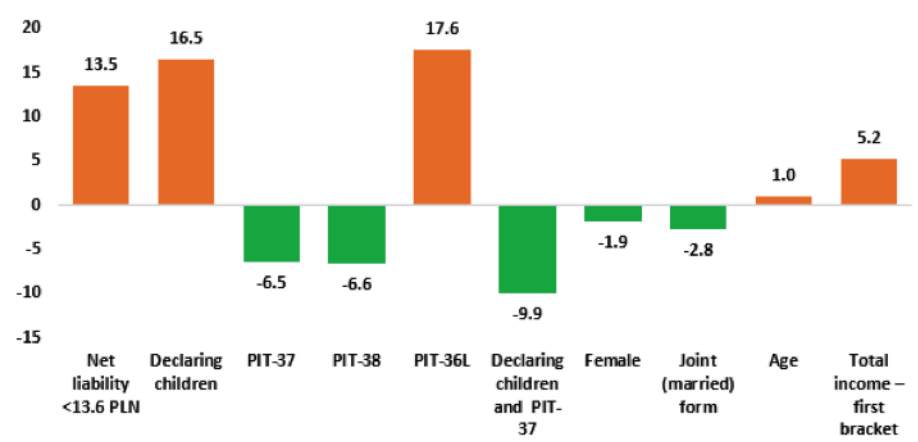
Figure 2.3: Determinants of Noncompliance by Polish Taxpayers, Percent Source: Polish authorities. Note: Analysis refers to 2011 - 14 period. Bars above the axis indicate determinants that lower tax compliance; bars below indicate determinants that improve tax compliance.

2.1.3   Randomisation

This was a randomised controlled trial. Randomisation of treatment group occurred at the individual level and was stratified by those characteristics that had been found to be most important determinant of tax payment in data from the previous four fiscal years 2011-14.

Based on administrative data for 2011-14 it was possible to identify the key determinants of noncompliance by Polish PIT taxpayers. Figure ?? sketches the relevance of some of these characteristics. A number of factors had a statistically important impact on compliance. For example, low tax liability decreased the likelihood of payment by 13.5% and PIT-36L declaration decreased it by 17.6%. However, being female increased payment probability by 1.9% and declaring children increased it by 9.9%.

Stratification prevents imbalance between treatment groups for known factors that influence prognosis or treatment responsiveness. For this experiment, randomisation was conducted at the individual level and stratification was used for initial liability, tax code submitted, electronic submission, gender, parental status, marital status, administrative region (NUTS-1 level), and age group. Given the high dimensionality of the stratification, balance on each of these variables could not be perfect, but the sample was balanced across

the treatment groups, as confirmed in Table **??**.[5]

2.2 Treatments

The aim of this experiment was to check how different letter content affects compliance measures.

The sample covered 150,122 taxpayers in arrears each of whom had a tax liability of more than PLN 50. Taxpayers in this sample were randomly assigned to one of eleven treatment conditions, each of which received a letter from the tax authority.

The registered and regular dunning letters, as well as the standard behavioural letter (our control) are shown in Appendix A. The behavioural messages that are introduced as additions to the standard behavioural letter are reported in Table **??**.

Standard behavioural letter: A substantial literature is evolving on the application of behavioural science to social policy (e.g., Sunstein 2015). Such terms as defaults, status quo bias, peer effects, simplicity, salience, immediate gratification, and reciprocity have entered the dialog of policy-makers. Prominent in the list of applications is tax compliance: see Hallsworth et al. (2017) for an overview of early work in the United Kingdom and elsewhere.

The standard dunning letter was therefore rewritten to conform to many of the best practices from previous research:

- It begins with a very clear description of the purpose of the letter and a "call-to-action" at the top, explaining what the taxpayer must do and giving a deadline: "Please pay your income tax by June 3, 2016."

- It minimises formal legal language to the extent possible, partly so that recipients will understand better but partly just to encourage them to read it.

- It specifies very concrete next steps regarding what to do and how to do it.

- In a separate table, it lists the interest due on each day until the deadline so that taxpayers do not need to calculate it themselves and have additional incentive to pay the liability before the deadline.

The letter now comes across as reader-friendly. The sentences are much shorter, the language is simple, and the messages are very clear. The letter is highly prescriptive in terms of what is being asked and clearly outlines the consequences of compliant and noncompliant behaviour. It also provides contact information

---

[5]Full randomisation was not possible because the covariates were derived from a different data sample (all taxpayers vs. the late taxpayers in the trial sample). Other differences were tax liability (sample included only taxpayers who owed more than 50PLN at the date of the experiment), and outliers were removed.

for the relevant tax authority so taxpayers can notify or seek clarification quickly without needing to look up the phone number themselves (a possible cognitive barrier). It can thus be seen as a plain reminder that addresses nonpayment due to forgetfulness or oversight.

The other letters consist of the behavioural baseline letter augmented with a brief section that includes a behavioural message. In addition to the plain reminders, the behavioural messages are meant to address different motives for nonpayment - economic, moral, and so on. Here, for explanatory clarity, the behavioural messages are provided in a box before the discussion of each.

Registered dunning letter: This is the original letter that the Polish Tax Office had been sending to remind taxpayers to pay their taxes. The dunning letter is normally sent by registered mail. The letter has a formal tone and sets out the legal basis for this type of communication.[6] It states the tax liability and asks the taxpayer to calculate the accrued interest; it provides general guidelines for the calculation but does not provide an example. The letter has a deterrent message, highlighting that "failure to perform [payment] by the indicated deadline shall result in referral of the case to execution proceedings, thereby generating costs of execution proceedings to be covered first."

The letter comes across as very formal, and the language is convoluted, bureaucratic, and legalistic. It might be very difficult for people with average reading skills to understand. Similarly, the explanation of how to calculate the interest rate calculation is difficult to follow.

Regular mail dunning letter: The method of communicating to the taxpayer might matter to compliance in several ways. The traditional letter is sent via registered mail, as that is the necessary legal basis to start proceedings. The registered letter is signed for by the recipient, and if the recipient is absent needs to be collected at the post office. Finally, it charges the recipient for the cost of the special delivery method (11.60PLN). An alternative treatment consisted of sending the same letter via ordinary mail. This intervention could affect compliance in many different ways. From the logistical point of view, regular mail is not tracked and is thus more likely, in principle, to be lost (data are not available, to the best of our knowledge). On the other hand, for taxpayers that are not at home when the letter is delivered via registered mail, the cost of opening the letter (by going to the post office) would be higher, while the cost of paying conditional of having opened the letter might be lower, as the taxpayers would find themselves already at the post office. Finally, the behavioural reaction to the regular mail letter might discourage taxpayers from paying back, as they might perceive that the letter is not authentic, or that the change in the delivery method underlies financial constraints on the government's side and thus less resources to be devoted to enforcement of the law. Finally, from a merely economic point of view, receiving the letter via ordinary mail decreases

---

[6]The Resolution of the Ministry of Finance from May 20, 2014 specifies procedures for creditors of monetary claims in undertaking actions aimed at the application of enforcement measures (Journal of Laws of 2014, item 656).

the amount owed by 11.60 PLN. While this amount is negligible, at the margin it could bring taxpayers to postpone payment, if their principle is to pay when they have enough money to fully repay their liability. All in all, therefore, the effect of simplification of the delivery method is hard to predict, and any size of the effect will not be able to confirm nor disconfirm any of the above mechanisms.

Social Norm letter: The Social Norm letter is based on the observation that people generally have a natural preference to do what their peers are doing. For instance, Gerber and Rogers (2009) found that informing citizens that 71 percent of their compatriots had voted in the previous election increased voter turnout. For taxes, Hallsworth et al. (2017) describe the successful use of social norms to improve compliance in the UK, as Kettle et al. (2016) did for Guatemala, where although the true rate of payment is only 64.5 percent, saying that in the letter increased both rates of payments and average amounts paid. One additional aspect of the letter used here is that it gives the actual rates by region rather than just nationwide; some research findings indicate that the closer the reference or comparison group, the stronger the effects of this type of norm (Goldstein et al. 2008).

Public Good Positive letter: The Public Good Positive letter is based on the assumption that people who knew what expenditures are financed from tax revenues might be more likely to pay their overdue taxes. Therefore, one objective of this letter was simply to inform citizens how some of their taxes are spent and to remind them that many services they use are available only because of tax revenues. This in itself is not necessarily behavioural, but several other elements incorporate behavioural science knowledge. The accurate and unusually precise figure of 37.79% focuses attention on what comes next, unlike a generic claim about taxes providing benefits. It also suggests that the government is tracking everything closely. The link to municipal services brings everything closer to recipient and family. The whole effort stimulates feelings of reciprocity: one should do something in return for all the benefits being received. Finally, the last sentence engages a perceived identity for the recipients as potentially responsible taxpayers rather than individuals who are letting down their communities and neighbours; this pivots the message toward the sense of self, not just the outward action.

Public Good Negative letter: The Public Good Negative letter is based on the same assumption as the previous letter but the behavioural framing is different. The motivation behind this "negative" letter was to harness all the stimulative aspects of the positive public good letter and in addition to frame a loss. It is well-established that decision makers respond more strongly to perceived losses related to a status quo than to perceived gains. In this case the implicit threat is that without sufficient tax revenue, all the municipal benefits currently enjoyed may be lost.

Deterrence letter: The Deterrence letter uses the behavioural letter with the addition of the deterrence message. This message serves three goals. The first sentence is meant to evoke a negative self-conscious emotion of guilt, which has been proven to be a powerful mediator to motivate moral action (Hoffman 1982a, 1982b; H. B. Lewis 1971; M. Lewis 1993). The second sentence is intended to create a sense that the deterrence threat is serious, and the third sentence gives concrete examples of the punishment actions. The examples are likely to evoke mental imagery that enhances realism and therefore behavioural intentions (see, e.g., Miller and Marks 1997; Yoo and Kim 2014). Thus, together the last two sentences are meant to create a sense of fear of possible consequences, which can be an important cause of law-abiding or norm-respecting behaviour (Haidt 2003).

Deterrence + Execution Order: The Deterrence + Execution Order letter reinforces the threat of punishment for noncompliance and makes the punishment more palpable by providing the actual Execution Order Form that non-compliers receive. [7] In addition, the execution form that people would have to fill out makes the time and effort cost associated with filling out the form more obvious, which should further encourage taxpayers to comply with their tax obligations.

Omission Administration: The Omission Administration letter's message is that non-compliance is a deliberate choice. Most real decisions have a status quo alternative: doing nothing or maintaining one's current or previous decision. Numerous studies have found that individuals tend disproportionately to stick with the status quo (Samuelson and Zeckhauser 1988 and Anderson 2003), for two reasons: (1) the losses from acting may loom larger than the gains (Tversky and Kahneman 1981); and (2) moral violations tend to be judged less harshly when the violation results from inaction rather than action (DeScioli et al. 2012). Individuals may use the ambiguity and uncertainty associated with acts of omission in order to minimise future psychological costs arising from the threat to self-image of acting dishonestly (Hallsworth 2013; Mazar et al. 2008).

This message is intended to overcome the status quo bias toward doing nothing - by framing the dishonest behaviour as a deliberate choice - an action in itself. By notifying the reader that failure to comply will be treated as a deliberate choice to be a dishonest taxpayer, this messages aims to eliminate omission as an excuse for noncompliance, thus increasing both the cognitive dissonance around the taxpayer's self-image of an honest person and the perceived cost of paying later. The wording also gives taxpayers an exemption for not previously declaring, which introduces an element of reciprocity, because the implication is that the

---

[7]The Execution Order is defined in the Resolution of the Ministry of Finance from May 16, 2014, (J. L. of 2014, item 650).

taxpayers have been granted a favour or shown good will. The text is also worded to give the impression that taxpayer behaviour is being closely monitored, which may heighten the perceived threat of subsequent actions against noncompliance. Thus it acts as a mild deterrent. A similar approach was shown to be effective in Guatemala (Kettle et al. 2016).

Omission Administration + Deterrence: The Omission Administration + Deterrence letter is a variant of the behavioural letter. Both omission and deterrence interventions were included to test whether there are additive effects if both are used together. That is, would combining the two interventions be more effective than using each separately?

Omission Taxpayer: The Omission Taxpayer letter is a slight but important variation to the previous letter. In particular, the perspective is changed from the tax authority to the taxpayer. Recent research on moral judgment and decision making suggests that in addition to the deterrence threat (whether that is mild or strong), individuals care about their moral self-image - they want to retain a positive view of themselves, and a threat to their moral self-image can be a significant motivator of honest behaviour even beyond the deterrence threat (Mazar et al. 2008). This intervention was meant to test the strength of this in the context of tax payments (see also Shu et al., 2012).

Letter	Contents
Social Norm	According to our records, [8] out of 10 residents in [REGION OF THE TAXPAYER] have already paid their income tax for 2015. You are part of a minority that has not yet fulfilled that duty.
Public Good Positive	Are you aware that 37.79% of your personal income tax goes to your municipality? From this income, your municipality finances preschools, schools, roads, and safety, benefiting everyone in your municipality including yourself and your family. Don't be an irresponsible inhabitant of your municipality and pay your delinquent taxes!
Public Good Negative	Are you aware that 37.79% of your personal income tax goes to your municipality? Without this income, your municipality cannot finance preschools, schools, roads, and safety, damaging everyone in your municipality including yourself and your family. Don't be an irresponsible inhabitant of your municipality and pay your delinquent taxes!
Deterrence	Not paying taxes places an unfair burden on all other taxpayers, who have honestly fulfilled their duty. We are therefore determined, more than ever, to collect taxes from those, who avoid paying them. As part of the execution procedures, we can, for example, block your bank account or salary, and, in addition, you will have to cover all execution expenses that arise.
Deterrence + Execution Order	[*Same as deterrence message above plus:*] We attach a sample Execution Order Form which we send to taxpayers that have not paid their taxes due.
Omission	So far, we have thought of your payment delay to be accidental. However, if you disregard this notice, we will consider it an intentional choice of yours and think of you as a dishonest taxpayer.
Omission + Deterrence	[*Same as omission message above plus:*] As part of the execution procedures, we can, for example, block your bank account, salary, and, in addition, you will have to cover all execution expenses that arise.
Omission Taxpayer-Perspective	So far, you might have thought of your payment delay to be accidental. However, if you disregard this notice, you should consider it an intentional choice of yours and think of yourself as a dishonest taxpayer.

Table 2.2:   Contents of the message in each behavioural letter. Underlines are added by the authors to highlight differences relative to comparable messages. Italics are added to report content that is common across different messages. In the letter, the content of the behavioural message is bold.

	Dunning registered	Dunning regular	Behavioural	Social norms	Public good positive	Public good negative	Deterrence	Deterrence + Execution	Omission Admin	Omission Admin + Deterrence	Omission Taxpayer	F p-val
Age	43.57 [13.56]	43.75 [13.54]	43.83 [13.62]	43.61 [13.57]	43.85 [13.62]	43.56 * [13.51]	43.73 [13.63]	43.71 [13.59]	43.83 [13.57]	43.66 [13.53]	43.85 [13.54]	.51
Female	.3529 [.4779]	.36 [.48]	.3507 [.4772]	.3535 [.4781]	.358 [.4794]	.3526 [.4778]	.3522 [.4777]	.3586 [.4796]	.3592 [.4798]	.3581 [.4794]	.3541 [.4783]	.79
Married	.1464 [.3535]	.1518 [.3589]	.1489 [.356]	.1498 [.3569]	.1529 [.3599]	.1516 [.3586]	.1522 [.3592]	.1493 [.3564]	.1492 [.3563]	.15 [.3571]	.1494 [.3564]	.98
Reports children	.1061 [.3081]	.0966 * [.2954]	.1043 [.3057]	.1014 [.3019]	.1001 [.3001]	.1034 [.3045]	.105 [.3065]	.1024 [.3032]	.1005 [.3006]	.1011 [.3015]	.1021 [.3027]	.73
Liability (PLN)	8138 [58901]	7255 [27779]	7671 [43922]	7420 [26599]	8083 [65646]	8351 [85971]	7269 [24071]	7593 [31401]	7680 [36730]	7538 [26417]	9214 [150196]	.36
PIT-36	.3962 [.4892]	.393 [.4884]	.4004 [.49]	.3974 [.4894]	.3965 [.4892]	.3993 [.4898]	.3962 [.4891]	.3985 [.4896]	.3994 [.4898]	.3989 [.4897]	.3966 [.4892]	1
PIT-36L	.2242 [.4171]	.2233 [.4165]	.2225 [.4159]	.2239 [.4168]	.2206 [.4146]	.2229 [.4162]	.2244 [.4172]	.2216 [.4154]	.222 [.4156]	.2207 [.4148]	.2231 [.4163]	1
PIT-37	.3402 [.4738]	.343 [.4747]	.3353 [.4721]	.3368 [.4726]	.341 [.4741]	.3371 [.4727]	.3362 [.4724]	.3392 [.4735]	.3364 [.4725]	.3395 [.4735]	.3394 [.4735]	.99
PIT-38	.0352 [.1842]	.0333 [.1794]	.0373 [.1896]	.0373 [.1896]	.0387 [.1928]	.0367 [.1881]	.0384 [.1922]	.0355 [.185]	.0377 [.1905]	.0364 [.1873]	.0353 [.1844]	.66
PIT-39	.0163 [.1267]	.0162 [.1264]	.0161 [.1258]	.0152 [.1224]	.0143 [.1185]	.0143 [.1186]	.0142 [.1183]	.0151 [.1218]	.0153 [.1229]	.0158 [.1249]	.0162 [.1261]	.82
N.	5888	6100	15244	15510	15434	15360	15435	15327	15267	15271	15286	

Table 2.3: Summary statistics. Mean [standard deviation] of main explanatory variables. Stars indicate significant differences from the Behavioural control group. * p<.10, ** p<.05, *** p<.01. F p-val column reports the p-value of the null hypothesis that all treatment groups have equal mean to the Behavioural control group. Sample: 150,122 late taxpayers with positive liability as of 23 May 2016 and who have been sent a letter by the Polish tax authorities.

Married and "Reports children" are self-reported. PIT-XX are dummies equal to 1 if the taxpayer has submitted a PIT-XX tax return. Of 150,122 taxpayers, 1,575 (1.05%) have submitted more than one PIT- document, of which 1,561 have submitted two and 14 have submitted three. PIT-37 refers to income received from a Polish payer, e.g. employer or client. PIT-36 and PIT-36L refer to income arising from non-Polish payers, with PIT-36L having a linear tax rate and PIT-36 a progressive one. PIT-38 refers to financial transactions and PIT-39 to proceeds generated from the sale of property.

2.3 Results

Aggregate differences in payment rate and amount, by treatment, are reported in Table **??** and Figure **??**. Tables **??** to **??** report estimates from hurdle and OLS models, with and without controls.

The estimated hurdle model is

$$P_i^* = X_i\delta + \eta_i \quad,$$

$$P_i = 1\left(P_i^* \geq 0\right) \quad,$$

$$Y_i^* = T_i\beta + X_i\gamma + u_i \quad,$$

$$Y_i = \max\left\{\log(50); Y_i^*\right\} \quad \text{if } P_i = 1 \quad.$$

where P_i is a dummy for payment, Y_i is the observed log payment amount, T_i is a treatment vector, X_i is a vector of control variables comprising the characteristics of taxpayers (amount of tax liability, declaring to be have children, type of tax form, gender, age (squared), liability decile), u_i and η_i are independent error terms, with $\eta_i|X_i, T_i, u_i$ distributed according to a standard logistic c.d.f., $u_i|X_i, T_i, \eta_i$ distributed according to a normal distribution with mean 0 and variance σ_i^2. The identifying assumption is that the two error terms are independent of treatments and covariates and conditionally uncorrelated.

The estimated OLS model is instead

$$Pr(P_i = 1) = X_i\delta + \eta_i \quad,$$

$$Y_i = T_i\beta + X_i\gamma + u_i \quad \text{if } P_i = 1,$$

where η_i and u_i are independently distributed error terms, with $E(\eta_i|X_i, T_i) = E(u_i|X_i, T_i, \eta_i) = 0$.

Results from the hurdle models, without (model 1) and with controls (model 2), as well as OLS models, without (model 3) and with controls (model 4), are reported in Tables **??** to **??**.

2.3.1 Payment rate

The first and main outcome of interest is the payment rate, which is the standard compliance measure in similar studies in the literature. A taxpayer is reported as paying if they pay any amount in the time period considered.

In the traditional treatment group (dunning registered) payment rate in the given period is as low as

40.90%, compared to 46.51% repayment rate among the standard behavioural treatment (Table **??** and Figure **??**). In the different versions of the behavioural letter repayment rate ranges from 43.29% (negative public goods) to 48.96% (omission admin with deterrence). Simplification can thus play a large role in increasing repayment rates, but as much of a difference can be made by choosing the most appropriate message in a given context.

When using more precise estimates and controlling for other covariates (Table **??**, column 2)[8] , we find the following results:

1. Language simplification significantly improves payment by as much as 6.84 pp;

2. Including a public good moral appeal or a social norms message significantly decreases payment;

3. Deterrence significantly decreases payment unless it is paired with a cue that makes it more salient or more likely;

4. Messages based on omission significantly increase repayment only when they are expressed as representative of the tax authority's point of view as opposed to when they are presented as moral suasion messages;

5. The most effective treatment is the omission message, expressed from the administration's point of view, paired with a deterrence message.

Overall, the treatment effects of behavioural variations over the standard letter range from -3.42 to +2.70 pp.

Table **??** further compares treatment effects from a logit regression, with and without controls.

We find the following:

1. messages appealing to public goods within a positive (gain) framework work significantly better than messages representing public goods with a negative (loss) framework. This runs contrary to the prior expectation that taxpayers would be encouraged to pay when compliance is presented as averting a loss relatively to when compliance is presented as creating a gain. One potential explanation is that the type of framework is not neutral to the likelihood of the behaviour: when forced to think of non-compliance as causing the end of provision of public goods, a taxpayer might be drawn into thinking about how likely their own behaviour is to cause such a big effect. Such thinking might be stronger if there is self-serving bias. If the taxpayer is a repeat offender, furthermore, as is often the case with

---

[8]We take model 2 as our preferred one, as it jointly makes use of information arising from the extensive as well as the intensive margins, and it adds control that improve the estimates' precision. Estimated treatment effects are consistent across the four models reported in the Table.

late taxpayers, the status quo that they have always perceived is the one whereby they have not paid taxes and the level of public goods was given. A message framing the status quo as compliance will thus have no byte on such recidive taxpayers.

Another potential explanation is that for taxpayers public services decrease utility and they perceive their private money as better spent on private goods rather than public goods or publicly provided services. In such case the positive public good framework would actually be a loss scenario, where taxpayers would be brought to consider non-compliance as the behaviour that averts the provision of public goods (a loss).

2. The delivery method does not significantly affect compliance. This non-significance result could mask a variety of effects going in opposite directions. As speculated above, ordinary mail might make a taxpayer doubt of the authenticity of the letter and thus reduce repayment. Similarly, it might be less likely to be actually delivered, depending on the quality of the ordinary mail delivery system. On the other hand, it might be more likely to be opened, as taxpayers who are absent do not have to go to the post office and collect it, and it does not charge the taxpayer the 11.60 PLN delivery costs that taxpayers are asked to repay when receiving the registered mail dunning letter.

3. The inclusion of the executive order makes deterrence more effective. This could be interpreted as evidence of the degree of salience of the deterrence, which is definitely higher when a copy of the execution order is received. It could also be interpreted, however, as evidence that appeals to deterrence backfire when there is no evidence of the punishment that a non-compliant taxpayer could incur. The execution order might then have made the deterrence letter more credible. To test which channel was affected, additional survey evidence would be needed.

4. Although omission messages that are framed from the taxpayer's point of view are not significantly better than the control, they are also not significantly worse than omission messages framed from the tax authority's point of view. This leaves us unable to exclude that the omission messages work regardless of how threatening they look. The theory, advanced by Hallsworth et al. (2015), that omission messages work because of their threatening power and the perceived higher likelihood of punishment, cannot be confirmed from a comparison of the two different omission treatments. To give a final answer, however, more data on how threatening the two treatments are perceived by the taxpayers in our experiment would be needed: it is in fact possible that, despite the design of the Omission Taxpayer message was aimed at creating a sense of moral suasion, as opposed to deterrence, this was not actually picked up by the recipients. A survey would be crucial to assess this question.

5. Although the deterrence message has a negative effect on compliance if added to the standard behavioural letter, the opposite happens if it is added to the Omission Administration letter. As the last row of the table shows, there is a strong positive interaction effect between Deterrence and Omission Administration. This shows how the same message can have different effects depending on the context where it is used. The specific reason why the interaction with Omission Admin is positive is hard to pin down without additional survey evidence. If the deterrence message interacts with the Omission Admin message, however, there should be an overlap between the channels affected by the two messages. It seems natural to believe that the deterrence message's effectiveness depends on how credible the message is perceived to be, how likely deterrence is perceived to be implemented, and how salient the punishment scenario is. The complementarity with the Omission Admin letter can then be explained by the effect that the latter has on one or more of the three channels described above. One explanation is given by Hallsworth et al. (2015), whose survey results suggest that an omission-based message raises the perceived probability of punishment.

2.3.2 Payment amount

Although payment amount is shown for completeness and for policy purposes, from an economic point of view that is harder to interpret and would require separate analysis. While the treatment effect should ideally represent the effect of treatments on the desired payment on the intensive margin, the identifying assumption to estimate this effect is very strong and hard to defend. In comparing taxpayers that have been encouraged (or not discouraged) to pay by different treatments there is the potential for selection bias. Using observed data on payers as informative of the underlying behavioural mechanisms could be misleading at best. Selection should be dealt with with appropriate experimental designs or with econometric techniques, that are however not the purpose of this paper.

Table ??, however, helps uncover some selection mechanisms. If we were willing to assume exogeneity of treatment conditional on liability decile, age, gender, parental self-reported status, we could interpret results from columns (1) and (3) as driven by selection and results (2) and (4) as the estimates of the underlying model once selection is netted out. This reveals, for instance, that the lower amount raised among compliant taxpayers from the dunning letters is fully explained by other covariates. This would mean that the dunning letters discourage the payment among those taxpayers that would otherwise pay more. Alternatively, the behavioural letter increases payment among taxpayers by encouraging compliance among taxpayers that, due to their observable characteristics, are expected to pay less. The positive public good message, pretty much like the dunning letters, discourages payment among those whose gender, age and other observable

characteristics are associated with higher payment, either because they discourage taxpayers with higher liability or because they discourage taxpayers wit better tax morale. Once controls are taken into account, we find significant and positive increases in the payment amount among taxpayers in the Social Norms and in the Deterrence treatment groups. These results somewhat counteract the negative effects that these treatments have on the payment rate, begging the question - from a policymaker's perspective - of whether, unconditionally of having paid, revenues per taxpayer are higher or lower after this treatment.

The treatment effect on payment rate seems to be the dominant effect, as suggested by data on unconditional (log) payment per late taxpayer shown in Figure **??** and Table **??**. This provides one more practical reason to focus on the payment rate as an outcome, as opposed to the payment amount. The hurdle model estimated in column 2 of Table **??** displays the following effects (evaluated at the fifth income decile and at the average of other covariates): a 20.26% decrease in government revenues per taxpayer when choosing the complicated language over the simplified language displayed in the standard behavioural letter, and an 8.11% increase in revenues over the standard behavioural letter following the addition of the Omission Administration + Deterrence message. These effects amount to about 11% and 4.5% of a standard deviation of log payment, respectively.
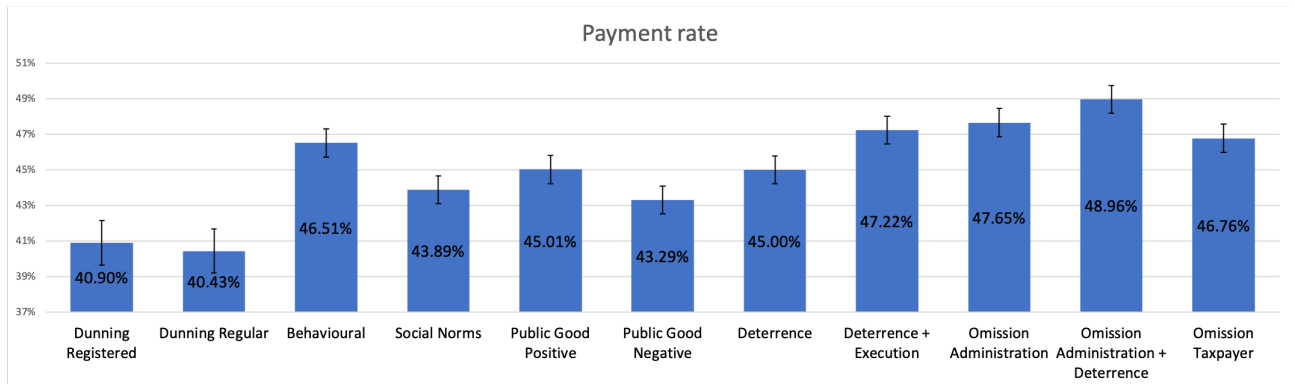
(a) Average payment rate by treatment group. Error bars indicate 95% confidence intervals.
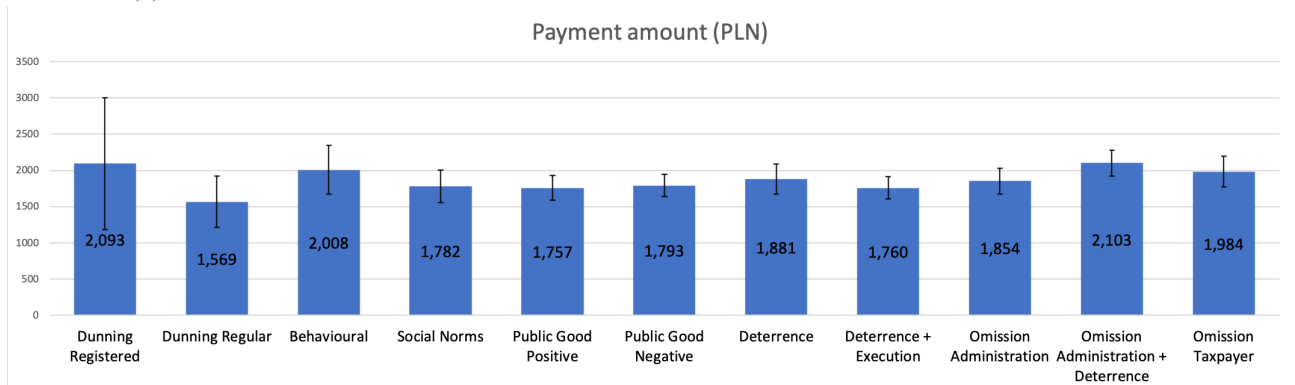
(b) Average payment (PLN) by treatment group. Error bars indicate 95% confidence intervals.
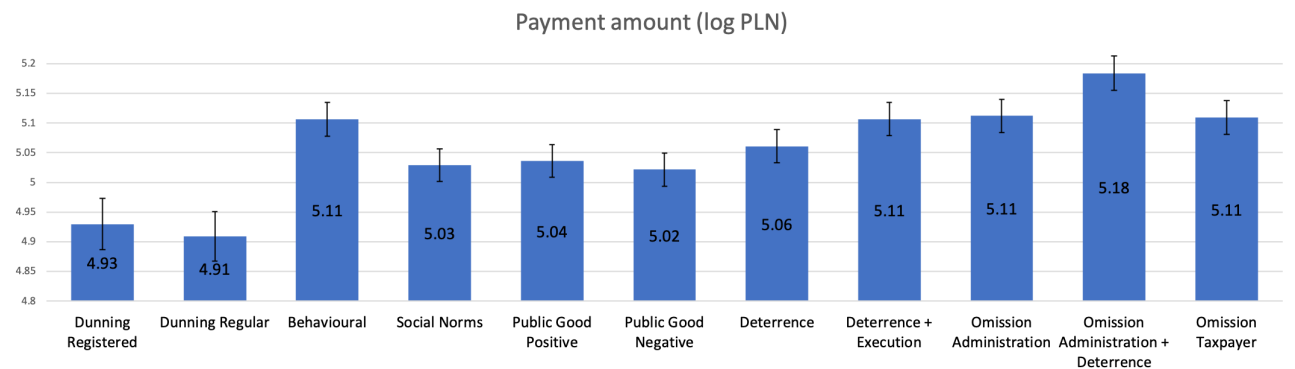
(c) Average payment (log) by treatment group. Error bars indicate 95% confidence intervals.

Figure 2.4: Summary statistics from the sample of 150,22 letters recipients.

	Dunning registered	Dunning regular	Behavioural	Social norms	Public good positive	Public good negative	Deterrence	Deterrence + Execution	Omission Admin	Omission Admin + Deterrence	Omission Taxpayer	F p-val
Paid (0-1)	.4090 *** [.4917]	.4043 *** [.4908]	.4651 [.4988]	.4389 *** [.4963]	.4501 *** [.4975]	.4329 *** [.4955]	.45 *** [.4975]	.4722 [.4992]	.4765 ** [.4995]	.4896 *** [.4999]	.4676 [.499]	0 ***
Payment (PLN)	2093 [35460]	1569 [14111]	2008 [21332]	1782 [14398]	1757 [10775]	1793 [9740]	1881 [13217]	1760 [9575]	1854 [11319]	2103 [11192]	1984 [13249]	.24
Payment (log PLN)	4.9297 *** [1.6986]	4.9089 *** [1.6749]	5.1064 [1.7866]	5.029 *** [1.7534]	5.036 *** [1.7446]	5.0216 *** [1.757]	5.0606 ** [1.7696]	5.1064 [1.7708]	5.1121 [1.7767]	5.1837 *** [1.8303]	5.1093 [1.7887]	0 ***
Outstanding liability (PLN)	6010 [28737]	5650 [23696]	5643 [37947]	5612 [22226]	6304 [64700]	6532 [85378]	5361 [20021]	5811 [29605]	5801 [34595]	5407 [23402]	7206 [149577]	.33
N.	5888	6100	15244	15510	15434	15360	15435	15327	15267	15271	15286	

Table 2.4: Summary statistics. Mean [standard deviation] of compliance variables as of 13 June 2016. Stars indicate significant differences from the Behavioural control group. * p<.10, ** p<.05, *** p<.01. F p-val column reports the p-value of the null hypothesis that all treatment groups have equal mean to the Behavioural control group. Sample: 150,122 late taxpayers with positive liability as of 23 May 2016 and who have been sent a letter by the Polish tax authorities.

	Pr(paid)			
	(1) Hurdle1	(2) Hurdle2	(3) OLS1	(4) OLS2
Dunning registered	-.0568*** (.0078)	-.0598*** (.008)	-.0561*** (.0076)	-.0551*** (.0073)
Dunning regular	-.0617*** (.0077)	-.0684*** (.008)	-.0608*** (.0075)	-.0629*** (.0073)
Social norms	-.0263*** (.0057)	-.0283*** (.0059)	-.0262*** (.0057)	-.0263*** (.0055)
Public good positive	-.0150*** (.0057)	-.0176*** (.0059)	-.0150*** (.0057)	-.0163*** (.0055)
Public good negative	-.0324*** (.0057)	-.0342*** (.006)	-.0322*** (.0057)	-.0317*** (.0055)
Deterrence	-.0152*** (.0057)	-.0167*** (.0059)	-.0151*** (.0057)	-.0156*** (.0055)
Deterrence + Execution	.0071 (.0057)	.0058 (.0059)	.0071 (.0057)	.0054 (.0055)
Omission Administration	.0113** (.0057)	.0104* (.0059)	.0114** (.0057)	.0097* (.0055)
Omission Administration + Deterrence	.0245*** (.0057)	.0270*** (.0059)	.0245*** (.0057)	.0252*** (.0055)
Omission Taxpayer	.0025 (.0057)	.003 (.0059)	.0025 (.0057)	.0027 (.0055)
Controls	No	Yes	No	Yes
N.	150,122	150,122	150,122	150,122

Table 2.5: Treatment effects on the probability of paying some taxes. Controls consist of age (squared), gender, a dummy for decile of tax liability. Columns (1) and (2) report the marginal effects of hurdle models evaluated at the control "behavioural letter" condition and, limited to column (2), at mean age (and its squared) and fifth liability decile.
Robust standard errors in parentheses. * p<.10, ** p<.05, *** p<.01.

	E(ln payment | paid)			
	(1) Hurdle1	(2) Hurdle2	(3) OLS1	(4) OLS2
Dunning registered	-.0745* (.0404)	-.0061 (.0161)	-.0797* (.0434)	-.0067 (.0178)
Dunning regular	-.0952** (.0396)	-.0213 (.0159)	-.102** (.0426)	-.0236 (.0176)
Social norms	-.0214 (.0291)	.0186* (.0106)	-.0228 (.0311)	.0206* (.0117)
Public good positive	-.0662** (.0288)	.0023 (.0108)	-.0708** (.0308)	.00259 (.0119)
Public good negative	-.0045 (.0294)	.0095 (.011)	-.00483 (.0313)	.0106 (.0121)
Deterrence	-.0143 (.029)	.0235** (.0106)	-.0153 (.0309)	.0261** (.0117)
Deterrence + Execution	-.0363 (.0284)	.0071 (.0106)	-.0388 (.0304)	.0080 (.0117)
Omission Administration	-.0461 (.0284)	.0017 (.0107)	-.0492 (.0304)	.0022 (.0118)
Omission Administration + Deterrence	.0276 (.0286)	.0109 (.0107)	.0293 (.0304)	.0121 (.0118)
Omission Taxpayer	-.0071 (.0288)	.0186* (.0106)	-.0075 (.0306)	.0205* (.0117)
Controls	No	Yes	No	Yes

Table 2.6: Treatment effects on the amount paid, conditional on the subject paying. Log payment is winsorised at log(50) for any amount lower than 50 PLN. Controls consist of age (squared), gender, a dummy for decile of tax liability. Columns (1) and (2) report the marginal effects of hurdle models evaluated at the control "behavioural letter" condition and, limited to column (2), at mean age (and its squared) and fifth liability decile.
Robust standard errors in parentheses. * p<.10, ** p<.05, *** p<.01.

	E(ln payment)			
	(1) Hurdle1	(2) Hurdle2	(3) OLS1	(4) OLS2
Dunning registered	-.1832*** (.0282)	-.1711*** (.0239)	-.177*** (.0264)	-.171*** (.0256)
Dunning regular	-.2055*** (.0279)	-.2026*** (.0238)	-.197*** (.0259)	-.187*** (.0250)
Social norms	-.0788*** (.0205)	-.0706*** (.0175)	-.0774*** (.0202)	-.0744*** (.0194)
Public good positive	-.0701*** (.0204)	-.0483*** (.0175)	-.0704*** (.0202)	-.0648*** (.0194)
Public good negative	-.0868*** (.0206)	-.0915*** (.0176)	-.0848*** (.0203)	-.0842*** (.0194)
Deterrence	-.0464** (.0203)	-.0358** (.0174)	-.0458** (.0203)	-.0413** (.0195)
Deterrence + Execution	.0017 (.0199)	.0197 (.0174)	.0000 (.0203)	.0071 (.0195)
Omission Administration	.0082 (.0199)	.0299* (.0174)	.0057 (.0204)	.0109 (.0195)
Omission Administration + Deterrence	.0768*** (.0197)	.0811*** (.0173)	.0773*** (.0207)	.0755*** (.0198)
Omission Taxpayer	.0033 (.0200)	.0172 (.0173)	.0029 (.0205)	.0025 (.0196)
Controls	No	Yes	No	Yes

Table 2.7: Treatment effects on the average amount paid (in log), unconditional of whether the subject pays or not. Log payment is winsorised at log(50) for any amount lower than 50 PLN. Controls consist of age (squared), gender, a dummy for decile of tax liability. Columns (1) and (2) report the marginal effects of hurdle models evaluated at the control "behavioural letter" condition and, limited to column (2), at mean age (and its squared) and fifth liability decile.
Robust standard errors in parentheses. * p<.10, ** p<.05, *** p<.01.

	(1) Logit, no controls	(2) Logit, controls
Public good negative	-.0173***	-.0165***
- public good positive	(.0057)	(.0059)
Registered dunning	.0047	.0084
- regular dunning	(.0089)	(.0093)
Deterrence with exec order	.0222***	.0225***
- Deterrence only	(.0056)	(.0059)
Omission taxpayer	-.0088	-.0074
- omission administration	(.0057)	(.0059)
Omission administration cum deterrence	.0132**	.0167***
- omission administration	(.0057)	(.0059)
Deterrence	-.0152***	-.0167***
- standard behavioural	(.0057)	(.0059)
Deterrence	.0283***	.0334***
X omission admin	(.0081)	(.0084)

Table 2.8: Comparison of marginal effects in a logit regression of payment of any taxes on treatments. Robust standard errors in parentheses, evaluated through the delta methods. Results from column (2) are obtained controlling for age (squared), gender, and tax liabilitiy decile are included. Marginal effects are reported at mean age (and its squared) and fifth liability decile. * $p<.10$, ** $p<.05$, *** $p<.01$.

2.4 Conclusions

This policy experiment confirms the importance of simplification on improving compliance. We separately studied the effects of simplification in message contents and simplification in the delivery method. While simple message content improves compliance significantly, simplification in the delivery method did not affect taxpayers' behaviour significantly. This non-significance result, however, amounts to a net social benefit, as simplification of the delivery method, if applied to the whole sample of late taxpayers as is customary, would have resulted in savings of PLN 1,741,415.

Among behaviourally informed, simplified, messages, we found that details matter highly and there can be very important differences. Differently to what has been found in similar policy experiments on late taxpayers, messages based on moral suasion and encouraging cooperation with the tax authority by highlighting the relevance of public goods or social norms have significant negative effects. This result shows the importance of catering policy interventions to the context. In this context, explanations for the different result can be due to the perceived quality of public goods, different attitudes to the tax administration, and different cultural values than in the contexts where similar experiments have been conducted. The experience of Poland as a post-socialist society probably needs to be taken into consideration, as well as the existence of pre-existing standard practices on the government's side based on very formal communication style. Against this background, letters based on moral suasion might fail to be interpreted as credible or might signal a weakening of the government's ability to enforce policies and lower resources.

Further research should aim at addressing the effects of moral suasion messages on the credibility of the tax authority and on the perceived probability of detection and sanction.

Omission-commission messages are found to increase compliance by telling the taxpayer that, from then on, the tax authority considers non-payment as an active choice not to comply with the law. Such messages are found (Hallsworth et al. 2017) to increase the perceived likelihood of punishment. While this has been found as evidence that the response to omission-commission messages is not driven by behavioural biases but by strategic responses, in our experiment we introduce a variation that is designed to provide the same message as a case of moral suasion and not of enforcement. We find no significant difference between the two treatments, suggesting that the effect of omission-commission messages on compliance could be a combination of increased probability of punishment as well as a framing effect.

A simple deterrence message alone, interestingly, decreased the payment rate when added to the baseline behavioural letter, but increased it when associated with the omission-commission one. This difference reveals the importance of substantiating deterrence messages with credible threats. In the lack of credibility, threats might even encourage the opposite response. Similarly, the negative effect of deterrence messages

alone was undone by the addition of a sample execution form, that made the perceived threat arguably more realistic and likely to arise.

That the sample is made up of late taxpayers needs particular consideration: this sample is extremely self-selected. Late taxpayers might represent the less pro-socially minded part of society, which would then be less responsive to messages based on moral suasion and more reactive to deterrence.

Secondly, for some taxpayers this intervention was a stark change of approach on the tax authorities' side. It is possible that some of the effects we analyse here are simply due to the novelty of the intervention. To understand the effects of a change in the communication style on taxpayers' behaviour, we would want to study the effect of repeat intervention over the years. To the best of our knowledge there is not research yet on this kind of long-term effects.

In the context of the research on tax compliance, our results confirm the general result that deterrence messages outperform interventions based on moral suasion and that different geographical contexts lead to very different responses from taxpayers. We also confirm results on the importance of simplicity of language. We add a dimension that can be simplified upon, that is delivery method, and do not find any net results. Different contexts and different methods of delivery could be considered, as - despite the less than phenomenal returns that coule be expected, based on the range of eff

We find that different messages can interact in interesting and possibly unexpected ways, and we highlight the importance of credibility of threats and how that needs to be taken into account when designing an experiment.

Although our results seem to confirm the idea that taxpayers respond to incentives more than to behavioural cues, we cannot rule out that some framing effects persist, in particular in the way taxpayers respond to omission-commission types of messages.

Clearly, cooperation with the tax authorities and the government does not seem to be the main driver of the payment decision within this self-selected sample of late taxpayers.

2.A   Control Letters and Experiment Letter Variants

This section displays the samples of the letters that were sent out to the taxpayers in the experiment.

REGISTERED Dunning Letter – ENGLISH translation

HEAD OF TAX OFFICE IN ŻARY
ŻARY, OSADNIKÓW WOJSKOWYCH 3
Phone number: 6844560500
………………………………………………………

NIP:………..

Receipt confirmation

DUNNING LETTER
NO. 466/15
OF 25.03.2015.

Pursuant to art. 15 § 1 of the Act of June 17, 1966 on execution proceedings in administration (J. L. of 2014, item 1619 – uniform wording) this is to call for payment of, as follows:

No.	Liability	Term	Amount of liability in PLN	*Interest as on the date of issue hereof in PLN	Total
1.	PIT (ZOB-D)	1/2014	30.00	3.00	33.00
2.	PIT (ZOB-D)	2/2014	123.00	11.00	134.00
3.	Dunning letter costs	X	X	X	11.60
4.	Total	X	X	X	178.60

Type of interest – tax interest
Rate applicable to further interest accrual – 8%

This is to call for performance of duty referred to herein within 7 days from delivery of this dunning letter. The amount due with interest accrued until the day of payment and costs of dunning letter should be paid to the cash desk, through the postal service of Poczta Polska or to the bank account NBP/O/OKR in Zielona Góra no. 12101017040055262223000000.

Failure to perform the said duty by indicated deadline shall result in referral of the case to execution proceedings thereby generating costs of execution proceedings to be covered first.

*Interest has been accrued as on the date hereof. When making payment please add interest accrued against the amount due from the day immediately following the date hereof until the payment date.

When making the payment in full amount of liability together with interest and dunning letter costs, round up the amount to full zloty. Interest shall not be collected if the amount of interest, as on the payment date, is lower or equal to 8.70 PLN.

(signature and personal stamp bearing name,
surname and official position)

REGULAR MAIL Dunning Letter – ENGLISH translation

XXXXXX, 05.23.16 r.

HEAD OF TAX OFFICE IN ŻARY
ŻARY, OSADNIKÓW WOJSKOWYCH 3
Phone number: 6844560500
……………………………………………………….

Jan Nowak
ul. Białobrzeska 40 m. 14
62-544 Warszawa

Pursuant to art. 15 § 1 of the Act of June 17, 1966 on execution proceedings in administration (J. L. of 2014, item 1619 – uniform wording) this is to call for payment of, as follows:

No.	Liability	Term	Amount of liability in PLN	*Interest as on the date of issue hereof in PLN	Total
1.	PIT (ZOB-D)	1/2014	30.00	3.00	33.00
2.	PIT (ZOB-D)	2/2014	123.00	11.00	134.00
4.	Total	X	X	X	167.00

Type of interest – tax interest
Rate applicable to further interest accrual – 8%

This is to call for performance of duty referred to herein within 7 days from delivery of this letter. The amount due with interest accrued until the day of payment and costs of this letter should be paid to the cash desk, through the postal service of Poczta Polska or to the bank account NBP/O/OKR in Zielona Góra no. 12101017040055262223000000.

Failure to perform the said duty by indicated deadline may result in referral of the case to execution proceedings thereby generating costs of execution proceedings to be covered first.

*Interest has been accrued as on the date hereof. When making payment please add interest accrued against the amount due from the day immediately following the date hereof until the payment date.

When making the payment in full amount of liability together with interest, round up the amount to full zloty. Interest shall not be collected if the amount of interest, as on the payment date, is lower or equal to 8.70 PLN.

(signature and personal stamp bearing name,
surname and official position)

BEHAVIORAL BASELINE LETTER AND VARIATIONS – ENGLISH translation

XXXXXXXX, May 23, 2016

Tax
Administration

Jan Kowalski
Naczelnik Urzędu Skarbowego
Urząd Skarbowy XXXXXXXXXX
ul. XXXXXXXXXX

Mr.
Jan Nowak
ul. Białobrzeska 40 m. 14
62-544 Warszawa

Please pay your income tax due by June 3, 2016

Dear Sir,

According to our records, you have not paid your income tax for 2015.

[BEHAVIORAL MESSAGE, IN VARIATIONS OVER THE BEHAVIORAL BASELINE LETTER, IS INCLUDED HERE]

If you do not pay your liability of PLN XXXX + any accumulated interest by June 3, 2016, you will be subject to execution proceedings.

Please pay your total amount due as specified in the table on the back of this page by bank transfer to the account XX XXXX XXXX XXXX XXXX XXXX XXXX, or by visiting your bank, postal service, or tax office.

If you are not able to pay your total amount due at this time or you have any questions, please call us urgently at XX XXX XX XX.

We will monitor your reaction to this letter.

Sincerely,

(Signature and stamp of Head of Tax Office)

Chapter 3

Career Choice and Cooperation in a Pro-social Team Task

3.1 Introduction

An organisation is "a means of achieving the benefits of collective action in situations where the price system fails" (Arrow, 1974). There is a strong interdependence between the actions taken by those who engage in this collective action, yet these people are usually very different with respect to their motivation and career drive. Understanding how cooperation within the organisation is affected by the encounter of people with different motivation is therefore key to understanding the success.

This issue is most visible in pro-social organisations, which often face a tension between recruiting workers by offering careers that appeal to financial drive and careers that appeal to motivation for the cause (including, but not limited to, volunteering). The tension is created by a combination of demand (the need for talented as well as the need for motivated workers) and supply factors (a substantive number of people is willing to forgo, fully or in part, their earnings in order to contribute some work to a pro-social organisation). In particular, the trade off faced by organisations is traditionally understood as one between recruiting motivated workers and more financially driven ones. This paper argues that an additional outcome should be considered in assessing this trade off, namely how different types of career affect cooperation in the workplace.

To understand the relevance of this question, consider a simple example. A charity is running a fundraiser at a sports event, where they are selling food and the proceeds go to the charity. The product sold is a combination sandwich and fries. At the venue there are two workers, one who makes sandwiches and one

who makes fries. It takes each of them five minutes to perform the task, yet if they exert some high effort they can speed it up to three minutes. A customer walks in, asks for the combination, but only has four minutes before the game resumes. After that time, they will leave and buy the product only if they get the full combination, sandwich and fries. The person who is making the sandwiches is a volunteer for the charity and needs to decide whether to apply high effort in fulfilling the order or not. They cannot speak with the other worker and cannot see whether they are making an effort or not. Yet they know that the other worker is not a volunteer, but is working for the catering company at a daily wage. Will the person who makes sandwiches take this information into account when deciding whether or not to make an effort?

This paper thus asks whether, in a pro-social team task, learning that a teammate has chosen a less financially rewarding career affects own effort. It does so through a lab experiment whereby participants are faced with a team task with perfect complementarity in effort. Subjects are asked to participate in a team task requiring filling envelopes for an Italian LGBT+ charity. At registration they are required to make a career choice on whether they want to volunteer part of their payment to the same charity, or not. They are then paired in teams with randomised partner assignment, informed of the partner's career choice and asked to sign in to a lab session, where the main task occurs. Effort is measured as the number of envelopes filled by each participant and time spent in the lab task. The lab setting allowed to design a task where complementarities of production and simultaneity of actions are present and made salient to the subjects, while at the same time keeping the task effective and generalisable to other activities that pro-social organisations and charities ask to do on a daily basis. The randomisation of the treatment allows to isolate the career choice from differences in skills or tenure, which would highly correlate to different career choices within the same company in an observational study.

The experimental design further allows to gain reliable insight into subjects' expectations on their partner, and gather evidence that learning about a teammate's career choice affects beliefs on that teammate's effort. In addition, I extend the understanding of beliefs by separately eliciting unconditional beliefs about all people who choose one career versus the other. This provides a more complete picture of the effects of being partnered with a volunteer on subjects' beliefs.

I find that those who are matched to a volunteer exert up to 19% more effort than those who are matched to a non-volunteer. I further find that this effect is mostly present among participants who exhibit higher motivation, such as those who made a non-binding request to be considered for additional volunteer work in the future. Finally, I find that expectations about the partner's effort increase when the assigned partner is a volunteer, and find evidence suggesting that this is the driving mechanism driving the effect of partner's career choice.

The results of the paper contribute to the economic literature by showing that workers and volunteers

are affected by the very selection process that brings them to work with a team or another, and attach informative power to their colleagues' career choice as a signal of their motivation. It also shows that the effect is driven by this informative power, and it is concentrated only among the more motivated agents. In the context of a pro-social organisation, this would suggest that mixing individuals with different career choice might be detrimental, as the loss in effort faced among the more motivated is not compensated by any increase in effort among the less motivated. However, when this mixing is needed, heterogeneity in career choice can be compensated with mechanisms ensuring that the more financially motivated agents are highly attached to the cause anyway.

Charities working on the field are aware of the difficulties of managing such a heterogeneous workforce, and are constantly assessing the right mix of staff and volunteers in their workforce. Other issues can arise from the interaction of these very different groups, such as fear of displacement, differential attrition rates, and the formation of social ties that might create resistance to change in the company's strategy. The unavailability of financial rewards and punishments for volunteers is another contributing factor. More research needs to be done on this topic.

Economic research has explored the topic of how career motivation can affect the match between a worker in a pro-social organisation and their performance. There are two main strands of literature: one highlighting the selective role of career on agents' type; another one highlighting the behavioural effect of career incentives on workers' performance in a pro-social organisation.

On the former side, Besley and Ghatak (2005) highlight the trade-off between high-powered economic incentives and matching with motivated agents, within mission-oriented organisations. Dal Bò et al. (2013) provide empirical evidence from a randomised controlled trial in public sector in Mexico, showing that financial incentives attract more qualified candidates. This general finding is confirmed in Ashraf et al. (2018) and Deserranno (2019), which however provide conflicting results on the possibility that financial incentives attract less motivated workers.

As far as the effect of career type and incentives on motivation, psychological literature has moved first and economics has built up on that more recently. Deci (1971), Cameron and Pierce (1994), Eisenberger and Cameron (1996) discuss the theory of crowding out by which financial incentives could displace intrinsic motivation for an activity. In their meta-research, however, Eisenberger and Cameron (1996) come to the conclusion that the scope for crowding out is very limited. In economics, however, the idea gained traction with Bénabou and Tirole (2003, 2006)'s theory. Ashraf et al (2014) and Gneezy and Rustichini (2000) in turn found some confirming evidence.

Both strands of literature, however, treat the role of motivation in a two-ended model, where the organisation stands on one side and the agent stands on the other. There is little study of how (or even whether)

the career tracks offered to worker (and accepted by them) affect the social interactions within the organisation. Yet the interpersonal dimension within the organisation is very important on many dimensions. The presence of conditional cooperators (Kessler 2017, Fischbacher et al. 2001, Bernheim 1994, Sugden 1984), whose cooperative behaviour is chosen as a positive function of teammates' cooperative behaviour, means that who is selected to take part in a job matters for performance and behaviour within the job.

Also the incentive structure is shown to affect relationships within the workplace. Lazear (1989) argues that individual incentives could create anti-cooperative behaviour that disrupts output, while Bandiera et al. (2005) to the contrary find that relative incentives can give rise to cooperation between workers at the expense of the firm. Independently of incentives, coworkers affect each other even when the production function or the incentive structure do not require cooperation (Mas and Moretti 2009, Bandiera et al. 2010) as they might be willing to conform to their group's standard norm (Douglas 1994). Even organisational identity (Akerlof and Kranton 2000) is an essentially communitarian aspect and relationship between coworkers is an essential part of it. Incentive schemes can affect how workers identify with the organisation (Akerlof and Kranton 2005), but in theory how workers identify with their role. Bringing workers with different roles together can make the respective identities more salient and facilitate aderence to one's own identity or be a source of conflict (Akerlof and Kranton 2000), as anecdotal evidence preliminary to this chapter has confirmed.

The results from this paper thus bring together the social dimension of the workplace and the selective role of incentives, to find that participants in team work actively form expectations on their partners' motivation based on how they responded to monetary incentives and respond to those expectations.

The rest of the paper is organised as follows. Section ?? shows a stylised model of the theory that is being tested to highlight its main predictions. Section ?? describes the experiment and the design. Section ?? shows evidence that a teammate's career choice has effects on individual behaviour. Section ?? shows that partner assignment matters more strongly among participants that exhibit higher levels of motivation. Section ?? shows how expectations are affected by partner's career choice and discusses that they could be the main channel through which partner assignment affects effort. Section ?? discusses and concludes.

3.2 Model

The stylised model in this section models the effect of partner's career choice on an agent's expectations and effort. It does so by clarifying the mechanism behind the career choice and by defining a heuristic that agents use to form their beliefs. This heuristic is supported by observational evidence in section ??. The model provides three testable predictions, namely that volunteers are expected to exert more effort, that

individuals increase their effort when their partner is a volunteer, and that this response is stronger among the most motivated ones.

3.2.1 The setting

There are two agents indexed by $i \in \{1, 2\}$, who are participating in a joint activity for a charity, whose output y depends on the two agents' efforts $e_1, e_2 \geq 0$ according to a Leontief production function $y(e_1, e_2) = \min\{e_1, e_2\}$. Effort cost is $c(e_i)$, increasing and convex, for both agents. Participation in the activity is rewarded $f > 0$ as reimbursement and $x > 0$ for taking time in the task, however after choosing whether to participate agents have the option to forego part of the payment and let it stay with the charity.

Each agent's final utility is thus represented by

$$u_i(e_i, v_i; e_j, v_j)$$

$$= \alpha_i \left(\lambda y(e_i, e_j) + xv_i \right) + (\beta_i + bv_j)e_i - c(e_i) + f + x - xv_i \quad ,$$

where $v_i \in \{0, 1\}$ represents the choice to volunteer the time spent in the task, $e_i \in \mathbb{R}^+$ is each agent's effort.

The parameter $\alpha_i \geq 0$ is agent i's preference for charitable output; $\lambda \geq 0$ is the MRS of one unit of output for money that goes directly the charity; β_i is the agent's intrinsic benefit from their own effort. Parameter $b \in \mathbb{R}$ allows for direct effects of partner's volunteering decision on each agents return from effort.

At time 0 values of α_i and β_i are drawn from commonly known distribution $F_{\alpha,\beta}$ and privately observed.[1] Parameters b and λ is instead equal and commonly known. This means that subjects who volunteer do not systematically differ from non-volunteers in terms of how much they care about output vis a vis effort. It is a strong assumption and worth keeping in mind when analysing the predictions.

At time 1 agents simultaneously choose whether to volunteer; at time 2 agents' decision to volunteer is disclosed [2] and they simultaneously decide their level of effort e_i. At time 4 agents' effort is revealed and payoffs are realised.

---

[1] We could relax this assumption and assume the distribution of the parameters is not known but agents have a prior about it and then use their observation about the partner's choice to donate to update the prior. That should justify why volunteers might lower their expectation on the volunteering effect when their partner is a volunteer, in the experiment.

[2] Note that we don't need to assume that agents don't expect this, as the utility function is separable in y and v_i; otherwise expectations on payoff from donating might have strategically depended on expectations on other agent's volunteering, e.g. if the partner is thought to be highly likely to volunteer, the relative incentive to volunteer is lower relatively to the incentive to put in effort. On a similar note, additivity of output and donation makes results independent of whether partner's contribution to the charity enters one agent's utility, although it may well be that knowing that the partner contributed makes one marginally less willing to put effort.

3.2.2 Solution concept

The solution is a profile of strategies and conjectures such that each player's strategies are rational in all subgames according to their conjectures and the conjectures are consistent with each player i believing that the other player j is rational and j's conjecture on i's effort does not depend on i's, nor on j's type. That is, each player i believes that j is rational but does not believe that j believes that i is rational.

3.2.3 Solution

At time 2, each agent makes a conjecture about the partner's effort c.d.f, $G_i(e_j|v_j)$. This conjecture **does not** depend on the agent's own volunteering decision, i.e. each agent does not believe that the partner's decision depends on their own donation decision. Agents instead expect their partner to make a default conjecture $G(e)$.[3] At time 2 agent i thus maximises

$$\max_{e_i} \alpha_i \left(\lambda \int_0^\infty \min\{e_i, e_j\} dG_i(e_j|v_j) + xv_i \right) + (\beta_i + bv_j)e_i - c(e_i) + f + x - xv_i \quad .$$

The first order condition thus defines agent i's optimal effort e_i^* is

$$\alpha_i \lambda \left[1 - G_i(e_i^*|v_j) \right] + (\beta_i + bv_j) - c'(e_i^*) = 0 \quad . \tag{3.1}$$

While it is immediate to see that optimal effort is increasing in α_i and β_i, this equation shows an undetermined relationship between partner's volunteering status and own effort. This depends on agents' expectations about their partner and the direct effect that their partner's volunteering decision has on their willingness to exert effort.

At time 1, by separability of the utility function in v_i and e_i, we can tell that agents choose $v_i = \mathbf{1}(\alpha_i \geq 1)$, that is if they care about one monetary unit given to the charity more than one unit kept for themselves. Volunteering is thus a signal of motivation.

Let us look directly into agents' beliefs over their partner. Each agent i believes that a partner j who made volunteering decision v_j maximises the following function at stage 2:

$$\max_{e_j} \alpha_j \left(\lambda_j \int_0^\infty \min\{e_i, e_j\} dG(e_j) + xv_j \right) + \beta_j e_j - c(e_j) + f + x - xv_i \quad . \tag{3.2}$$

Let us call $\tilde{e}(\alpha, \beta)$ the solution to the problem that each agent believes the partner is solving. This is

---

[3]This limitation simplifies the solution and is quite accurate in describing the sophistication of a sizeable share of experiment participants, as shown in Table **??**. Indeed most of them were not sophisticated enough to draw a distinction between the unconditional effort of volunteers/non-volunteers and the effort chosen by their partner, which is consistent with this strong restriction.

increasing in α_j and β_j We want agents to form beliefs that are consistent with this heuristic maximisation problem, that is, for each agent i, $G_i(e_j|v_j) = Pr\left(\tilde{e}(\alpha_j, \beta_j) \leq e_j|v_j\right)$.

3.2.4   Predictions

If α_j is uncorrelated or "positively correlated" with β_j (more precisely, if $F_{\beta|\alpha}(\beta_j\alpha_j)$ is decreasing in α_j), then it can be shown $G_i(e_j|1)$ first order stochastically dominates $G_i(e_j|0)$.

This is because, as $v_j = \mathbf{1}(\alpha_j \geq 1)$, also α_j is expected to be higher when $v_j = 1$. Then

$$
\begin{aligned}
G_i(e_j|v_j = 1) &= Pr\left(\tilde{e}(\alpha, \beta) \leq e_j|\alpha_j \geq 1\right) \\
&= \frac{1}{1 - F_\alpha(1)} \int_1^\infty Pr\left(\tilde{e}(\alpha_j, \beta_j) \leq e_j|\alpha_j\right) dF_\alpha(\alpha_j) \\
&= \frac{1}{1 - F_\alpha(1)} \int_1^\infty F_{\beta|\alpha}\left(\tilde{\beta}(e_j, \alpha_j)|\alpha_j\right) dF_\alpha(\alpha_j)
\end{aligned}
$$

where $\tilde{\beta}$ is the inverse of $\tilde{e}$: $\tilde{e}(\alpha_j, \tilde{\beta}(e_j, \alpha_j)) = e_j$.

Assuming $F_{\beta|\alpha}$ is non-increasing in α_j, and knowing that $\tilde{\beta}$ is non-increasing in α_j, we conclude that the argument of the integral is maximised at $F_{\beta|\alpha}\left(\tilde{\beta}(e_j, \alpha_j = 1)|\alpha_j = 1\right)$ and we thus obtain the following inequality:

$$
\begin{aligned}
G_i(e_j|v_j = 1) &\leq \frac{1}{1 - F_\alpha(1)} \int_1^\infty F_{\beta|\alpha}\left(\tilde{\beta}(e_j, 1)|1\right) dF_\alpha(\alpha_j) \\
&= F_{\beta|\alpha}\left(\tilde{\beta}(e_j, 1)|1\right) \\
&= \frac{1}{F_\alpha(1)} \int_{-\infty}^1 F_{\beta|\alpha}\left(\tilde{\beta}(e_j, 1)|1\right) dF_\alpha(\alpha_j) \\
&\leq \frac{1}{F_\alpha(1)} \int_{-\infty}^1 F_{\beta|\alpha}\left(\tilde{\beta}(e_j, \alpha_j)|\alpha_j\right) dF_\alpha(\alpha_j) \\
&= Pr\left(\beta_j \leq \tilde{\beta}(e_j, \alpha_j)|\alpha_j \leq 1\right) \\
&= G_i(e_j|v_j = 0) \quad .
\end{aligned}
$$

In such case, being partnered up with a volunteer will raise an agent's effort as long as $b \geq 0$, and this increase will be larger for agents with higher motivation α_i. Intuitively, as motivation increases so does the effect of information on one's own behaviour.

More formally, equation (**??**) implicitly defines a function $e^*(\alpha_i, \beta_i, G_i(\cdot|v_j))$. Applying the implicit function theorem, and making use of the envelope theorem, obtains

$$
\frac{\partial e_i^*(\alpha_i, \beta_i, G_i(\cdot|v_j))}{\partial G_i(\cdot|v_j)} = \frac{-\alpha_i \lambda}{c''(e_i^*)} < 0
$$

and

$$\frac{\partial^2 e_i^*(\alpha_i, \beta_i, G_i(\cdot|v_j))}{\partial G_i(\cdot|v_j)\partial \alpha_i} = \frac{-\lambda}{c''(e_i^*)} < 0 \quad .$$

This result implies the following predictions:

Prediction 1:

Agents expect volunteers to exert more effort than non-volunteers.

Prediction 2:

Agents who are matched with a volunteer exert more effort than agents who are matched with a non-volunteer.

Prediction 3:

Agents who are more motivated are more responsive to the information about their partner's career choice.

Prediction 4:

The effect that partner's career has on effort is fully explained by the effect that partner's career has on expectations on partner's effort.

3.3 The experiment

To test the predictions of the theoretical model I ran a lab-in-the-field experiment. The task exhibits Leontief production function and involves the production of real pro-social output, in a controlled setting and in a simultaneous way. This section describes the experiment in detail and shows how it was run.

3.3.1 The experimental task

The experimental task involved production of real pro-social output, namely envelopes for fundraising letters in support of an Italian LGBTI charity, *Avvocatura per I Diritti LGBTI - Rete Lenford* (henceforth, Rete Lenford). Team's were composed of two members, who would not know each other's identity nor communicate, either before or after the experiment. Team production was split so to maximise complementarities between the effort of team members. To do so, the output produced by a team would be a large envelope containing an A4 letter, with a smaller envelope inside that, in turn, contained a small paper sheet reading a Thank You note. Both the outer, large, envelope, and the inner, small one, would need to read the name and address of the recipient. One team member was given empty large envelopes, A4 sheets reading the letter, and a list of forty addresses from which to copy the name of recipients, envelope by envelope. This member's task would be to fold the A4 letter in two, insert it in the envelope, and then copy the address

from the list onto the back of the large envelope. The other team member, in turn, would be given a deck of smaller envelopes and smaller sheets reading a Thank you note. They were asked to fill each small envelopes with a Thank you notes and write addresses in order. The list provided to the two team-members was the same and it was reporting a clear ordering of the addresses that had to be written in each envelope. For this reason, only the minimum of the amount of envelopes filled by each team member will be sent out. Team members are clearly explained that any exceeding amount will be removed. Each team member was asked to fill at least twelve envelopes, but could fill as many as they wanted, up to forty envelopes as the list had forty addresses. These instructions were followed precisely by 95.7% of participants in the lab.

In order to stress the interactive nature of the exercise and make anticipation as relevant as possible, several aspects of the design were put in place. First of all, the complementary nature of the production process was stressed at all stages, from the preliminary survey to the declaration of consent before they entered the room, to the content of the instructions paper in the room and the setup of the material in the room (see below for details). Secondly, the cost of effort was higher by requiring subjects to perform the task in a small room with closed doors and no windows. Third, anticipation about the partner's decisions was made as needed as possible as subjects were not able to communicate nor know the identity of their partner, and were reminded that their partner might or might have not filled more than the minimum required amount of envelopes. This was achieved in particular by asking subjects to put their filled envelopes in two separate boxes: one designed for the minimum required amount of envelopes, and the other for any extra envelopes that subjects would want to fill. The words "Required" and "Optional", as well as the word "partner", were visible on coloured post-its that were stuck to the two envelopes.

3.3.2   The treatment

Before attending the lab, in a preliminary survey, potential participants were asked if they wanted to volunteer part of their compensation for attending the lab to Rete Lenford. This is what constitutes the *career choice* made by subjects, and is described in the next subsection. As subjects were partnered up with other potential participants, the treatment consisted of the type of career chosen by their partner. This created a *volunteer partner* condition and a *non-volunteer partner condition*. Details of the randomisation procedure are explained below.

3.3.3   The timeline

The experiment took place between 30 January and 23 February 2016 and it involved two stages: a preliminary online survey and a lab task.

Subjects from the mailing list of the LSE Behavioural Research Lab (BRL) were invited to take part in a five-minute online survey, preliminary to the final lab task, rewarded with a £2 Amazon voucher and participation in the lottery draw for a £250 Amazon voucher. At the end of the survey, they were then asked to participate in the experimental task in the lab. This task would be paid £7, of which £4 were explained to be a travel compensation fee, and £3 were considered as payment for the time spent in the lab.

However, at the survey stage, subjects were given an option to give up the £3 and volunteer their time in the lab, thus receiving only £4. If they decided so, the £3 would be sent directly to Rete Lenford. After making the career choice, subjects were then randomly assigned a partner and told the whether they were matched to a volunteer or a non-volunteer. They would then register online to book a session in the lab that would take place within two weeks' time. In the meantime teams were formed, so that if one team member did not participate the whole output from the team would be null. At the agreed lab session, those subjects that attended the session were given instructions out loud, tested for comprehension of the production function, and accompanied to the assigned room by the RA. In the room, with closed doors, subjects were asked to fill the envelopes. Once participants decided that they wanted to finish and leave the lab, they went to the RA and were asked to answer a debriefing questionnaire eliciting beliefs on the behaviour of their partner and of other experimental subjects. After this, they could collect the payment and leave.

3.3.4 The online survey

All students who registered with the BRL mailing list and had not participated in a pilot version of the experiment were invited to take part in the online survey. The questionnaire was designed to take less than five minutes and was preliminary to participation in the lab session within the following two weeks. However, subjects were also told they would not have to participate in the lab session after the online questionnaire. The posting of the study, which was designed to leverage both financial and intrinsic motivation, can be found in Figure **??**.

Invitations were staggered in such a way that about 300 contacts would receive an invitation in each week the online survey was run. Students in each week were randomly drawn from the initial sample of students registered with the BRL, but as new students signed up for that they were added to the survey, so selection into different weeks was not completely random.

Each week on Thursday students received the first email inviting them to participate, and on the following Monday were sent a reminder. After this, they would not be invited again even if they had filled the questionnaire but missed the registration.

Once a contact was invited once, they would not be invited again even if they had not filled the questionnaire at a previous time.

Participation in the survey was told to last about five minutes and is rewarded with a £2 Amazon voucher, to be sent to the participants. On top of this, respondents to the survey would be in for a £250 lottery draw. The survey was composed of:

- a preliminary section with demographic and more general questions related to gender, sexual orientation, knowledge, support, care and involvement with LGBT+ rights (Figures **??** to **??** show the relevant survey screenshots);

- a description of the experimental task with a comprehension test to check specifically for whether respondents understood the perfectly complementary nature of input in the task; respondents had two chances to get the comprehension test right (Figures **??** shows the relevant survey screenshot);

- a section where respondents chose whether to participate and, if so, whether they would donate the £3 to Rete Lenford (Figures **??** to **??** show the relevant survey screenshots);

- a message notifying them of the pre-assigned partner's career choice (volunteer vs. non-volunteer) and reminds them of the importance of complementarity and of registering, before redirecting them to the registration portal (Figures **??** and **??** show the relevant survey screenshots);

- a page showing information needed to register for the lab experiment (Figure **??**);

- a question asking respondents to express the desire to volunteer for the same charity in the future (Figure **??** shows the relevant screenshot).

At the end of the survey, respondents were redirected to the BRL's website where they could sign up for a session. Depending on whether they donated or not, they had two different types of listings, which were designed so that the volunteers and the non-volunteers would get equal access to the type of session they wanted. Indeed, given that the group of volunteers was one third as large as the group of non-volunteers, most volunteers would suffer limited access to the booking for the lab than most non-volunteers. This might have been a considerable source of selection bias regarding the decision to participate.

To gather better information about the respondents, their data were matched with their responses filled at the time of registration with the BRL (or at the beginning of the academic year for those who had registered earlier). Those variables mostly relate to demographic characteristics, socio-economic status, native language.

Students who responded to the survey, passed the comprehension test and then declared that they intended to participate in the lab task constitute the *respondents' sample*. The sample comprises 190 respondents, of which 50 declared they would volunteer and 140 declared they wouldn't. Of the volunteers, 27 were matched to non-volunteers and 23 were matched to volunteers. Of the 140 non-volunteers, 113 were matched to non-volunteers and 27 to volunteers. Table **??** reports own career choice and partner's choice among survey respondents.

The randomisation was conducted at the survey stage through qualtrics, which - together with the relatively small sample size - made it difficult to stratify among key variables. Balance tests provided in Tables **??** through **??** show that the sample is balanced in each subgroup. Given the different probability of being assigned a partner of one type depending on one's initial choice, balance has to be tested separately among non-volunteers and among volunteers, respectively. Among non-volunteers, balance is not reached only for socio-economic class, being native Italian speaker and (possibly correlated) having spoken the english language for over five years. Among volunteers, atheist or non-religious people were significantly more likely to be found among those partnered with a volunteer than those partnered with a non-volunteer (and, by consequence, christians were less common in the former group), and higher knowledge of LGBT+ rights was on average higher among those matched with a volunteer.

Overall the balance tests are satisfactory and inform us of the good quality of the randomisation.

3.3.5 The randomisation

The main treatment is the career choice (volunteer or non-volunteer) of a subject's assigned partner. This was notified at the survey stage, right after a respondent's decision to participate in the lab and to choose their career, but before they was register for the lab. This design feature was such that the very fact of signing up could be seen as a measure of effort and as an experimental outcome. It came to the cost of potentially inducing selective attrition, but this issue was dealt with separately, in ways that are explained below in the paper.

Partner's career choice was randomly assigned by the online survey platform, Qualtrics. To maximise power, the probability of matching with a volunteer was such that the smallest group (volunteers) would be equally split between those matched with volunteers and those matched with non-volunteers. Probabilities were computed based on the results from the last pilot version before the final experiment. The probability of being matched to a volunteer would be one half for a volunteer and one sixth for a non-volunteer.

At the end of the survey partner were asked to register, with the caution that - if they or their partner did not attend - the whole team output would be null.

The decision to assign the treatment before subjects would register was taken so that registration and attendance in the lab could be considered as outcomes, since attendance and dropout in the workforce are a relevant concern for many charities and other pro-social organisations, as preliminary interviews with CEOs from charities have shown.

3.3.6 The lab task

Each session lasted from about 30 minutes to one hour maximum. Pairs were formed in the morning before all sessions started. Subjects were not necessarily paired up with team mates from the same session. Lab participants arrived at the given time, were given five minutes to read the instructions and the consent form, were asked to sign the consent form and answer a comprehension check designed to ensure understanding of the production function, and were then taken to the room were the task would take place. The research assistant accompanied each participant to their room and informed them of the career choice of their partner (volunteer/non-volunteer).

In the room, participants found a pen, a list of addresses, a bunch of envelopes, and two boxes. Whether a participant was given the large or the small envelopes in their team, was established randomly as well as the room number where the experiment would take place. The task consisted of filling the envelopes and writing addresses on each envelope. At a later stage, smaller envelopes that were filled by one team member will be matched with larger envelopes filled by the other team member and inserted into them. For this reason, the production function is Leontief. Participants were required to fill at least twelve envelopes but had the option to fill more and put them in a separate box. The optional extra envelopes would be taken into account only insofar as the other team member would fill some extra envelopes, too. To remind participants of this feature of the production function, each box had its own label, one reading "Minimal required amount" and the other reading "Optional extra envelopes. Will be considered only if your partner fills some extra envelopes too". To make the partner's status more salient, a red post-it sticker was stuck to the latter box reading "Your partner has (not) donated" in accordance to the type of the partner.

Participants were then left alone and could leave when they preferred after filling at least twelve envelopes. Upon leaving, the RA recorded their exit time, asked them to fill a debriefing questionnaire eliciting beliefs on the partner, and gave them the compensation for attending.

The debriefing questionnaire elicited expectations on the number of envelopes filled by, respectively, the pre-assigned partner; the average volunteer participating in the experiment; the average non-volunteer participating in the experiment. Answers were incentive compatible as each correct guess (up to a 0.5 tolerance margin) doubled the chances of winning the lottery prize that students were entered by answering

the survey.

The *sample of participants* was thus composed of those respondents who registered online and attended the lab session within two weeks of the survey. Some respondents attended the lab session more than two weeks after the survey and were counted as not participating and not producing envelopes. Results are robust to this change, which involves 9 subjects or 4.74% of the respondents' sample.

Within the participants' sample, outcome variables are the number of envelopes filled, making errors in filling the envelopes with respect to the guidelines, and the time spent in the room.

Assignment of treatment was randomised at the survey response time, which ensured the exogeneity of treatment assignment. However, some variables were collected in the lab, at the participation stage. Expectations on other participants' effort, for instance, were only collected after subjects performed the task in the lab. This ensured that expectations could be measured as closely as possible to the time at which the subjects would choose their effort level. However, it means that any comparison between subjects in the sample of participants requires additional assumptions that are not ensured by randomisation alone. Namely we need to assume that treatment is exogenous, conditional on observables, in the participants' sample. That is, in expectation there must be no unobservable difference between those who attend and are treated and those who attend and are not.

Table **??** shows that there is suggestive evidence of differential attrition.

Tables **??** to **??**, however, show that balance on observables was preserved in the sample of those attending.

These data suggest that, while we will have to be extremely cautious in comparisons among those attending the lab experiment, the scope for imbalance is very limited, at least when it comes to observables. Given the wide range of categories covered by the observables, some of which are highly related to motivation over the specific cause, this is reassuring about possible self-selection on unobservables.

3.3.7   Methodological notes

The choice of the task is motivated by the need to generate real output in a way that neatly follows a clear production function, so to ensure that the task is regarded by all subjects as one with strategic complementarities. The fact that subjects are asked to fill envelopes for a charity has the advantage of creating real output that can be of benefit to the charity. This choice, however, comes at a few costs. The first one is that, as the monetary benefit from each envelope is not known and expectation is subjective, different subjects with the same degree of motivation might have different marginal benefits from the filling of each envelope. Effort in the task might thus reflect not just motivation, but dislike for the activity. Subjects who dislike the activity might turn to volunteering part of their payment as a substitute for effort in the

task. The choice of volunteering the payment, therefore, might not necessarily be positively correlated with effort. This issue, however, is only adding to the complexity of the study and analysing behaviour and, if anything, adds to the external validity of the results from the study, which allow us to directly see what subjects believe about the difference between volunteers and non-volunteers. Secondly, the splitting of a task such as filling envelopes into two separate parts, each done by a separate subject, and its potential for waste generation, might sound too artificial and defeat the purpose of a task that mimics real-life charitable activities. This concern was considered at the piloting stage, where subjects were invited in one-to-one and group interviews, and asked why did they believe that the task was split in two. Most subjects at that stage mentioned efficiency as the reason for the task being split in two. One subject did mention that this was done to measure individual contribution, but no subject answered that they believed the split was created specifically to introduce complementarities.

The choice of charity deserves separate discussion. The choice was mostly driven by ease of access and pre-existing contacts with the management of the charity. One's concern is that an Italian LGBTI charity might be too distant from subjects' experience and values to make a big difference. The concern is that the results could not easily be extended to other organisations. There are merits to these concerns: in particular it is possible that motivation might play a bigger role in activities for a well-known charity or for one that is perceived as closer to the participant. However, once we take into account that the effect might be somewhat weakened by the distance of the charity, it is hard to hypothesise how the treatment effect might be different with another type of charity. On the other hand, the choice of a distant charity has the practical advantage of being hardly substitutable with other charities. Subjects who are motivated by the cause will have a hard time finding close substitutes to an Italian LGBTI charity: effort in the experiment therefore becomes more valuable to such subjects as there are few make-up options outside of the experiment. If the experiment had been conducted with a more easily accessible charity, instead, subjects might have more easily lowered their response to information simply because the charity would offer other opportunities of volunteering.

Finally, the experiment was not pre-registered because of time considerations and the little availability of the LSE BRL, which did not allow for precise planning ahead of the experiment. In particular, it was not possible to pre-determine when the pilot fase would end and when the experiment would start. The main concerns that could be raised by the absence of pre-registration are related to measures of effort. However, the design of the experiment is clear itself that the main outcome of individual effort is the number of envelopes filled by subject. Our analysis is therefore focused on that measure, and on time spent in the lab as a secondary measure. An alternative interpretation of effort could involve the correct filling of envelopes according to the rules. Indeed, a sizeable share of subjects did not follow the guidelines, making the matching impossible and disrupting their team's effort. There is then a clear distinction between the

number of envelopes filled as a measure for the amount of effort, and following the guidelines as a measure of the quality of effort. This is not the main outcome of interest and we had no null hypothesis to test about this trade off. This measure of "quality of effort" will therefore be discarded from the main analysis. Results for it and for other outcome measures will still be shown, not as tests to the theory but as informative statistics that are worth exploring.

3.4   Effort and the effect of partner's career choice

Effort, in the form of envelopes filled, increases among lab participants that are "treated" with a volunteer partner. The estimated treatment effect in the experiment is of 3.35 envelopes or an increase of effort by 13.4%. Effort measured as time spent in the activity increases but not significantly.

Tables **??** and **??** report all outcomes by career choice and partner's career, in the samples of survey respondents and lab participants respectively. A few patterns emerge.

First of all, the volunteering-partner treatment is negatively associated with attendance in the lab in the sample of volunteers, as seen in Table **??**. This effect, however, is the result of the addition of two separate stages: the first, of registration, where respondents sign up into their lab session right after being informed of their partner's career choice and being reminded of the importance of attending, as the team has already been formed. The second stage of attrition occurs between sign-up and attendance in the lab. While the compound attendance rate in the lab is significantly different among the two treatment cells, at the sign-up stage there is no significant difference between the two groups. This suggests that the correlation is not interpretable as causal, as the effect of partner's career choice is arguably less relevant after registration, when the information on the partner's career has had more time to be forgotten.

Secondly, volunteers do not seem to exert more effort than non-volunteers, as measured as number of filled envelopes or time spent in the task. This suggests that the choice of volunteering is only partly related to motivation, or that subjects might consider their choice to volunteer part of the payment as a substitute for filing envelopes.

Third, participants that are partnered up with a volunteer exert on average more effort than those who are not "treated" with a volunteer partner. A simple comparison of the treatment groups suggests that the effects are of similar size among volunteers and non-volunteers. The distribution of effort is also more dispersed among those who are "treated" than among the "untreated".

Among non-volunteers, those matched with a volunteer are significantly more likely to make an error in following the guidelines, and therefore the difference in terms of correctly filled envelopes loses significance. As this finding is incidental, this outcome is not considered as a measure of effort; yet this observation could

suggest a potential trade off between quality and quantity of effort, and further research understand how the trade off is affected by characteristics of the environment.

Finally, subjects that are treated by being assigned a volunteer partner are not faster than the untreated and spend more time in the task, albeit not significantly in a statistical way.

Figures **??** and **??** confirm this picture by showing the distribution of effort among those who attend the lab experiment. The distribution has a longer upper tail among the "treated" than the untreated. This observation also explains the higher standard deviation of effort in the group of treated, as reported in Table **??**.

3.4.1 Econometric analysis

The econometric approach to estimating the effect of matching relies mainly on the identifying assumption of exogeneity of treatment. As partner type is assigned randomly, this assumption can be safely accepted. Furthermore, the summary statistics shown in Tables **??** to **??** confirm that observables are balanced across treatment conditions.

Table **??** reports the marginal effects from maximum likelihood estimation of four models of effort, named v1 to v4. Model v1 sees as its outcome the number of envelopes, and models separately the hurdle of attending the lab session and the choice of how much effort to exert. The model accounts for the discrete nature of the number of envelopes as well as the censoring that happens at the minimum required amount of envelopes filled (12) and at the upper bound (40). The latent variable can thus be interpreted as the desired amount of effort that would have taken place but for these restrictions. Models v and v2 thus model the number of envelopes and the log number of envelopes, respectively, by taking into account that each variable is censored to be between 12 and 40, or $\ln(12)$ and $\ln(40)$ respectively. Letting $\underline{y}$ and $\bar{y}$ be the lower and upper bound of each variable, models v1 and v2 are represented by the following system of equations:

$$
\begin{cases}
A_i = \mathbb{I}(\beta_0 V_i + \gamma_0 PV_i + \delta_0 V_i PV_i + Z_i \theta_0 + \nu_i \geq 0) \\
y_i^* = \beta V_i + \gamma PV_i + \delta V_i PV_i + X_i \theta + \epsilon_i \\
y_i = \underline{y}\mathbb{I}(y_i^* \leq \underline{y}) + floor(y_i^*)\mathbb{I}(\underline{y} < y_i^* < \bar{y}) + \bar{y}\mathbb{I}(y_i^* \geq \bar{y}) \quad \text{if } A_i = 1
\end{cases},
$$

where y_i is the outcome of interest, A_i is a dummy for attending, V_i a dummy for being a volunteer, PV_i a dummy for being assigned a volunteer partner, Z_i a set of controls for attendance, X_i a set of controls for effort. The error terms are $\nu_i | V_i, PV_i, Z_i \overset{iid}{\sim} N(0,1)$ and $\epsilon_i | V_i, PV_i, X_i \overset{iid}{\sim} N(0, \sigma_i)$. The error terms are assumed to be independent of one another: conditional on treatment and covariates, unobservable characteristics that affect attendance are not related to any unobservables that affect effort.

Models v3 and v4, instead, model the dependent variable (time spent in minutes and log minutes, respectively) as unbounded:

$$
\begin{cases}
A_i = \mathbb{I}(\beta_0 V_i + \gamma_0 PV_i + \delta_0 V_i PV_i + Z_i \theta_0 + \nu_i \geq 0) \\
y_i = \beta V_i + \gamma PV_i + \delta V_i PV_i + X_i \theta + \epsilon_i
\end{cases}
.
$$

The error terms are $\nu_i | V_i, PV_i, Z_i \overset{iid}{\sim} N(0,1)$ and $\epsilon_i | V_i, PV_i, X_i \overset{iid}{\sim} N(0, \sigma_i)$. As in models v1 and v2, the error terms are assumed to be independent.

The results of these models are reported in Table **??**. Column (1) reports marginal effects at the mean for determinants of attendance. Of all covariates, the only significant determinant is personal income: at the sample average, a subject with personal income above the median in the sample is 32 percentage points less likely to attend than one with personal income below. Most importantly, however, there is no significant effect on the probability of attending associated with being a volunteer. On average, being partnered with a volunteer for a non-volunteer reduces participation, not significantly, by 1 percentage point. This negative effect is much stronger (25 percentage points) among volunteers. Despite this large differential, the hypothesis that the treatment effect is the same among volunteers and non-volunteers cannot be rejected.

Models v1 to v4 show consistently that volunteers do not exert more effort than their counterparts. Being partnered up with a volunteer, however, has a strong significant effect on the number of envelopes filled, although not a significant one on the time spent in the task. If we considered the latent variable, i.e. the desired number of envelopes but-for the restriction on minimum and maximum amount, the estimate is that being matched with a volunteer increases effort by 4.13 envelopes and 15.8 percent. The estimated effect on the observed number of envelopes is obviously smaller but still significant, at 3.35 envelopes and 13.4 percent. There is no significant difference in this treatment effect between volunteers and non-volunteers.

There is, however, no significant treatment effect on time spent.

Interestingly, self-reported variables related to motivation do not have a strong effect on effort, nor does sexual orientation. A very strong predictor of effort is, instead, the reported interest in additional volunteering for the charity. This variable is consistently associated to higher effort in terms of envelopes filled as well as time spent in the task.

Finally, religious differences do not seem to be significant. Gender is not associated to higher envelopes filled, although it is associated with time spent doing the task, with men spending on average 3.8 more minutes.

Similar results are obtained when restricting the attention on attending participants and performing OLS

regression. Such results are shown in Table **??**, that applies to the participants' sample the regression model

$$y_i = \beta V_i + \gamma PV_i + \delta V_i PV_i + X_i \theta + \epsilon_i \quad , \tag{3.3}$$

where y_i is the outcome of interest for respondent i, V_i is the dummy equal to one if person i decided to volunteer; PV_i is the dummy equal to one if i's assigned partner is a volunteer, X_i is a vector of controls.

Controls include gender, care for LGBT rights and knowledge of them, personal income above the median, sexual orientation, number of other people attending the same session.

The identifying assumption that $cov(\epsilon_i, PV_i | V_i, X_i) = 0$ is satisfied if we assume away selection on unobservables, that is $E(\epsilon_i | PV_i, V_i, X_i) = 0$. This assumption implies that, despite the sizeable (yet, not significant as seen above) difference in attendance between the two treatment groups, these do not systematically differ in their selection patterns based on other observables. Appendix **??** discusses the assumption and shows that there is no systematic difference in selection on observables between treatment arms. This leads to conclude that differential attrition is not an invalidating issue for an OLS analysis of outcomes among participants.

The first result is that, when matched with non-volunteers, volunteers do not significantly differ from non-volunteers, as all coefficients in the top row of Table **??** show.

Secondly, the partner's career choice matters. Being assigned a partner who is a volunteer is estimated to raise the number of envelopes by 3.2 (or 18.5% of the unconditional mean among non-volunteers, or 55% of a standard deviation among non-volunteers). However, among non-volunteers there is a significant increase in the percentage of those that make mistakes in filling the envelopes (a 13.2 percentage points increase) which reduces the effect of partner's career choice on the number of envelopes filled, once incorrect envelopes are counted away. The increase in the time spent in the lab is sizeable yet not significant. This is partly due to measurement error, as time was measured manually by a research assistant as subjects entered the room and as they left to collect their payment, with chances of queues forming and ensuing mismeasurement. On the other hand, however, it looks like partner's career choice does not have a significant effect on the speed with which envelopes are filled. It is thus hard to give an ultimate answer to the question of whether treatment increases effort by increasing productivity or by making people spend more time on the task. The answer is probably a combination of both, and given the context in which the experiment was conducted it is not too relevant: at the margin, the opportunity cost of filling an envelope (about one minute and twenty seconds) is not too high, especially for students who are mostly not engaged in paid work.

Finally, the effect of the partner's career choice on effort measures is homogeneous across volunteers and non-volunteers. However, being paired with a volunteer increases the probability of error among non-

volunteers. If making an error in following guidelines is determinet by worse attention to instructions, this could suggest a trade off between quality and quantity of work. How partner assignment affects this trade off is beyond the purpose of this paper, but an interesting avenue for future research.

3.4.2 Distribution of effort

The theoretical model implies that the effect of partner assignment is stronger for those agents whose motivation is higher. If motivation is a determinant of effort, and if the treatment effect of partner assignment matters more to those who are more motivated, then the effect of partner assignment should spread the distribution of effort by raising it at the top. This effect is in line with previous findings: namely, that the standard deviation of envelopes filled is higher among those treated with a volunteering partner (Table **??**) and that the distribution of effort has a longer tail among the treated, as seen in Figures **??** and **??**.

As the theory thus suggests that the treatment is stronger on the upper tail of the effort distribution, we should expect to find a stronger effect on higher quantiles of effort. Figures **??** and **??** confirm this prediction by reporting estimates from quantile regressions of the type

$$Q_{y_i|V_i,PV_i}(\tau) = \alpha_\tau + \beta_\tau V_i + \gamma_\tau PV_i + \delta_\tau V_i PV_i \quad , \tag{3.4}$$

where τ is the quantile of interest. The regressions are limited to the sample of lab participants.

Figure **??** (b) shows that the number of envelopes increases considerably and significantly on and above the seventh decile of the distribution when the assigned partner is a volunteer. Figure **??** (b) shows that the increase in time spent, instead, is more widespread across the different quantiles and concentrated in the lower half of the distribution.

Figures **??** (c) and **??** (c) instead show that at no quantile do volunteers respond differently from non-volunteers. This could detract from the theory that partner effect is more concentrated among the more motivated, or arguably be evidence that volunteers are not significantly more motivated than non-volunteers.

3.5 Heterogeneous effects of partner's career

This section provides stronger evidence in support of Prediction 4, that more motivated agents are more responsive to the treatment than the less motivated ones. The chosen proxies for motivation are three variables: reporting interest in volunteering again in the future, reported care for LGBT+ rights, and sexual orientation.

Tables **??**, **??** and **??** report the OLS estimates of coefficients from equation

$$y_i = \alpha + \beta V_i + \gamma PV_i + \delta V_i PV_i + \lambda M_i + \xi M_i PV_i + \epsilon_i \quad , \tag{3.5}$$

where y_i is an effort outcome, and M_i is the relevant motivation variable. Having ruled out the potential for selective attrition, the identifying assumption of exogeneity can be applied to the sample of lab participants.

Estimates of coefficient ξ are generally positive and significant for interest in further volunteering and care for LGBT+ rights, while not significant for sexual orientation, which however does not seem to be a determinant of effort in the first place.

Table **??** reports coefficient estimates among respondents in the survey. Among those who report more interest in volunteering, the effect of being assigned a volunteer partner is larger by 9.6 envelopes and 8.7 minutes in the lab. Column (1) shows that the effect of being assigned a partner who is a volunteer is almost entirely concentrated among those who care about volunteering for the cause in the future. There is a similar, although less significant, effect on time spent.

This proxy for motivation could however conflate three factors: one's own motivation, low cost of effort or low opportunity cost of time. To isolate the first of these effect, another proxy for motivation is care for LGBT+ rights, used for heterogeneity analysis in Table **??**. This proxy, however, is a purely self-reported variable and could make for a less reliable signal of motivation. Being one standard deviation more caring of LGBT+ rights than the mean, makes one person more responsive to the assignment of a volunteer partner by 3.76 envelopes, or altrnatively it as an additional 20% effect on top of the treatment.

Finally, Table **??** shows that sexual orientation per se is not an important source of heterogeneity. This seems to be related to the fact that being heterosexual is not, per se, a highly driver of effort, as one might expect.

Incidentally, one result is interesting and it appears from column (3) of Tables **??** and **??**. Those who care more about the cause respond to being assigned a volunteer partner by exerting more effort, but are also more likely to commit mistakes in interpreting the instructions. This would be consistent with a model where agents have a fixed amount of energy that can be devoted to producing output or to reading carefully through the instructions. Overall, being assigned a partner who is a volunteer raises the effort spent on this whole consumption of energy, but is also makes the production of output more salient than the correct interpretation of instructions. One might see the interaction coefficient for the *wanting to volunteer* variable as indicative that those people who want to volunteer have less skills and are therefore facing a more binding constraint in their energy consumption, so will need to divert their attention away from envelopes in a more consistent manner. Yet this interpretation is farfetched as it does not explain why a similar pattern is

observed with the *care for LGBT+ rights* variable.

Overall, we can conclude that the effect of partner assignment is concentrated among more motivated subjects. This result is very important as it calls for higher attention to team composition precisely in those settings where motivation is more relevant and heterogeneous, such as charities that can be joined by people motivated by career or by the cause.

3.6   Expectations

An important aspect of the theory outlined in this paper is that the career chosen by the assigned partner affects effort by changing expectations on the partner's own effort. Ideally we would then like to add exogenously vary expected partner's effort, and test whether this decreases the effect of partner assignment, as would be expected from the model. Unfortunately, however, the size of the sample was not large enough to allow an additional dimension of treatment. An additional test would be the inclusion of expectations as a "bad control" and the test for whether their inclusion decreases the statistical significance of treatment. The problem, however, is that expectations on partner's effort are collected after the treatment is assigned, so they are probably affected by the treatment themselves. A less ambitious, yet very informative, exercise, is to learn as much as possible about expectation formation as an outcome per se.

Section **??** uses elicited expectations to understand the mechanisms of expectation formation. It shows that being mached with a volunteer increases expectations on partner's effort, yet affects beliefs in more complicated ways.

Section **??**, instead, uses mediation analysis to establish that the effect of partner's career on individual effort is mediated by expectations. This rules out the alternative explanation that a partner's identity may directly affect an agent's returns to effort.

The data collected from this part of the experiment are summarised in Table **??** and discussed in Appendix **??**, which further explains the definition of some variables included in Table **??**.

3.6.1   Measurement of expectations

To gain as much information as possible on how expectations are affected by partner's career choice, at the end of the lab session participants were asked to answer a few questions about their expectations and their reactions to them. The questionnaire that they were asked to fill is shown in Figure **??**. The timing of the questionnaire was designed to to make sure their expectations would be informed by their experience of filling the envelopes and that the question would not bias the participants' behaviour in the task.

First, respondents' beliefs are elicited through incentive compatible questions. The first question is

the expectation on how many envelopes has the partner filled. This question, however, depends on the partner's career choice. To obtain information about the counterfactual expectation, two more expectations are elicited: that on all volunteers and that on all non-volunteers. I argue that these variables provide a reasonable approximation of the counterfactual, thus providing useful information on how expectations respond to partner's career. There are, indeed, two possible mechanisms through which partner's career choice can affect expectations. Learning that one's partner is a volunteer might, on the one hand, increase the expectation on the partner's effort (assuming one expects volunteers to exert more effort, which is true for most subjects). On the other hand, it might as well inform that volunteers are more common than otherwise expected, therefore that volunteering is not such a relevant correlate of motivation, thus lowering one's expectation over their partner. These two effects cannot be disentangled when only asking for expectations on the partner, while asking for expectations on volunteers and non-volunteers separately helps separate them out.

Therefore, subjects are asked separately about their expectations on the number of envelopes filled by their partner, by non-volunteers, and by volunteers. The correct answer is, respectively, the number of envelopes filled by the partner assigned; the average number of envelopes filled by non-volunteers (13.14) throughout the experiment (13.36); the average number of envelopes filled by non-volunteers throughout the experiment. As teams were formed before the lab sessions, a subject who does not attend is recorded as filling zero envelopes for the computation purposes. For each correct response (up to a 0.5 error) the probability of being drawn in the lottery would be doubled. A subject who correctly answered all questions will thus be eight times as likely to win as a subject who did not answer correctly any of them.

Finally, to get more understanding of participants' thought process, we asked them directly what they remembered about their partner's career choice, whether they were affected by their partner's career choice and whether that information affected their behaviour.

3.6.2 The effect of partner's career on expectations

It has been noted from the above empirical analysis that there is no significant difference in effort exerted by volunteers and non-volunteers. The theory, however, posits that subjects respond to partner assignment precisely by updating expectations on effort. If subjects had correct expectatoins on the partner, in this particular context, they would not change their expectations on the partner. The question thus become how do subject update their expectations on the partner when they are matched with a volunteer.

Table **??** shows that all subjects that are matched with a volunteer have higher expectations on their partner's effort, and this effect is similar between volunteers and non-volunteers: volunteers raise their

expectation on partner's effort on average from 14.6 to 16.35; volunteers raise their expectations from 14.17 to 18. The difference is significant in both groups. Figure **??** (a) further shows that partner information shifts the whole distribution of expectations on partner's effort up, with first order stochastic dominance of the distribution obtained after matching with a volunteer.

These results mask, however, interesting differences in unconditional beliefs on volunteers and non-volunteers.

Table **??** shows that being matched with a volunteer significantly increases expectation of effort exerted by one's own group: from 13.98 to 15.32 envelopes for non-volunteers; from 15.87 to 18.15 envelopes for volunteers. Figures **??** (b) and **??** (c) confirm this pattern.

The net effect on the expected "volunteer effect" (subjects' expected difference between volunteers' and non-volunteers' filled envelopes) is not clear. Figure **??** (d) shows. While the percentage of people who report to believe that volunteers exert more effort than non-volunteers stays constant, conditional expectation on this volunteer effect becomes stronger among volunteers. On the other hand, expectations on the volunteer effect on the non-positive end of the distribution become significantly more negative, both among volunteers and non-volunteers.

While these effects seem hard to square with any given theory, it is not the purpose of this paper to explain them. It is sufficient for the above analysis to note that, when asked about partner's effort, subjects matched with a volunteer consistently have higher expectations at all points of the distribution. Appendix **??** proposes a theory and assesses how it fits the observed data.

3.6.3 Expectations as a channel

The theory highlighted in this paper is that partner's career affects effort by directly affecting expectations on the partner's effort. This is not the only reason why information on partner's choices could affect own effort in a pro-social task. We could allow, for instance, for information on partner's career choice to directly affect one's utility by inspiring pro-social feelings, through conformity or by signalling higher quality of the project. Under this alternative theory, there would be potential for long term benefits from being assigned to a more motivated team, that would go beyond effort in the team task and affect individuals' attachment to the same or other causes. It is thus relevant, from a policy as well as an academic perspective, to distinguish between effects of partner assignment that are mediated by expectations on partners' effort, and other indirect effects, for instance on the utility function.

To do so, a mediation analysis exercise is needed, by estimating equations of the form

$$y_i = \alpha + \alpha_0 V_i + \beta_0 PV_i + \beta_i V_i PV_i + \mathbf{EP_i}\gamma_0 + V_i\mathbf{EP_i}\gamma_1 + \epsilon_i \quad , \tag{3.6}$$

where $\mathbf{EP_i}$ is a vector of variables representing expectations on partner's effort.

Table ?? shows the results of this mediation analysis using different specifications for $\mathbf{EP_i}$. The assumption required to interpret these results is sequential ignorability (Heckman et al, 2015), i.e. the requirement that the mediator be independent of potential outcomes, conditional on treatment and covariates. This is unlikely if the mediator is the belief about the partner. Guess on partner's effort could, for instance, be correlated to general optimism about people's motivation, which would be in turn affected by the treatment. For this reason it is useful to include, in the list of mediators, other proxies that are available that capture parts of the process of expectation formation. In columns (4) and (5) this is done by including controls for unconditional expectations on donors and non-donors. The assumption now becomes that, conditional on treatment, covariates and beliefs on the population of volunteers and non-volunteers, the expectation on partner's effort is independent of potential outcomes.

In all specifications, after accounting for expectations on partner's effort, being matched with a volunteer loses statistical significance as an explanatory variable. Under the assumption of sequential ignorability, this statistical resul can be interpreted as evidence that there is no direct effect of partner assignment on effort occurring outside the hypothesised channel of expectation formation (prediction 4 of the model). In particular, there seems to be no inherent relationship between the incentive to cooperate and the type of group the partner belongs to. In this context, channels such as identity formation can therefore be ruled out.

3.7 Conclusions

The lab-in-the-field experiment discussed in this paper provides evidence in support of a theory whereby cooperation in a team is affected by expectations about other team members' contribution, in a framework whereby the effort levels of different agents are complementary and unobservable, while their career choice is common knowledge.

The methodological innovation is two-fold. First, the experimental setting allows to study perfect complementarities with a well known type of task while modifying it to account for complementarities in production. One might be worried that this variation might sound slightly artificial and unclear to participants. In particular, participants might wonder why the spare envelopes produced by a person in a team could not be matched by participants in a latter stage. Yet, individual and panel interviews (available upon request) conducted among participants in pilot version of the experiment show that the structure of the production function was not considered surprising for the task at hand. Participants made sense of it by making reference to "efficiency" and other logistical reasons, or simply by inferring that they had to be measured with

respect to their own individual effort.

The second methodological innovation is that, differently from other experiments in the literature (most notably Kessler, 2013), beliefs are elicited not only on the factual behaviour of one's partner, but also on counterfactual behaviours. That is, participants are asked to make a general prediction about donors and non-donors. This added richness of detail allows to study the mediating effect of beliefs by making less restrictive assumptions than otherwise done. For instance, being informed that one's partner is a volunteer could affect one's beliefs about their partner directly, but also by convincing them that volunteers are less rare and not as highly motivated as otherwise thought. This would, in turn, affect beliefs about the whole population. Without controlling for beliefs on the whole population of volunteers and non-volunteers, we might underestimate the direct channel that leads from partner's volunteering to own effort via expectations on the partner. This might lead to overstate the direct effect of partner's career choice on effort, while to the contrary the whole effect would be driven by expectations - just in two different ways.

The experiment finds that, consistent with the theory, participants in the experiment generally believe that volunteers exert more effort than non-volunteers, and anyway believe that their partner exerts more effort when they are matched with a volunteer. While the former result is less general, the latter is generally found across all participants. Being assigned a volunteer does not have the same effect on the individual's expectations of the volunteer effect, however. Among volunteers, for instance, being partnered with a volunteer has a polarising effect: some increase their expectations on the differential between volunteers and non-volunteers; some instead start believing that non-volunteers actually increase their effort compared to volunteers. These results suggest that the mechanism of expectation formation needs more understanding and additional research.

Participants, consistent with the theory, also adjust their contribution to the team after receiving information on the career chosen by the partner they are matched with. Indeed, being matched with a volunteer considerably raises effort in the task. The distributional effect is higher at the upper tale, suggesting that those who would put high levels of effort were they matched with a volunteer become discouraged when they are told that their partner is a non-volunteer, and so cut down their effort levels.

The more responsive, again consistent with the theory, are those who care more about the cause and about volunteering. LGBT+ participants, however, are not more responsive, perhaps as people from this group already have other outlets to contribute to the cause, than the specific task performed in the lab.

A mediation analysis exercise has provided evidence in support of the theory that partner's career choice affects effort by increasing expectations on a partner's effort. A direct channel from partner's career to own effort is instead excluded. In this experiment, then, participants do not have idiosyncratic preferences for being matched with people from their own group, or if they do, these do not affect the marginal cost of

effort.

The results imply that agents take into account information about their partner when they are choosing how much effort to contribute to their team while unable to observe their partner's effort. In particular, being told that one's partner is a volunteer raises the effort and therefore cooperation with the team. This effect is, however, concentrated among those who are more motivated, and absent among those who do not care about the output.

The most obvious application of this result is to a charity recruiting volunteers and non-volunteers for a similar job and having them work side by side. As anecdotal evidence in the research leading up to this paper has shown, oftentimes - when starting performing new activities and recruiting new figures - start from hiring professionals and only later, after they have gained experience on the field and are ready to scale up, recruit volunteers with larger campaigns. Volunteers are a good resource for many reasons, and allow the charity to expand fast because of the reduced direct compensation that they require. However, this paper shows that mixing volunteers with professionals during the expansion process might demotivate volunteers to start with, and yield to outcomes below the potential. This is the case especially if non-volunteers are less motivated, as the results show that the less motivated agents will not increase their effort in response to being matched with highly motivated volunteers, while the highly motivated volunteers will indeed decrease theirs even before learning that their partners are indeed less motivated. Homogeneity in the dimension of mission alignment is, therefore, key.

It is highly possible that different career tracks can be sustained as long as agents from the less mission-oriented track are selected more carefully. When mixing is needed, such as in charities that are highly financially constrained, the results from the experiment suggest two possible avenues. One is the manipulation of information about other teammates. For instance, collecting statistics about the professionals working from a charity and selecting the information that makes them appear more motivated. Ethical issues aside, this approach is not sustainable in the long run if agents are not naive and if they can easily learn the true motivation of their less motivated partners. This will indeed be easy to infer because they will not change their behaviour regardless of any information they receive. A second avenue is to add high-powered incentives to the pay of professionals. Taking the results to the letter, this should work because it will increase effort among professionals so to match the higher levels by volunteers.

Concerns about fairness and other-regarding preferences (Fehr and Schimdt, 1999), however, can alter the policy implications discussed above as well as the very results from the paper. Being matched with someone who works on the same task but do not voluntarily choose to be paid less, would lead to inequality that would almost certainly discourage even the most motivated of the volunteers. There must thus be socially acceptable reasons to create different career tracks within a company. Such reasons can include

differences in the task, in seniority levels, or in the degree of responsibility. People who work for charities by promoting direct debit donations in public venues, for instance, have a monetary pay. Interviews with charity representatives in the preliminary work for this project suggests that this is a form of efficiency wage to discourage shirking, given the highly sensible information that these workers have to gather. The results from this paper, and the ensuing policy implications, must be carefully weighted against any concerns for other-regarding preferences.

One key element of the experimental design was indeed that subjects could self-select into the choice of volunteering or being professionals. The application of the results to reality need careful consideration. Who would be a volunteer and who would not be a volunteer in a charity? The CEO whose best alternative is to lead a for-profit, non pro-social, company, with a higher bonus, or the volunteers? And in a group of volunteers performing the same task, e.g. counting money for a fundraiser, whom should we think of as a volunteer in the sense of this experiment? The pensioner, with nil opportunity cost, or the young professional, with a higher opportunity cost? Tu understand reality in light of this experiment, it seems necessary to consider the opportunity cost as a proxy for volunteering.

Another specificity of the experiment, as well as the model, is that in this framework agents are not told that their volunteering decision will be revealed. In the theory, this is modelled as agents actively not expecting this information to be revealed. In the experimental application, however, things are more nuanced because participants are told that they will learn "some information" (Figure **??**) about their partner before making their own decision, but not exactly what information. If they understand that the screen they are shown is the same as that shown to their partner, they can understand that their partner will learn some information about them and, possibly, information about their volunteering choice. Unfortunately the respondents' interpretation of this question was not tested, and we cannot tell. If volunteering could be thought of as a strategic choice, however, the effect would depend on the expectations that people have and how they change them.

Once all the limits and concurring factors are taken into account, the results can generalise to all organisations with a strong identity and to work activities that entail high scope for intrinsic motivation. Education and healthcare are obvious examples. In that case, offering one single career track seems to be the most sensible solution. Besley and Ghatak (2005) argue that mission oriented organisations will match with mission-oriented workers, and vice versa. This paper adds on to that argument by making the case for a homogeneous workforce and by highlighting the asymmetric impact that heterogeneity has on the more and the less motivated.

To some extent, these results can be extended to goods and services with network externalities or externalities of consumption. Anecdotal personal evidence of attendance of large concerts of indie bands gone "too

mainstream", for instance, shows that moshing is far less likely to occur at large and cheaper concerts than in carefully selected, more exclusive venues. In this example, moshing would be the cooperative behaviour and the volunteers would be those who are willing to undergo higher sacrifice to attend the concert. Of course the problem is solved if the hardcore fans can queue early in the morning to find their space in the mosh pit.

Many questions remain to be answered. First of all, what happens in a long-run relationship: can other means of communication facilitate cooperation in teams whose agents have different attachment to the cause? Could differences in motivation be negotiated with other aspects of the relationship between coworkers? What happens in relations that are closer than the one studied in this experiment: will differences in identity come back to the surface, after having been excluded in the experiment, as agents are now socially closer? Do agents actively prefer to be part of homogeneous over heterogeneous teams? Is homogeneity of a team a benefit that firms and charities offer to their employees and extract surplus from?

3.A Tables

	Partner is not a volunteer	Partner is a volunteer	Total
Non-volunteer	113	27	140
Volunteer	27	23	50
Total	140	50	190

Table 3.1: Volunteering decision and assigned partner's status. Sample of survey respondents.

	Partner is not a volunteer	Partner is a volunteer	Total
Non-volunteer	82	20	102
Volunteer	23	13	36
Total	105	33	138

Table 3.2: Volunteering decision and assigned partner's status. Sample of lab participants.

| | Non-volunteers | | | | Volunteers | | | |
| | Partner not vol. | | Partner vol. | | Partner not vol. | | Partner vol. | |
	Mean [Std. dev.]	N.	Mean [Std. dev.]	N.	Mean [Std. dev.]	N.	Mean [Std. Dev.]	N.
Age	22.95 [4.27]	112	22.85 [3.44]	27	23.37 [4.05]	27	24.52 [5.44]	23
Female (0-1)	.6991 [.4607]	113	.6667 [.4804]	27	.7037 [.4653]	27	.6957 [.4705]	23
Heterosexual (0-1)	.8761 [.3309]	113	.8519 [.362]	27	.7778 [.4237]	27	.6522 [.487]	23
Gay/Lesbian (0-1)	.0354 [.1856]	113	.0741 [.2669]	27	.037 [.1925]	27	.087 [.2881]	23
Bisex (0-1)	.0708 [.2576]	113	.037 [.1925]	27	.0741 [.2669]	27	.1304 [.3443]	23
Asexual (0-1)	0 [0]	113	0 [0]	27	.037 [.1925]	27	.0435 [.2085]	23
Other sexual orientation (0-1)	.0088 [.0941]	113	0 [0]	27	.037 [.1925]	27	.0435 [.2085]	23
Student (0-1)	1 [0]	112	1 [0]	27	1 [0]	27	1 [0]	23
LSE student (0-1)	.75 [.435]	112	.7778 [.4237]	27	.7778 [.4237]	27	.7826 [.4217]	23
Econ student (0-1)	.1339 [.3421]	112	.1111 [.3202]	27	.037 [.1925]	27	.1739 [.3876]	23

Table 3.3: Summary statistics for the sample of survey respondents. Stars indicate significant difference between subjects partnered with a volunteer and other subjects partnered with a non-volunteer within the same volunteering/non-volunteering group. * = p-val<.10, ** = p-val <.05, *** = p-val<.01.

| | Non-volunteers | | | | Volunteers | | | |
| | Partner not vol. | | Partner vol. | | Partner not vol. | | Partner vol. | |
	Mean [Std. dev.]	N.	Mean [Std. dev.]	N.	Mean [Std. dev.]	N.	Mean [Std. Dev.]	N.
Working (0-1)	.1607 [.3689]	112	.2222 [.4237]	27	.2592 [.4466]	27	.1739 [.3876]	23
Full time work (0-1)	.0089 [.0945]	112	0 [0]	27	.037 [.1925]	27	.0435 [.2085]	23
Part time work (0-1)	.1518 [.3604]	112	.2222 [.4237]	27	.2222 [.4237]	27	.1304 [.3443]	23
Family income above median (0-1)	.2768 [.4494]	112	.2222 [.4237]	27	.3333 [.4804]	27	.3913 [.499]	23
Family income not reported (0-1)	.3571 [.4813]	112	.3704 [.4921]	27	.3704 [.4921]	27	.2174 [.4217]	23
Personal income above median (0-1)	.1161 [.3217]	112	.0741 [.2669]	27	.037 [.1925]	27	.1739 [.3876]	23
Personal income not reported (0-1)	.3036 [.4619]	112	.4074 [.5007]	27	.3704 [.4921]	27	.3043 [.4705]	23
Poor or working class (0-1)	.1071 [.3107]	112	.2222 [.4237]	27	.1481 [.362]	27	.0435 [.2085]	23
Lower middle class (0-1)	.3839 [.4885]	112	.1852* [.3958]	27	.2592 [.4466]	27	.2609 [.449]	23
Middle upper	.375 [.4863]	112	.4074 [.5007]	27	.4815 [.5092]	27	.6087 [.499]	23
Social class undisclosed (0-1)	.1339 [.3421]	112	.1852 [.3958]	27	.1111 [.3202]	27	.087 [.2881]	23

Table 3.4: Summary statistics for the sample of survey respondents. Stars indicate significant difference between subjects partnered with a volunteer and other subjects partnered with a non-volunteer within the same volunteering/non-volunteering group. * = p-val<.10, ** = p-val <.05, *** = p-val<.01.

| | Non-volunteers | | | | Volunteers | | | |
| | Partner not vol. | | Partner vol. | | Partner not vol. | | Partner vol. | |
	Mean [Std. dev.]	N.	Mean [Std. dev.]	N.	Mean [Std. dev.]	N.	Mean [Std. Dev.]	N.
EN language (0-1)	.5179 [.5019]	112	.4815 [.5092]	27	.4815 [.5092]	27	.3913 [.499]	23
IT language (0-1)	0 [0]	112	.0741* [.2669]	27	.037 [.1925]	27	.0435 [.2085]	23
Asian language (0-1)	.1071 [.3107]	112	.037 [.1925]	27	.1852 [.3958]	27	.2609 [.449]	23
Over five years learning EN (0-1)	.9821 [.133]	112	.8889* [.3202]	27	.8889 [.3202]	27	1 [0]	23
Atheist (0-1)	.3125 [.4656]	112	.2592 [.4466]	27	.2592 [.4466]	27	.5652** [.5069]	23
Christian (0-1)	.2678 [.4448]	112	.3704 [.4921]	27	.2592 [.4466]	27	.0435** [.2085]	23
Religion undisclosed (0-1)	.0893 [.2864]	112	.1852 [.3958]	27	.0741 [.2669]	27	.0435 [.2085]	23
LGBT+ rights knowledge (1-6)	3.6 [1.2]	112	3.41 [1.01]	27	3.7 [.87]	27	4.39* [.94]	23
LGBT+ rights involvement (1-6)	2.12 [1.16]	111	1.96 [1.05]	27	2.04 [1.22]	27	2.61 [1.23]	23
LGBT+ rights, care for (1-6)	4.28 [1.57]	111	4.26 [1.35]	27	5.19 [1.08]	27	5.18 [.8]	22
Interested in more volunteering (0-1)	.2743 [.4482]	113	.2222 [.4237]	27	.4444 [.5064]	27	.3478 [.487]	23

Table 3.5: Summary statistics for the sample of survey respondents. Stars indicate significant difference between subjects partnered with a volunteer and other subjects partnered with a non-volunteer within the same volunteering/non-volunteering group. * = p-val<.10, ** = p-val <.05, *** = p-val<.01.

| | Non-volunteers | | | | Volunteers | | | |
| | Partner not vol. | | Partner vol. | | Partner not vol. | | Partner vol. | |
	Mean [Std. dev.]	N.	Mean [Std. dev.]	N.	Mean [Std. dev.]	N.	Mean [Std. Dev.]	N.
Age	22.67 [3.73]	82	22.2 [3.22]	20	23.52 [3.98]	23	23.08 [4.03]	13
Female (0-1)	.7073 [.4578]	82	.55 [.5104]	20	.6522 [.487]	23	.7692 [.4385]	13
Heterosexual (0-1)	.8415 [.3675]	82	.85 [.3663]	20	.7391 [.449]	23	.6923 [.4804]	13
Gay/Lesbian (0-1)	.0488 [.2167]	82	.1 [.3078]	20	.0435 [.2085]	23	.0769 [.2773]	13
Bisex (0-1)	.0854 [.2811]	82	0 [0]	20	.087 [.2881]	23	.1538 [.3755]	13
Asexual (0-1)	0 [0]	82	0 [0]	20	.0435 [.2085]	23	0 [0]	13
Other sexual orientation (0-1)	.0122 [.1104]	82	0 [0]	20	.0435 [.2085]	23	0 [0]	13
Student (0-1)	1 [0]	82	1 [0]	20	1 [0]	23	1 [0]	13
LSE student (0-1)	.7805 [.4165]	82	.85 [.3663]	20	.7826 [.4217]	23	.8462 [.3755]	13
Econ student (0-1)	.1585 [.3675]	82	.1 [.3078]	20	.0435 [.2085]	23	.1538 [.3755]	13

Table 3.6: Summary statistics for the sample of participants in the lab session. Stars indicate significant difference between subjects partnered with a volunteer and other subjects partnered with a non-volunteer within the same volunteering/non-volunteering group. * = p-val<.10, ** = p-val <.05, *** = p-val<.01.

| | Non-volunteers | | | | Volunteers | | | |
| | Partner not vol. | | Partner vol. | | Partner not vol. | | Partner vol. | |
	Mean [Std. dev.]	N.	Mean [Std. dev.]	N.	Mean [Std. dev.]	N.	Mean [Std. Dev.]	N.
Working (0-1)	.1463 [.3556]	82	.2 [.4104]	20	.1739 [.3876]	23	.0769 [.2773]	13
Full time work (0-1)	0 [0]	82	0 [0]	20	0 [0]	23	0 [0]	13
Part time work (0-1)	.1463 [.3556]	82	.2 [.4104]	20	.1739 [.3876]	23	.0769 [.2773]	13
Family income above median (0-1)	.2805 [.452]	82	.15 [.3663]	20	.3913 [.499]	23	.3077 [.4804]	13
Family income not reported (0-1)	.3537 [.481]	82	.4 [.5026]	20	.3043 [.4705]	23	.3077 [.4804]	13
Personal income above median (0-1)	.0732 [.262]	82	.05 [.2236]	20	0 [0]	23	0 [0]	13
Personal income not reported (0-1)	.3171 [.4682]	82	.45 [.5104]	20	.4348 [.5069]	23	.4615 [.5189]	13
Poor or working class (0-1)	.0854 [.2811]	82	.2 [.4104]	20	.1304 [.3443]	23	.0769 [.2773]	13
Lower middle class (0-1)	.3902 [.4908]	82	.15** [.3663]	20	.2174 [.4217]	23	.2308 [.4385]	13
Middle upper	.378 [.4879]	82	.4 [.5026]	20	.5217 [.5108]	23	.6153 [.5064]	13
Social class undisclosed (0-1)	.1463 [.3556]	82	.25 [.4443]	20	.1304 [.3443]	23	.0769 [.2773]	13

Table 3.7: Summary statistics for the sample of participants in the lab session. Stars indicate significant difference between subjects partnered with a volunteer and other subjects partnered with a non-volunteer within the same volunteering/non-volunteering group. * = p-val<.10, ** = p-val <.05, *** = p-val<.01.

| | Non-volunteers | | | | Volunteers | | | |
| | Partner not vol. | | Partner vol. | | Partner not vol. | | Partner vol. | |
	Mean [Std. dev.]	N.	Mean [Std. dev.]	N.	Mean [Std. dev.]	N.	Mean [Std. Dev.]	N.
EN language (0-1)	.439 [.4993]	82	.5 [.513]	20	.4348 [.5069]	23	.3077 [.4804]	13
IT language (0-1)	0 [0]	82	.05** [.2236]	20	.0435 [.2085]	23	0 [0]	13
Asian language (0-1)	.1341 [.3429]	82	.05 [.2236]	20	.2174 [.4217]	23	.2308 [.4385]	13
Over five years learning EN (0-1)	.9756 [.1552]	82	.85** [.3663]	20	.8696 [.3443]	23	1 [0]	13
Atheist (0-1)	.3659 [.4846]	82	.25 [.4443]	20	.2174 [.4217]	23	.3846 [.5064]	13
Christian (0-1)	.2195 [.4165]	82	.4* [.5026]	20	.3043 [.4705]	23	.0769 [.2773]	13
Religion undisclosed (0-1)	.0854 [.2811]	82	.15 [.3663]	20	.087 [.2881]	23	.0769 [.2773]	13
LGBT+ rights knowledge (1-6)	3.5 [1.19]	82	3.45 [1.1]	20	3.74 [.81]	23	4 [.82]	13
LGBT+ rights involvement (1-6)	2.04 [1.2]	82	2 [1.08]	20	2 [1.21]	23	2.38 [1.19]	13
LGBT+ rights, care for (1-6)	4.17 [1.66]	82	4.25 [1.45]	20	5.26 [1.05]	23	5.15 [.8]	13
Interested in more volunteering (0-1)	.2805 [.452]	82	.3 [.4702]	20	.4348 [.5069]	23	.3846 [.5064]	13

Table 3.8: Summary statistics for the sample of participants in the lab session. Stars indicate significant difference between subjects partnered with a volunteer and other subjects partnered with a non-volunteer within the same volunteering/non-volunteering group. * = p-val<.10, ** = p-val <.05, *** = p-val<.01.

| | Non-volunteers | | | | Volunteers | | | |
| | Partner not vol. | | Partner vol. | | Partner not vol. | | Partner vol. | |
	Mean [Std. dev.]	N.	Mean [Std. dev.]	N.	Mean [Std. dev.]	N.	Mean [Std. Dev.]	N.
Signed up (0-1)	.823 [.3834]	113	.8148 [.3958]	27	.8889 [.3202]	27	.7391 [.449]	23
Attended (0-1)	.7257 [.4482]	113	.7407 [.4466]	27	.8519 [.362]	27	.5652** [.5069]	23
Later than 5 min or not attended	.2832 [.4526]	113	.2592 [.4466]	27	.1852 [.3958]	27	.4348* [.5069]	23
Number of envelopes (0 if not attended)	12.62 [9.22]	113	15.3 [11.91]	27	14.67 [7.04]	27	11.83 [12.55]	23
Time spent, min (0 if not attended)	17.06 [12.04]	113	19.41 [13.3]	27	20.74 [10.7]	27	15.7 [15.93]	23

Table 3.9: Summary statistics for the sample of survey respondents. Stars indicate significant difference between subjects partnered with a volunteer and other subjects partnered with a non-volunteer within the same volunteering/non-volunteering group. * = p-val<.10, ** = p-val <.05, *** = p-val<.01.

| | Non-volunteers | | | | Volunteers | | | |
| | Partner not vol. | | Partner vol. | | Partner not vol. | | Partner vol. | |
	Mean [Std. dev.]	N.	Mean [Std. dev.]	N.	Mean [Std. dev.]	N.	Mean [Std. Dev.]	N.
N. other people attending session	2.6 [.86]	82	2.3 [.73]	20	3.26 [.69]	23	3.38 [.65]	13
Later than 5 min or not attended	.0122 [.1104]	82	0 [0]	20	.0435 [.2085]	23	0 [0]	13
Envelopes	17.39 [5.79]	82	20.65** [8.81]	20	17.22 [3.55]	23	20.92* [9.1]	13
Envelopes (log)	2.81 [.28]	82	2.95* [.39]	20	2.83 [.21]	23	2.97 [.38]	13
Envelopes per min	.7584 [.1793]	82	.7923 [.232]	20	.7447 [.2035]	23	.7504 [.1712]	13
Envelopes per min (log)	-.3 [.24]	82	-.27 [.3]	20	-.33 [.25]	23	-.31 [.22]	13
Number of correctly filled envelopes	17.29 [5.34]	82	19.25 [9.56]	20	16.3 [4.96]	23	20.92* [9.1]	13
Number of correctly filled envelopes (log)	2.81 [.27]	82	2.93 [.38]	19	2.82 [.21]	22	2.97 [.38]	13
Some error (0-1)	.0244 [.1552]	82	.15** [.3663]	20	.0435 [.2085]	23	0 [0]	13
Time spent (min)	23.51 [6.84]	82	26.2 [7.39]	20	24.35 [6.6]	23	27.77 [10.1]	13
Time spent (log)	3.12 [.29]	82	3.23 [.29]	20	3.15 [.31]	23	3.28 [.3]	13

Table 3.10: Summary statistics for the sample of participants in the lab. Stars indicate significant difference between subjects partnered with a volunteer and other subjects partnered with a non-volunteer within the same volunteering/non-volunteering group. * = p-val<.10, ** = p-val <.05, *** = p-val<.01.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	m1-m4 Prob. attending	m1 Envelopes (latent var)	m1 Envelopes (observed)	m2 Log envelopes (latent var)	m2 Log envelopes (observed)	m3 Time spent	m4 Time spent (log)
Volunteer	-.01 (.07)	0.75 (1.41)	0.45 (1.10)	0.039 (0.060)	.026 (.046)	0.91 (1.65)	0.042 (.073)
Partner is a volunteer	-0.01 (.09)	4.13* (2.16)	3.35*** (1.79)	0.158* (0.085)	.134* (.069)	1.55 (1.73)	0.054 (.068)
Volunteer X Partner is a volunteer	-.25 (.17)	-0.61 (3.60)	-.45 (3.03)	-0.032 (0.139)	-.028 (.111)	2.45 (3.65)	0.098 (.131)
Interested in more volunteering	0.09 (.07)	3.75*** (1.43)	2.90*** (1.10)	0.150*** (0.057)	.125*** (.049)	2.83** (1.36)	0.124** (.055)
Care for LGBT+ rights (std dev)	0.02 (.04)	-0.03 (.66)	-.03 (.51)	-0.003 (0.029)	-.002 (.024)	0.89 (.63)	0.040 (.026)
Knowledge LGBT+ rights (std dev)	-0.05 (.05)	1.25 (.79)	.97 (.60)	0.056 (0.034)	.046 (.028)	0.67 (.79)	0.020 (.030)
Male	0.10 (.07)	0.70 (1.42)	.54 (1.10)	0.001 (0.054)	.001 (.045)	3.79*** (1.270)	.151*** (.050)
Straight	-0.13 (.09)	2.46 (1.60)	1.90 (1.22)	0.089 (0.064)	.075 (.053)	3.050 (1.60)	0.127* (.072)
Personal income above median	-0.32*** (.11)	3.30 (3.47)	2.55 (2.69)	0.109 (0.128)	.091 (.106)	-.225 (3.717)	-.044 (.142)
Christian	-0.05 (.09)	-1.27 (1.46)	-.98 (1.13)	-0.050 (0.062)	-.042 (.051)	-0.24 (1.42)	0.010 (.059)
Not religious	-0.00 (.08)	1.89 (1.53)	1.46 (1.17)	0.073 (0.060)	.061 (.050)	0.17 (1.37)	0.027 (.053)
Large envelopes		1.16 (1.18)	.90 (.91)	0.050 (0.048)	.042 (.040)	0.52 (1.15)	0.020 (.047)
N. other participants		-0.54 (.75)	-.41 (.59)	-0.024 (0.030)	-.020 (.025)	-1.46** (.68)	-0.075** (.030)

Table 3.11: Maximum likelihood estimates of hurdle models v1-v4, that separately model the decision to attend the experiment and behaviour in the lab. Robust standard errors in parentheses. Column (1) reports the marginal effects of each variable in the selection part of the model. The estimates were the same, up to the second decimal, for all four models. Models v1 and v2 additionally control for the censoring of the number of envelopes at the lower bound of 12 and the upper bound of 40. For each of those models, one column reports the coefficients for the latent variable and the other column reports marginal effects for the observed variable among lab participants. Marginal effects are evaluated at the sample mean of the control variables. In columns (1), (2) and (4), the marginal effects for being treated with a volunteer partner is the relevant marginal effect among donors at the unconditional sample mean. The interaction term "Volunteer X Partner is a Volunteer" reports the estimate of the difference between the marginal effect reported above and the corresponding effect for non-volunteers, each evaluated at the unconditional sample mean of other observables. Models v3 and v4 report the coefficients for the expected value of time spent in the lab in a hurdle model.

	Envelopes		Envelopes (log)		Error (0-1)		Correct envelopes		Time spent (min)		Time spent (log min)		Productivity (envelopes/min)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(11)	(12)	(13)	(14)	(15)	(16)
Volunteer	-0.173 (1.506)	-0.230 (1.621)	0.013 (0.071)	0.018 (0.076)	0.019 (0.048)	-0.004 (0.052)	-0.988 (1.523)	-0.851 (1.672)	0.836 (1.707)	0.812 (1.828)	0.035 (0.069)	0.036 (0.073)	-0.014 (0.045)	-0.018 (0.048)
Partner is a volunteer	3.260** (1.592)	3.222** (1.613)	0.139* (0.075)	0.136* (0.076)	0.126** (0.050)	0.132** (0.052)	1.957 (1.610)	1.826 (1.664)	2.688 (1.804)	1.564 (1.819)	0.110 (0.073)	0.056 (0.073)	0.034 (0.048)	0.074 (0.048)
Volunteer X Partner is a volunteer	0.446 (2.727)	0.651 (2.742)	0.003 (0.128)	0.014 (0.129)	-0.169* (0.086)	-0.171* (0.088)	2.661 (2.758)	3.031 (2.829)	0.734 (3.091)	2.483 (3.092)	0.015 (0.125)	0.097 (0.124)	-0.028 (0.082)	-0.085 (0.082)
Other controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
N	138	138	138	138	138	138	138	138	138	138	138	138	138	138

Table 3.12: OLS Results in the sample of lab participants. * = p-val<.10, ** = p-val <.05, *** = p-val<.01. Controls include gender, care for LGBT rights and knowledge of them, personal income above the median, heterosexual sexual orientation, number of other people attending the same session.

	(1) Env.	(2) Env. (log)	(3) Error (0-1)	(4) Correct env.	(5) Time (min)	(6) Time (log min)	(7) Prod. (env/min)
Volunteer	-0.387 (1.443)	0.002 (0.068)	0.012 (0.045)	-1.205 (1.504)	0.650 (1.700)	0.026 (0.069)	-0.015 (0.045)
Partner volunteer	0.991 (1.693)	0.042 (0.080)	0.052 (0.053)	0.501 (1.765)	1.255 (1.995)	0.065 (0.081)	-0.005 (0.053)
Volunteer X Partner volunteer	-0.089 (2.603)	-0.019 (0.123)	-0.187** (0.082)	2.357 (2.713)	0.420 (3.067)	0.007 (0.124)	-0.039 (0.081)
Interest in more volunteering	1.390 (1.286)	0.075 (0.061)	0.045 (0.040)	1.406 (1.340)	1.205 (1.515)	0.061 (0.061)	0.010 (0.040)
Interest in more volunteering X Partner volunteer	7.472*** (2.588)	0.319** (0.122)	0.243*** (0.082)	4.762* (2.697)	4.698 (3.049)	0.145 (0.124)	0.130 (0.081)
Constant	17.000*** (0.760)	2.791*** (0.036)	0.012 (0.024)	16.898*** (0.792)	23.174*** (0.895)	3.100*** (0.036)	0.755*** (0.024)
N	138	138	138	138	138	138	138

Table 3.13: OLS coefficients from sample of participants. No additional controls. * = p-val<.10, ** = p-val <.05, *** = p-val<.01.

	(1) Env.	(2) Env. (log)	(3) Error (0-1)	(4) Correct env.	(5) Time (min)	(6) Time (log min)	(7) Prod. (env/min)
Volunteer	0.039 (1.533)	0.027 (0.072)	0.032 (0.049)	-0.591 (1.556)	0.676 (1.773)	0.031 (0.071)	-0.001 (0.047)
Partner volunteer	4.056** (1.585)	0.180** (0.074)	0.149*** (0.050)	2.722* (1.609)	3.095* (1.833)	0.129* (0.074)	0.051 (0.048)
Volunteer X Partner volunteer	-2.032 (2.824)	-0.126 (0.132)	-0.244*** (0.090)	0.246 (2.867)	-0.487 (3.267)	-0.044 (0.132)	-0.082 (0.086)
LGBT+ rights, care for (st. d.)	-0.299 (0.591)	-0.019 (0.028)	-0.018 (0.019)	-0.560 (0.600)	0.225 (0.684)	0.005 (0.028)	-0.018 (0.018)
LGBT+ rights, care for (st. d.) X Partner volunteer	3.756*** (1.384)	0.195*** (0.065)	0.113** (0.044)	3.640** (1.405)	1.882 (1.601)	0.091 (0.065)	0.080* (0.042)
Constant	17.333*** (0.699)	2.809*** (0.033)	0.021 (0.022)	17.186*** (0.710)	23.555*** (0.809)	3.118*** (0.033)	0.755*** (0.021)
N	138	138	138	138	138	138	138

Table 3.14: OLS coefficients from sample of respondents. No additional controls. * = p-val<.10, ** = p-val <.05, *** = p-val<.01.

	(1) Env.	(2) Env. (log)	(3) Error (0-1)	(4) Correct env.	(5) Time (min)	(6) Time (log min)	(7) Prod. (env/min)
Volunteer	0.003	0.023	0.016	-0.718	0.940	0.041	-0.012
	(1.520)	(0.071)	(0.048)	(1.527)	(1.725)	(0.070)	(0.046)
Partner volunteer	3.962	0.216	0.190*	2.672	1.684	0.124	0.073
	(3.163)	(0.149)	(0.100)	(3.177)	(3.590)	(0.145)	(0.095)
Volunteer X Partner volunteer	0.408	-0.007	-0.183**	2.671	0.975	0.016	-0.034
	(2.775)	(0.131)	(0.088)	(2.788)	(3.150)	(0.127)	(0.083)
Heterosexual	1.716	0.095	-0.027	2.643	1.024	0.055	0.020
	(1.633)	(0.077)	(0.052)	(1.640)	(1.853)	(0.075)	(0.049)
Heterosexual X Partner volunteer	-0.843	-0.091	-0.076	-0.868	1.171	-0.017	-0.046
	(3.220)	(0.151)	(0.102)	(3.235)	(3.655)	(0.148)	(0.097)
Constant	15.947***	2.732***	0.047	15.069***	22.650***	3.070***	0.741***
	(1.545)	(0.073)	(0.049)	(1.552)	(1.754)	(0.071)	(0.046)
N	138	138	138	138	138	138	138

Table 3.15: OLS coefficients from sample of respondents. No additional controls. * = p-val<.10, ** = p-val <.05, *** = p-val<.01.

| | Non-volunteers | | | | Volunteers | | | |
| | Partner not vol. | | Partner vol. | | Partner not vol. | | Partner vol. | |
	Mean [Std. dev.]	N.	Mean [Std. dev.]	N.	Mean [Std. dev.]	N.	Mean [Std. Dev.]	N.
Correctly recalls partner type	.9836 [.128]	61	.8667** [.3519]	15	.875 [.3416]	16	.9091 [.3015]	11
Affected by partner (NA missing)	.3667 [.486]	60	.5333 [.5164]	15	.5625 [.5123]	16	.6364 [.5044]	11
Encouraged by partner (NA missing)	.3333 [.4851]	18	.8889* [.3333]	9	.2222 [.441]	9	.8571* [.378]	7
Guess on partner's envelopes	14.6 [3.8]	81	16.35* [3.6]	20	14.17 [2.44]	23	18** [5.96]	13
Guess on volunteers' envelopes	16.61 [5.09]	82	16.68 [4.38]	20	15.87 [2.47]	23	18.15* [5.46]	13
Guess on non-volunteers' envelopes	13.98 [2.89]	82	15.32* [3.13]	20	13.26 [2.03]	23	14.15 [2.12]	13
Guess on volunteer effect	2.63 [5.06]	82	1.36 [5.8]	20	2.61 [2.41]	23	4 [4.67]	13
Expects positive volunteer effect	.6585 [.4771]	82	.65 [.4894]	20	.7826 [.4217]	23	.7692 [.4385]	13
Expects negative volunteer effect	.1585 [.3675]	82	.3 [.4702]	20	.0435 [.2085]	23	.2308* [.4385]	13
Believes partner = average of partner's group	.4691 [.5022]	81	.2** [.4104]	20	.5217 [.5108]	23	.4615 [.5189]	13
Belief consistent with higher order	.1605 [.3694]	81	.65* [.4894]	20	.3913 [.499]	23	.0769** [.2773]	13
Believes to do more than own group	.9634 [.1889]	82	.8* [.4104]	20	.9565 [.2085]	23	.9231 [.2773]	13

Table 3.16: Summary statistics for the sample of those attending the lab session. Stars indicate significant difference between subjects partneres with a volunteer and other subjects partnered with a non-volunteer within the same volunteering/non-volunteering group. * = p-val<.10, ** = p-val <.05, *** = p-val<.01.

	(1) Envelopes	(2) Envelopes	(3) Envelopes	(4) Envelopes	(5) Envelopes
Volunteer	-0.230 (1.538)	0.245 (1.226)	0.647 (1.168)	-3.298 (3.825)	2.468 (6.633)
Partner Vol.	3.193** (1.597)	1.426 (1.288)	1.612 (1.228)	1.550 (1.294)	1.570 (1.196)
Partner Vol. X Vol.	0.222 (2.708)	-1.707 (2.168)	-1.596 (2.084)	-2.477 (2.307)	-2.066 (2.171)
Guess on partner		1.013*** (0.115)	0.627*** (0.162)	0.942*** (0.136)	0.470** (0.183)
Guess on partner X Volunteer				0.248 (0.254)	-0.026 (0.486)
Guess on non-volunteers			0.372* (0.210)		0.614*** (0.225)
Guess on non-volunteers X Volunteer					-1.132* (0.585)
Guess on volunteers			0.381*** (0.099)		0.310*** (0.101)
Guess on volunteers X Volunteer					0.864** (0.417)
Constant	17.457*** (0.711)	2.663 (1.771)	-3.267 (2.423)	3.704* (2.067)	-3.178 (2.572)
N	137	137	137	137	137

Table 3.17: Mediation analysis regressions. No controls added. Sample of participants in the lab who answered all questions on expectations. * = p-val<.10, ** = p-val <.05, *** = p-val<.01.

3.B   Figures

Study Name	170041 - Charitable activities for an LGBTI cause - part 1
Study Type	**Online Study** This study is an online study on another website. To participate, sign up, and then you will be given access to the website to participate in the study
Pay	£2 + Lottery--see description
Duration	5 minutes
Description	This is an online+lab project, and this is the study listing for part 1 - the online part. The online part of the project is a short questionnaire taking about five minutes, comprising a few questions and a test to check the comprehension of the description of the lab-based part of the project. All participants that complete the online questionnaire will receive a £2 Amazon voucher (emailed to the email address associated with your Sona account after the participation deadline) and be entered in a lottery draw for a £250 Amazon voucher. After participants complete the online part of the project, they will receive instructions on how to signup for the lab-based part of the project. The lab-based part of the project requires participants to perform a real charitable team-task in support of an LGBT+ charity. Participation in the lab will take approximately 30 minutes and is paid up to £7 (depending on choices made in the online survey), with an option for participants to make a donation to the LGBT+ charity.
Website	☑ View Study Website
Researchers	BRL . ✉
	Roberto Sormani ✉

Figure 3.1: Listing of the study as it appeared to eligible students who were invited.

Hello, the goal of this survey is to collect some general information and to prepare you for the lab project on charitable activities, where you will be asked to perform a team task for a charitable LGBT organisation.

You are welcome to respond to the survey irrespective of whether you want to participate in the project.

By participating in the survey you will earn **£2** and automatically enter the draw of **a £250 Amazon voucher**.

Please enter your lab ID here

I give my consent to participating in this questionnaire and to the use of these data for research purposes in accordance to the Data Protection Law.
I understand that all the information I am providing here is anonymous.

☐ Consent (*)

\>\>

Figure 3.2: The first screen subjects see in the online survey.

This section is about some personal information.

Answers to these questions are optional and anonymous, and will not affect your participation in the experiment.

What gender do you identify with?

◯ Female

◯ Male

◯ Other (please specify if you want)

◯ Rather not say

What is your sexual orientation?

◯ Asexual

◯ Heterosexual

◯ Gay/Lesbian

◯ Bisexual

◯ Other (please specify if you want)

◯ Rather not say

The lab project is about helping out a charity that promotes LGBT+ rights. We would thus like to ask you your own self-rating on the following respects.
The answers are **yours only** and will not affect your selection into the study.

How much would you rate the following, on a 0-5 scale, where 0 means "Not at all" and 5 means "Totally"?

How much do you care about LGBT+ rights?

How much do you know about LGBT+ rights?

How much are you actively involved with LGBT+ rights?

Figure 3.3: The second screen subjects see in the online survey.

INSTRUCTIONS AND COMPREHENSION CHECK

This section is about the lab project. You will read a description of the project and be asked to answer a comprehension check. If you answer the comprehension check correctly you will be eligible to participate in the lab project and select a slot.

- You will fill envelopes for **a real fundraising campaign** in support of an Italian LGBT charity.

- You will be put in a **team of 2** with another participant, who will be in a separate room.

- You will copy addresses **from top to bottom** on envelopes and put a letter inside each envelope.

- You and your partner will have **the same** list of addresses.

- One member of your team will have large envelopes and the other will have small envelopes.

- You and your partner will have to **fill at least 12 envelopes**.

- At a later stage, **we** will put the small envelopes into the large envelopes.

- So we will send out each envelope that you fill, **only if your partner fills their envelope and enters the same address**.

- We will then **throw away** those extra envelopes that don't find a match.

- Your **team output** will then be the lowest number between the number of envelopes that you fill and the number of envelopes that your partner fills.

- You may leave the room at any time after you've filled 12 envelopes.

- You can still choose to fill more envelopes, that will be used only if your partner fills some extra-envelopes too.

How many letters will be sent out

if you fill 15 envelopes and your partner fills 19 envelopes

if you fill 19 envelopes and your partner fills 15 envelopes

if your partner fills the minimum required amount of envelopes

Do you have any more questions about the lab project? We will try and address these questions as soon as possible.

>>

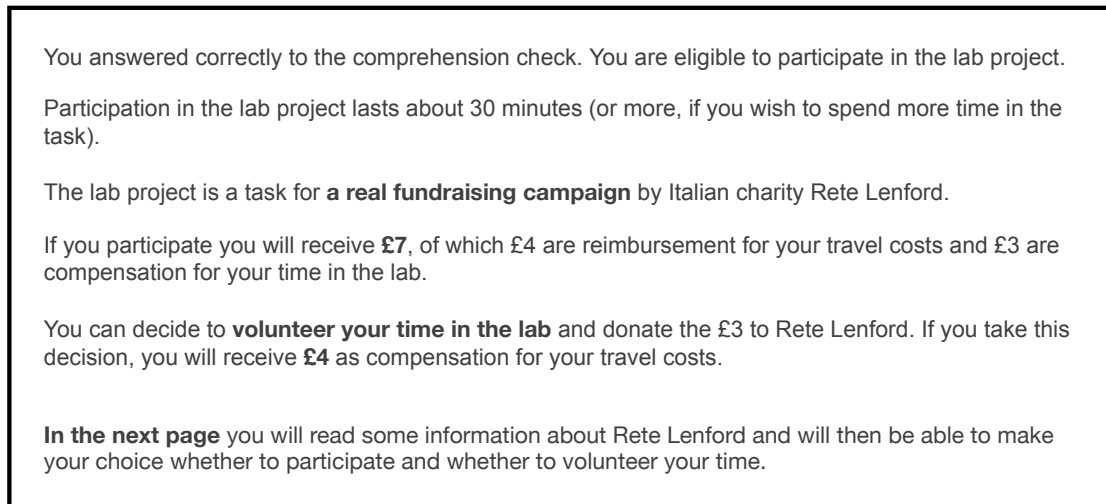Figure 3.4: The third screen subjects see in the online survey.

You answered correctly to the comprehension check. You are eligible to participate in the lab project.

Participation in the lab project lasts about 30 minutes (or more, if you wish to spend more time in the task).

The lab project is a task for **a real fundraising campaign** by Italian charity Rete Lenford.

If you participate you will receive **£7**, of which £4 are reimbursement for your travel costs and £3 are compensation for your time in the lab.

You can decide to **volunteer your time in the lab** and donate the £3 to Rete Lenford. If you take this decision, you will receive **£4** as compensation for your travel costs.

In the next page you will read some information about Rete Lenford and will then be able to make your choice whether to participate and whether to volunteer your time.

>>

Figure 3.5: The fourth screen subjects see in the online survey.

Rete Lenford is an Italian charity that provides **free legal assistance to LGBTI people** and initiates legal lawsuits for LGBTI rights. Among other things, they have appealed to the Constitutional Court asking for equal marriage rights of same-sex couples.

Rete Lenford is a network of **professional lawyers** who give their expertise for free to support and improve the lives of hundreds of LGBT people each year.

They support people in same-sex relationships who are not recognised as parents. They give free legal assistance in discrimination lawsuits. They promote the protection of children of same-sex couples.

in Italy there is no marriage equality; same-sex couples cannot adopt children (nor can single persons); a proposal to criminalise homophobic hate crime has been rejected under the attacks of the conservative movements, and intersex children often undergo gender reassignment at birth (also known as Intersex Genital Mutilation), despite the condemnation of this practice by the UN.

In the next page you will choose whether you participate and whether you want to volunteer your time.
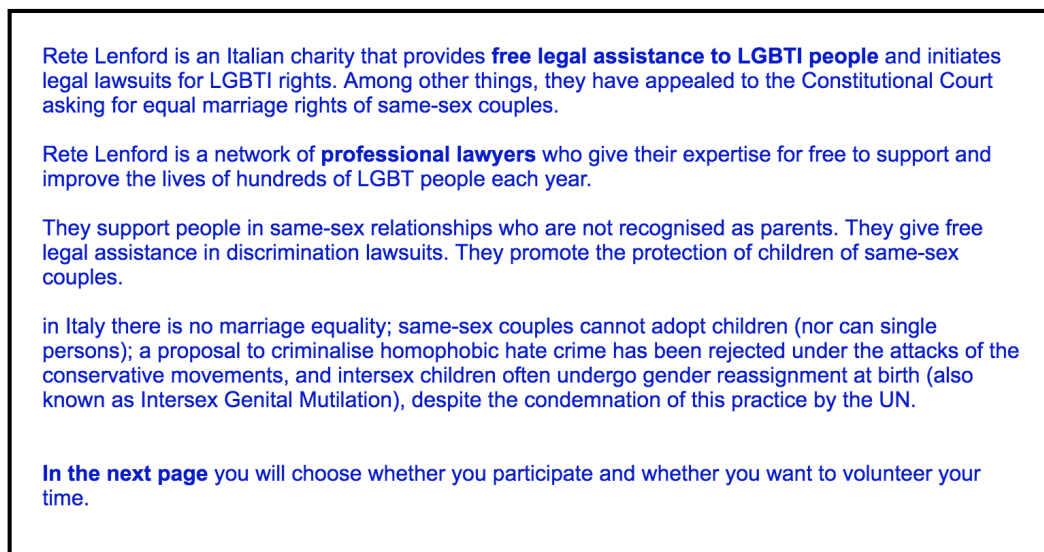
>>

Figure 3.6: The fifth screen subjects see in the online survey.

You can now decide whether you want to participate and, if so, whether you want to donate part of your compensation to Rete Lenford and support them in their effort to provide free legal assistance to LGBTI people.

If you choose to participate you will be assigned a partner. You will not personally meet your partner. However, if you choose to participate you are going to read some information about your partner in the next page.

○ I want to participate in the lab project and volunteer my time by donating £3 to Rete Lenford. I will then receive **£4**

○ I do not want to participate in the lab project

○ I want to participate in the lab project without volunteering my time. I will then receive **£7**

>>

Figure 3.7: The sixth screen subjects see in the online survey. The ordering of answers was randomised.

Thanks for choosing to participate.

You have now been matched with another participant in this study.

Although **you will not personally meet** your partner, we can provide you some information.

Your partner is **a student** who has participated in **this online survey** and **has chosen not to donate** the £3 to Rete Lenford.

In the next page you will read important information on how to reserve your slot in the lab.

Please do sign up and attend your session: since you have been paired up with a partner, if you or your partner do not show up in the lab, the whole team output will not be sent out.

>>

Figure 3.8: The seventh screen subjects see in the online survey, if they are matched with a non-volunteer.

Thanks for choosing to participate.

You have now been matched with another participant in this study.

Although **you will not personally meet** your partner, we can provide you some information.

Your partner is **a student** who has participated in **this online survey** and **has chosen to donate** the £3 to Rete Lenford.

In the next page you will read important information on how to reserve your slot in the lab.

Please do sign up and attend your session: since you have been paired up with a partner, if you or your partner do not show up in the lab, the whole team output will not be sent out.

Figure 3.9: The seventh screen subjects see in the online survey, if they are matched with a volunteer

Important information: please read carefully

Please press >> and sign up for a slot in the lab by registering for **part 2a** of this study on Sona.

You will need to enter the following code:

0NlVkMf8

Please store this code by <u>downloading</u> this information, and enter your email address if you would like to receive this information via email.

The code shall not be circulated.

Please **press Next** to answer one last question and go to the end of the survey.

Figure 3.10: The eigth screen subjects see in the online survey. The code given gives access to different listings for the lab session, conditional on whether the respondent has decided to volunteer or not. This was designed to ensure that non-volunteers, which were a smaller group, would not have bear a higher congestion cost in signing up, compared to non-volunteers.

After the lab project, we will arrange the matching of large and small envelopes, so that everything is ready for delivery.

We will recruit volunteers that will put the small envelopes into the large envelopes.

Would you like to register your interest to volunteer for that stage? If so, we will be in touch again after the lab project.

○ Yes, I'm interested

○ No, I'm not interested

Press **Next** to get your credits and sign up for the study.

>>

Figure 3.11: The ninth (and last) screen subjects see in the online survey. Pressing the forward button would automatically redirect to the page to sign up.

CONSENT FORM

We would like to invite you to take part in this research study, which examines involvement in collective action and LGBT+ rights. The Principal Investigator of this study is Roberto Claudio Sormani, PhD candidate at the Economics Department at the LSE.

If you agree to take part in this research, you will be asked to go to one of the meeting rooms in the lab, keep the door closed, fill **at least twelve envelopes** and write addresses on them, reading instructions carefully.

There are no risks to you from this research and there is a **direct monetary benefit of £4 as reimbursement for travel costs and £3 for your lab participation, that you might have donated to Rete Lenford depending on whether you accepted to volunteer your time in the lab**.
Any information from this study that is published or presented at scientific meetings will be completely anonymous.
Your participation in this research is voluntary. Whether or not you choose to take part in this research will have no bearing on your standing or academic grades.
It is very important to the significance of this project that information about the design is not shared with to future participants. For this reason you are asked to **respect the confidentiality** of this project and **not disclose information** on the setup and the content of this project.

Please write down your answer to the following questions and give your consent.

Question:
Suppose there is a team whereby one participant fills 13 envelopes and the other participants fills 17 envelopes. How many envelopes are going to be sent out?
_____

I have read the above document and give my consent to participation in the project.

_____
_____
NAME (PRINT)   Signature

Figure 3.12: The consent form that participants were asked to fill as they entered the lab.

Team task - instructions

- You will fill envelopes for **a real fundraising campaign** in support of an Italian LGBT charity (see box).

- You will be put in a **team of 2** with another participant, who will be in a separate room and might be attending this or another session of the study.

- You will copy addresses **from top to bottom** on envelopes and put a letter inside each envelope.

- You and your partner will have **the same** list of addresses.

- One member of your team will have large envelopes and the other will have small envelopes.

- You and your partner will have to **fill at least 12 envelopes**.

- At a later stage, **we** will put the small envelopes into the large envelopes.

- So we will send out each envelope that you fill, **only if your partner fills their envelope and enters the same address**.

- We will then **throw away** those extra envelopes that don't find a match.

- Your **team output** will then be the lowest number between the number of envelopes that you fill and the number of envelopes that your partner fills.

- You may leave the room at any time after you've filled 12 envelopes.

- You can still choose to fill more envelopes, that will be used only if your partner fills some extra-envelopes too.

- Remember: fill the envelopes and write the addresses on the envelopes.

- Due to technical reasons, we reassigned the teams, so the partner that you were matched with in the online part is not the same as the partner that is being assigned to you now. Your new partner is another person who participates in the lab part of the study, in this or another session.

- **Inside the room you will find some information about your partner**

What is Rete Lenford, and why care about Italian LGBTI rights?

Rete Lenford is an Italian charity that provides free legal assistance to LGBTI people and initiates legal lawsuits for LGBTI rights. Among other things, they have appealed to the Constitutional Court asking for equal marriage rights of same-sex couples.

in Italy there is no marriage equality; same-sex couples cannot adopt children (nor can single persons); a proposal to criminalise homophobic hate crime has been rejected under the attacks of the conservative movements, and intersex children often undergo gender reassignment at birth (also known as Intersex Genital Mutilation), despite the condemnation of this practice by the UN.

Example

If one of you fills 16 large envelopes writing the first 16 addresses on them, and the other one fills 13 small envelopes entering the first 13 addresses, the first 13 envelopes will match and we will send them out.

We will throw away the remaining 3 envelopes.

Figure 3.13: The sheet that participants would find in the room where the experiment took place. The instructions contained in the sheet were also read out loud by the Research Assistant as participants entered the lab.

Buongiorno,

Avvocatura per i Diritti LGBTI - Rete Lenford è un'associazione no-profit che dal 2007 assicura **tutela giudiziaria e assistenza legale** alle persone lesbiche, gay, bisessuali, transgender e intersessuali (LGBTI).

Ogni giorno offriamo la nostra professionalità **gratuitamente** per contrastare le discriminazioni ai danni delle persone LGBTI e delle loro famiglie, garantendo l'effettiva uguaglianza di diritti. Con il tuo **5x1000** o una **donazione online** puoi sostenere la nostra missione!

Finora abbiamo garantito migliaia di consulenze e **assistenza gratuita** a oltre 500 coppie dello stesso sesso che si sono sposate all'estero per esercitare un loro diritto fondamentale. Abbiamo lottato per il **matrimonio egualitario** presso la Corte Costituzionale. Abbiamo **ottenuto la trascrizione dei matrimoni *same-sex* contratti all'estero** in comuni come Roma, Milano e Napoli. Abbiamo **gestito corsi di formazione** per operatori giuridici e forze di polizia, perché lo Stato potesse **tutelare e rispettare** le persone LGBTI ogni giorno.

Per continuare la nostra azione abbiamo bisogno del vostro contributo.

Aiutaci a sostenere chi ha bisogno di noi!

Come puoi sostenere *Avvocatura per i Diritti LGBTI*

- Con il tuo **5x1000**, inserendo il codice **06006020488**
- Con un **bonifico** ad Avvocatura per i Diritti LGBTI,
  BANCA: UNICREDIT
  CONTO CORRENTE n. 000103570496
  IBAN IT63J0200811102000103570496
- Con una **donazione online o PayPal** sul sito www.retelenford.it

---

Figure 3.14: The letter that participants in charge of filling the large envelopes were asked to fold.

QUESTIONNAIRE

Lab ID number _____

Question 1: if you make a correct guess for this question, your chances of winning the £250 lottery will double.

How many envelopes has your partner filled? _____

Question 2: if you make a correct guess for this question (up to a 0.5 prediction error), your chances of winning the £250 lottery will double.

**Consider the participants who donated the £3 to the charity for this study. How many envelopes have they filled on average? _____**

Question 3: if you make a correct guess for this question (up to a 0.5 prediction error), your chances of winning the £250 lottery will double.

Consider the participants who did not donate the £3 to the charity for this study. How many envelopes have they filled on average? _____

Question 4: Has your partner donated the £3? (Yes/No) _____

Question 5: Did the information on your partner affect you? If so, how?

_____

Figure 3.15: Questionnaire that participants had to fill after the lab task and before receiving compensation.
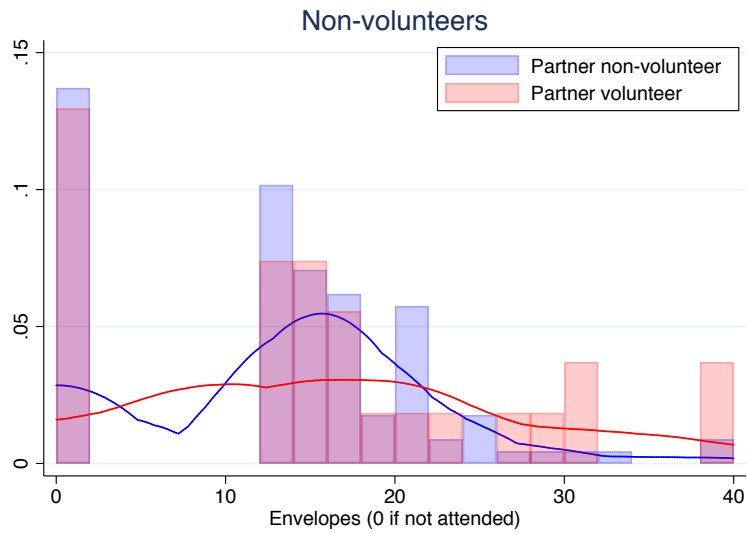
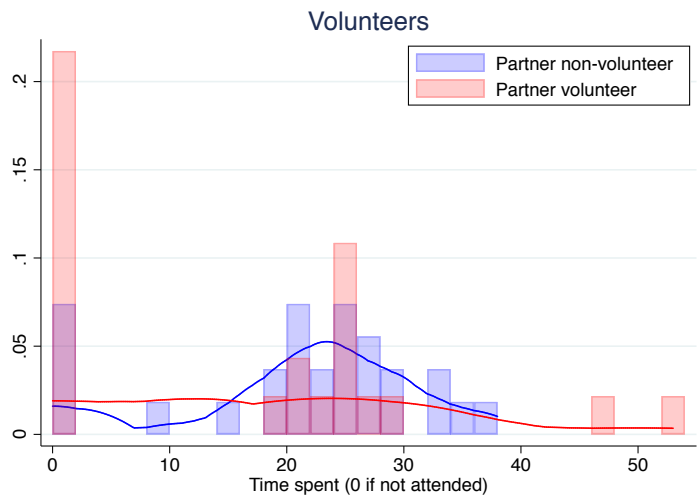(a) Envelopes filled by non-volunteers, by partner type
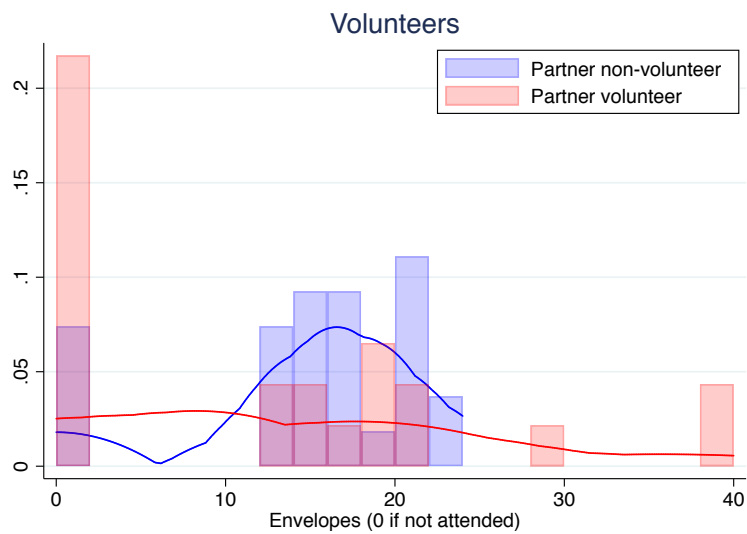
(b) Envelopes filled by volunteers, by partner type

Figure 3.16: Envelopes filled (0 if not attending), by volunteering decision and partner type, among all survey respondents. Histograms and kernel density estimates.
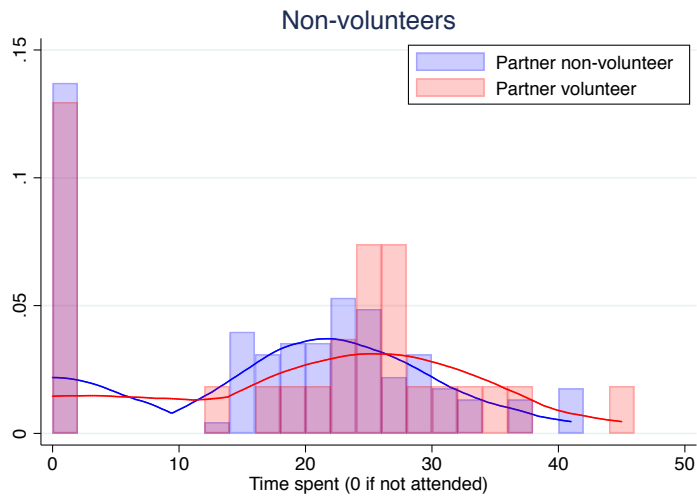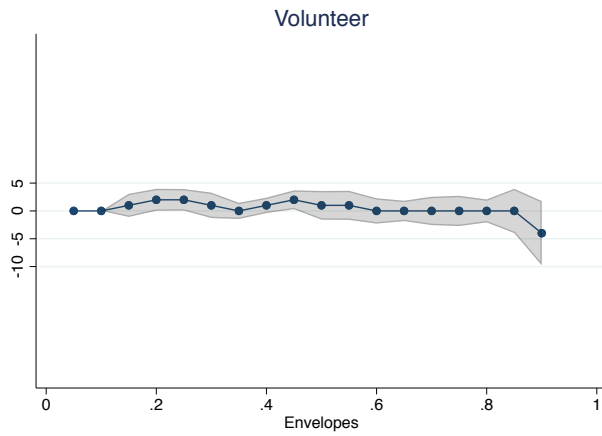
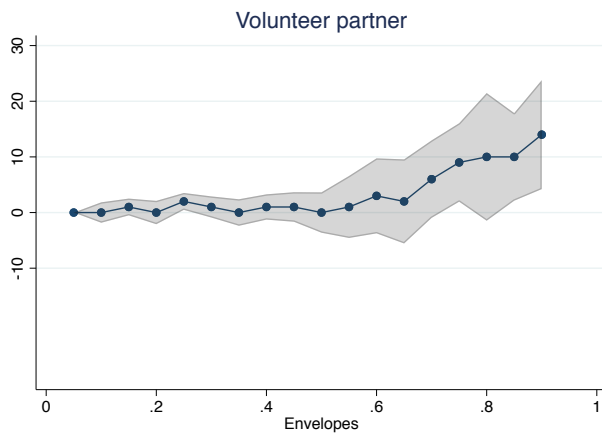(a) Time spent in the task by non-volunteers, by partner type

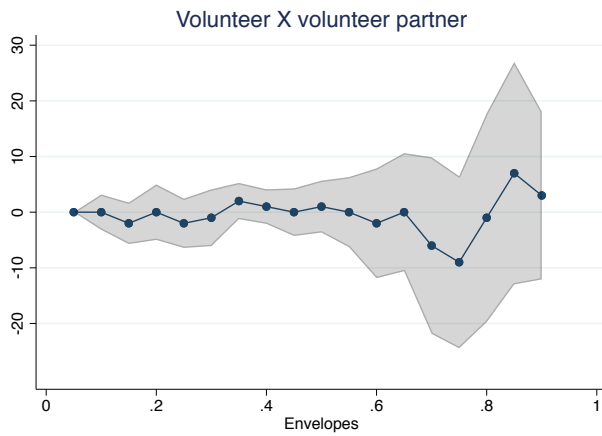(b) Time spent in the task by volunteers, by partner type

Figure 3.17: Time spent in the task, by volunteering decision and partner type (0 if not attending). Histograms and kernel density estimates.

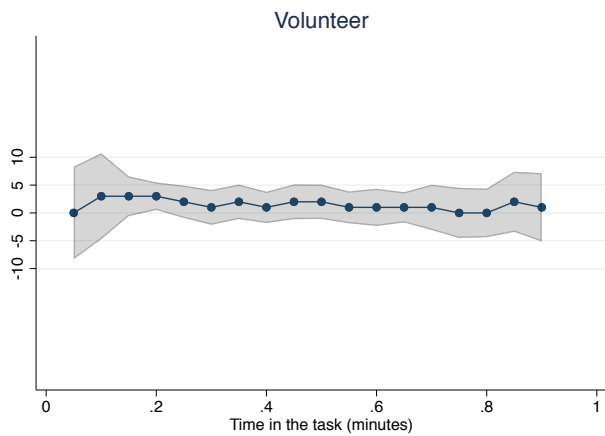(a) Quantile effect of being a volunteer, by quantile.

(b) Quantile treatment effect of being assigned a volunteer partner, by quantile.
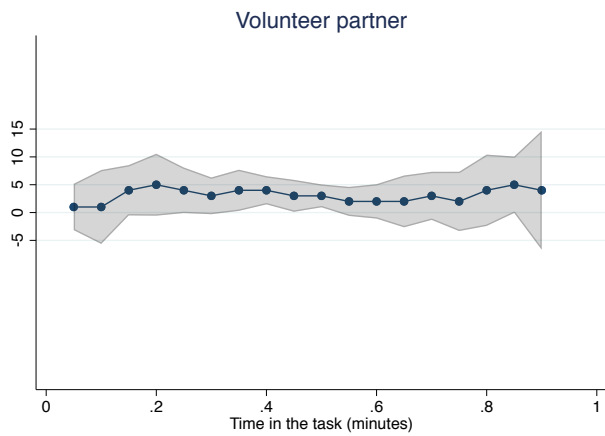

(c) Quantile interaction effect of being assigned a volunteer partner X being a volunteer, by quantile.
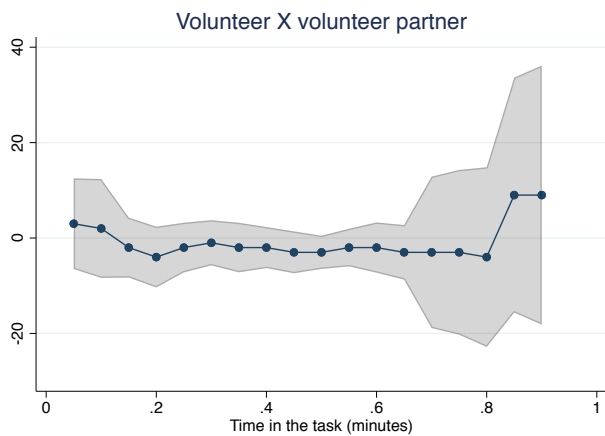
Figure 3.18: Coefficients of quantile regressions of envelopes in the sample of participants. Robust standard errors. 90% confidence intervals greyed out.

(a) Quantile effect of being a volunteer, by quantile.

(b) Quantile treatment effect of being assigned a volunteer partner, by quantile.
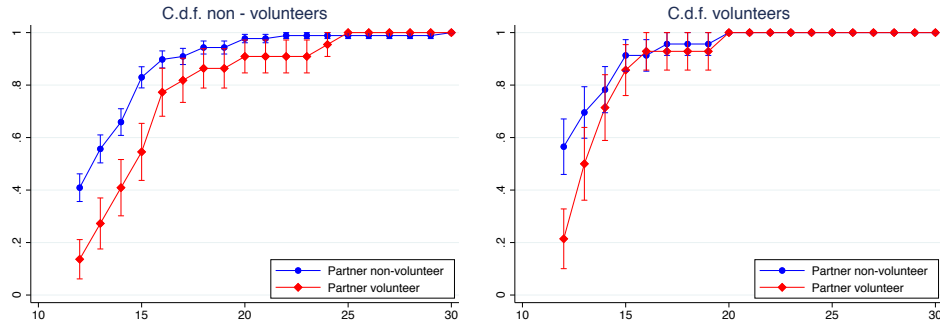

(c) Quantile interaction effect of being assigned a volunteer partner X being a volunteer, by quantile.
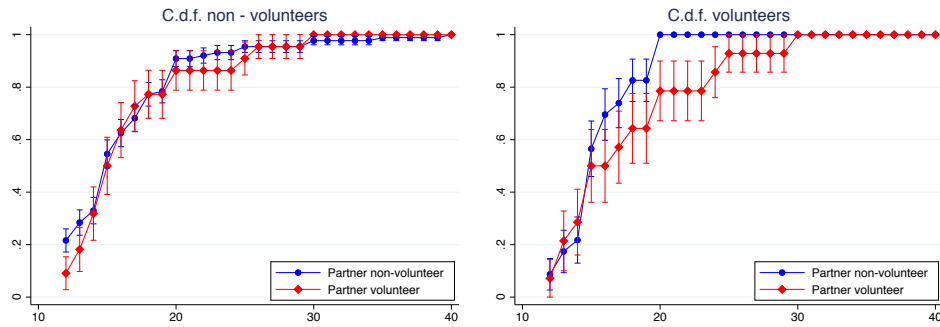
Figure 3.19:   Coefficients of quantile regressions of time spent (in minutes) in the sample of participants. No controls in the regressions. Robust standard errors. 90% confidence intervals greyed out.
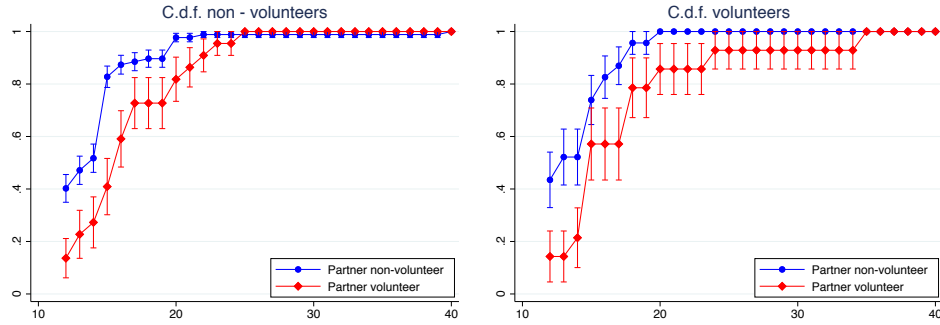
(a) Expected number of envelopes filled by partner

(b) Expected number of envelopes filled by non-volunteers

(c) Expected number of envelopes filled by volunteers
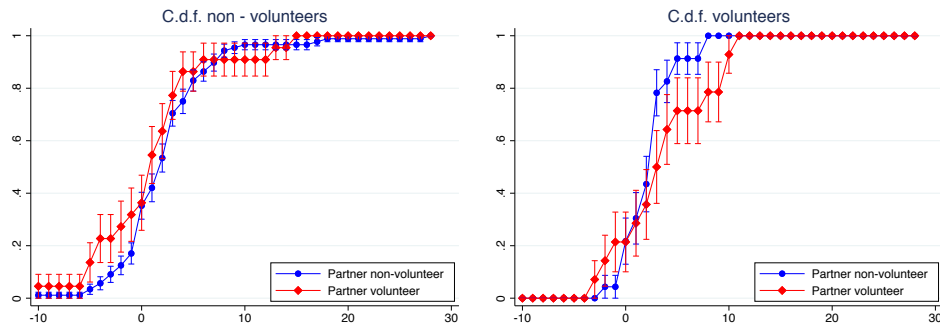
(d) Expected difference between envelopes filled by volunteers and envelopes filled by non-volunteers

Figure 3.20: Distribution of elicited expectations on other agents, among non-volunteers (left-hand side of the figure) and among volunteers (right hand side). Standard errors in bars.

3.C   Attrition

Table **??** suggests scope for selective attrition, among volunteers, between those matched with a volunteer and those matched with a non-volunteer. There are two stages where attrition can occur: one is right after the treatment assignment, when participants are asked to register because - if they do not - the whole team's output will be destroyed. Interestingly, the 15 percentage points difference in sign-up at this stage is not significant and might be attributed to chance. The overall difference, instead, is significant due to the compounding of the first stage with the decision to attend eventually. It would thus occur in the days after a participant has signed up.

It is difficult to see how the treatment decision might have had an impact in the days after it was taken, especially given that it was not notified again to those who signed up until they would attend the lab session.

One might be concerned that treatment induced different type of respondents to register and sign up, thus making comparisons between the two treatment conditions invalid. For this to be a cause for concern, it must be that the assignment of a non-volunteer partner discouraged sign-up and attendance among a selected group of people. For instance, one reason for lower sign-up among those treated with a volunteer partner is that the information that one's partner is a volunteer might signals that the volunteering decision is less rare and socially valuable than previously thought, and the whole task altogether seem less appealing. Those who are more discouraged by this effect could be those whose main drive is social approval. Among participants we would thus have two differently selected groups: those matched with a non-volunteer, who are relatively more driven by social approval, and those matched with a volunteer, who are those less driven by that (the others having been discouraged by the treatment itself). Comparisons in outcomes between these two groups will suffer from differences in potential outcomes between them.

Many more narratives could be found to argue that differential attrition could be a source of bias in this case. There are, however, good reasons to reject this concern.

First of all the sample of volunteers participating in the lab session is overall balanced on observables of different types, as Tables **??** to **??** strongly suggest.

Secondly, Tables **??** and **??** test for differential attrition by regressing attendance on each observable characteristic, volunteer-partner treatment, and their interaction term. In the null hypothesis where differential attrition does not cause selection bias, the interaction coefficient will be nil, as all characteristics will equally affect attendance in the two treatment conditions. Tables **??** and **??** report p-values of the interaction coefficients, confirming this.

Finally, there is a simpler more direct explanation for why volunteers who are treated with a non-volunteer partner are less likely to attend: the difference stems from imbalances that were naturally occurring at

the randomisation stage. The randomisation execution happened to assign more atheist, fewer christian people and people with more knowledge of LGBT-rights to the volunteer-partner condition. Interestingly and perhaps surprisingly, these characteristics are also negatively correlated with attendance; once they are controlled for, differences in attendance become smaller and less significant across the two treatments, among the volunteers.

Concerns for selective attrition in the group of non-volunteers are much less, first of all because the attrition rates are similar aross treatment conditions in this group, secondly because the two treatment conditions are balanced in the sample of participants (Tables ?? to ??), and thirdly because the tests carried out in Tables ?? and ?? confirm that there is no selective attrition on most observables.

Characteristic	Coefficient	Coefficient's p-val	Interaction	Interaction's p-val
Age (pre-screen)	-0.0023	0.79	-0.0401	0.14
Female (0-1)	0.0283	0.76	-0.4172	0.04
Heterosexual (0-1)	-0.2316	0.07	0.2207	0.42
Gay/Lesbian (0-1)	0.2844	0.21	-0.0044	0.99
Bisex (0-1)	0.1607	0.33	-0.9299	0.06
Student at LSE (0-1)	0.1412	0.14	0.1683	0.46
Economics student (0-1)	0.1626	0.19	-0.2459	0.42
English language (0-1)	-0.2157	0.01	0.2706	0.15
Asian language (0-1)	0.2137	0.12	0.0555	0.91
Over five years l earning english (0-1)	0.0606	0.82	-0.3523	0.36
Working (0-1)	-0.0702	0.55	-0.0251	0.92
Working part-time (0-1)	-0.0233	0.84	-0.0720	0.76
Family income above median (0-1)	0.0224	0.81	-0.3319	0.15
Family income not reported (0-1)	-0.0010	0.99	0.0951	0.64

Table 3.18: Test of selective attrition for non-volunteers. The table reports coefficients from a separate regression of attendance on each specific caracteristics, partner type, and an interaction term. Under the null hypothesis of lack of attrition bias, coefficients for the interaction term are null.

Characteristic	Coefficient	Coefficient's p-val	Interaction	Interaction's p-val
Personal income above median (0-1)	-0.2985	0.02	0.0385	0.91
Personal income not reported (0-1)	0.0558	0.55	0.0748	0.71
Poor or working class (0-1)	-0.1592	0.25	0.0640	0.80
Lower middle class (0-1)	0.0299	0.73	-0.2026	0.40
Middle to upper class (0-1)	0.0198	0.82	-0.0425	0.83
Social class unreported	0.0857	0.49	0.2325	0.36
Atheist (0-1)	0.1905	0.04	-0.2262	0.29
Christian (0-1)	-0.1711	0.07	0.2652	0.19
Religion not reported (0-1)	-0.0282	0.85	-0.1446	0.59
LGBT+ rights knowledge (1-6)	-0.0313	0.36	0.0634	0.50
LGBT+ rights involvement (1-6)	-0.0226	0.53	0.0482	0.60
LGBT+ rights, care for (1-6)	-0.0088	0.73	0.0049	0.95
Interested in volunteering (0-1)	0.0224	0.81	0.3109	0.17

Table 3.19: Test of selective attrition for non-volunteers. The table reports coefficients from a separate regression of attendance on each specific caracteristics, partner type, and an interaction term. Under the null hypothesis of lack of attrition bias, coefficients for the interaction term are null.

Characteristic	Coefficient	Coefficient's p-val	Interaction	Interaction's p-val
Age (pre-screen)	0.0082	0.70	-0.0370	0.17
Female (0-1)	-0.2105	0.25	0.4070	0.14
Heterosexual (0-1)	-0.1905	0.35	0.2905	0.30
Gay/Lesbian (0-1)	0.1538	0.73	-0.2253	0.69
Bisex (0-1)	0.1600	0.62	-0.0433	0.92
Asexual (0-1)	0.1538	0.73	-0.7448	0.24
Other sexual orientation (0-1)	0.1538	0.73	-0.7448	0.24
Student at LSE (0-1)	0.0238	0.91	0.1873	0.54
Economics student (0-1)	0.1538	0.73	-0.2328	0.65
English language (0-1)	-0.1593	0.35	-0.0391	0.88
Italian language (0-1)	0.1538	0.73	-0.7448	0.24
Asian language (0-1)	0.1818	0.41	-0.2701	0.38
Working (0-1)	-0.3786	0.04	-0.0030	0.99
Working full time (0-1)	-0.8846	0.04	0.2937	0.63
Working part-time (0-1)	-0.2381	0.24	-0.0286	0.93
Family income above median (0-1)	0.2222	0.21	-0.4206	0.11

Table 3.20: Test of selective attrition for volunteers. The table reports coefficients from a separate regression of attendance on each specific caracteristics, partner type, and an interaction term. Under the null hypothesis of lack of attrition bias, coefficients for the interaction term are null.

Characteristic	Coefficient	Coefficient's p-val	Interaction	Interaction's p-val
Family income not reported (0-1)	-0.2412	0.16	0.5412	0.05
Personal income above median (0-1)	-0.8846	0.03	0.2004	0.65
Personal income not reported (0-1)	0.2353	0.16	0.1843	0.46
Poor or working class (0-1)	-0.1196	0.62	0.5741	0.26
Lower middle class (0-1)	-0.1857	0.34	0.0975	0.73
Middle to upper class (0-1)	0.1374	0.42	-0.1215	0.63
Social class unreported	0.1667	0.54	-0.2381	0.58
Atheist (0-1)	-0.1857	0.31	-0.2297	0.37
Christian (0-1)	0.2000	0.30	0.2545	0.60
Religion not reported (0-1)	0.1600	0.62	0.2945	0.60
LGBT+ rights knowledge (1-6)	0.0415	0.66	-0.3027	0.03
LGBT+ rights involvement (1-6)	-0.0219	0.76	-0.0651	0.53
LGBT+ rights, care for (1-6)	0.0579	0.47	0.0080	0.94
Interested in volunteering (0-1)	-0.0333	0.85	0.1250	0.63

Table 3.21: Test of selective attrition for volunteers. The table reports coefficients from a separate regression of attendance on each specific caracteristics, partner type, and an interaction term. Under the null hypothesis of lack of attrition bias, coefficients for the interaction term are null.

3.D   Beliefs

This section documents and discusses in detail results from the elicitation of beliefs. Table **??** reports raw data and elaboration from responses to this questionnaire. Let us comment the main findings.

Self-reported behaviour.

Most respondents correctly remembered their partner's type. Non-volunteers had significantly worse recollection when matched with volunteers; volunteers had non-significantly better recollection when matched with volunteers. This is surprising as one might instead believe that partner's career choice would be more salient when it differs from own choice. On the other hand, it could be evidence of recollection bias or strategic manipulation of recollection.

Some respondents, though far from all, reported to be affected by information about their partner: among non-volunteers, as few as 36.7% of those matched with non-volunteers say they were affected by their partner, while 53% of those matched with volunteers responded they were affected by their partner. Higher numbers are observed among volunteers, suggesting that their higher motivation does indeed make them more responsive to information about the partner. In that case, too, the partner's influence was higher among those matched with a volunteer. Like with recollection, this result goes against the theory that a partner of a different career choice makes career choice more salient in the decision making process. It is instead consistent with a theory where not volunteering is seen as the norm while volunteering is seen as more striking and worthy.

Among those reporting that they were affected by their partner's career choice, some say that knowing their partner's choice encouraged them to make more effort. Among those who report being affected in the first place, the percentage of those who were encouraged by that information to exert more effort was always higher in the volunteer-partner treatment condition.

Expectations on partner.

In order to have more "hard" information about how people form their beliefs, we need to look at the guesses that they were asked to make in the post-experiment questionnaire.

It is clear that the type of partner assigned does matter and, among participants, those matched with a volunteer have higher expectations on their partner than those matched with a non-volunteer. The average difference in expectations is sizeable and significant: 1.75 for non-volunteers, 3.83 for volunteers. This difference is also equally spread across the distribution, as shown by the c.d.f. reported in the top row of Figure **??**.

Expectations on volunteers and non-volunteers.

When matched with a volunteer, participants acquire a better view of their own type. In all cases, on average,

they still believe that volunteers exert more effort than non-volunteers; however non-volunteers decrease their "expected volunteer effect" (that is, their conjecture on the average difference in effort between volunteers and non-volunteers) and volunteers increase it. This is consistent with the idea that agents have self-image concerns and that the volunteering decision is more salient when the partner is a volunteer. Among volunteers, this also contradicts the theory that agents use partner's career to update their expectations of how "selective" the volunteering decision is.

This, however, masks interesting differences in the distribution. The percentage of participants who believe that volunteers will make strictly more effort than non-volunteers is roughly equal across treatment groups, at about 66% among non-volunteers and 78% among volunteers. The percentage of people who believe that volunteers make strictly less effort than non-volunteers increases when the partner is a volunteer - the increase being significant only among volunteers. The bottom row in Figure **??** shows this effect.

Expectations on own performance compared to own group.

The overwhelming majority of participants has filled more envelopes than they believe their own group has. Among non-volunteers, however, this percentage drops significantly.

Expectation formation: first-order and higher-order beliefs.

In the theoretical model it was assumed that agents were somewhat naive and would not anticipate the effect of the own status on partner's behaviour. This assumption is actually consistent with the behaviour of a large group of participants, as highlighted by the percentage of respondents who believe that their partner fills as many envelopes as they believe the partner's group does. About half of participants matched with non-volunteers believe that their own partner will behave exactly as the average of their group. This figure includes those who believe that volunteers and non-volunteers have equal performance on average. However, among non-volunteers matched with non-volunteers, this figure is as low as 20%. For these subjects, beliefs on their partner are first order.

For a subject who does not classify as responding to first-order beliefs, beliefs could be rationalised by assuming that this subject believes that their partner holds the same belief on the sign of the "volunteer effect" as they do, and that the partner responds in line with these beliefs. Under these assumptions, a subject's beliefs are rationalisable at a higher order if the subject believes that their group is the most (least) productive, and if they believe that their partner is more (less) productive than the average in their partner's group is. For instance, a non-volunteer who believes that volunteers exert more effort than non-volunteers, and is matched with a volunteer, can be rationalised as having higher order beliefs if they believe that their partner does less effort than volunteers, on average, do. Table **??** shows that being assigned a partner from the opposite group significantly increases the likelihood of exerting second order reasoning. It also shows that, in a mixed team, the non-volunteer is more likely to exert second order reasoning than the volunteer.

These results imply that expectations on groups by career choice are not always equal to the counterfactual expectation on the partner, if the partner were from the other group.

3.E Conditional and unconditional expectations

Partner information could affect expectations in two ways: first, by affecting beliefs about the unconditional effort exerted by volunteers and non-volunteers, and secondly, by affecting beliefs on partner - taking unconditional beliefs as given. The ambiguity of the overall effect calls for elicitation of unconditional beliefs along with the conditional ones.

Figure **??** shows that being matched with a volunteer has a clear effect on expectations on partner's effort, with the distribution of expected effort among those assigned a volunteer first order stochastically domintating that among those matched with a non-volunteer. Yet, unconditional expectations are affected in different ways that are not immediate to explain using a standard framework.

To see the ambiguity of the effect, consider an agent who believes that more motivated agents are at the same time more likely to volunteer their compensation and to exert effort. Learning that the partner is a volunteer will raise expectations on the partner's effort. However, something else might happen. If an agent has a strong belief about the underlying distribution of motivation in the population, but is not totally sure of how much more motivation it takes to be volunteers, learning that one's partner is a volunteer will signal that the choice of volunteering is more common and less selective than previuosly thought. The perceived link between motivation and volunteering will become weaker and the expectation on effort exerted by the other partner will decrease. To test whether the second channel is at play, we asked subjects not only what effort they would expect of their partner, but also how much they would expect of volunteers and non-volunteers unconditionally.

A stylised model that explains this intuition is one where agents' motivation m_i is distributed according c.d.f. $F_M(\cdot)$. Agents decide to volunteer if motivation m is greater than an unknown threshold level λ_i over which everybody has a prior $\pi(\lambda)$. Effort in the task is a function $e_i = e(m_i)$ which is increasing in motivation. Agents observe i's volunteering decision v_i and then update their posterior c.d.f. of λ_i conditional on v_i $P_{\lambda|v}(\lambda_i|v_i)$, and c.d.f. $P_{M|v,\lambda}(m_i|\lambda_i,v_i)$ according to Bayes' rule.

We can think of λ_i as a threshold that depends on individual income. Observing an agent that volunteers might thus need that they are highly motivated or that their income is sufficiently large for them to just give up some money in support of a charity. Observing that the partner volunteers will make it more likely both that the partner is more motivated and that their threshold level is low enough. The inference on partner's motivation will thus be the result of these opposite forces.

Once we observe that agent i makes volunteering choices $v_i \in \{0,1\}$, expectations on their effort level are given by

$$E(e(m_i)|v_i) = E_{\lambda_i|v_i}\left[E(e(m_i)|\lambda_i, v_i)\right] \quad .$$

The effect of assigning a volunteer partner can be decomposed as

$$E(e(m_i)|1) - E(e(m_i)|0) =$$

$$E_{\lambda_i|1}\left[E(e(m_i)|\lambda_i, 1)\right] - E_{\lambda_i|0}\left[E(e(m_i)|\lambda_i, 0)\right] =$$

$$E_{\lambda_i|1}\left[E(e(m_i)|\lambda_i, 1)\right] - E_{\lambda_i|1}\left[E(e(m_i)|\lambda_i, 0)\right] + E_{\lambda_i|1}\left[E(e(m_i)|\lambda_i, 0)\right] - E_{\lambda_i|0}\left[E(e(m_i)|\lambda_i, 0)\right] =$$

$$\underbrace{\int_{\mathbb{R}}\left[E(e(m_i)|\lambda_i, 1) - E(e(m_i)|\lambda_i, 0)\right] dP(\lambda_i|1)}_{\text{Volunteering effect}} + \underbrace{\int_{\mathbb{R}} E(e(m_i)|\lambda_i, 0)\left[dP(\lambda_i|1) - dP(\lambda_i|0)\right]}_{\text{Learning effect}}$$

The volunteering effect is the change in expectations due to the higher motivation that is signalled by the volunteering decision, while the learning effect is the consequence of learning, from the choice of volunteering, that it might not have been as hard a decision as previously thought.

To show that the two effects are opposite in sign, consider that the volunteering effect is positive by the assumption that $e(\cdot)$ is an increasing function of m_i,

$$E(e(m_i)|\lambda_i, 1) = E(e(m_i)|\lambda_i, m_i \geq \lambda_i) > E(e(m_i)|\lambda_i, m_i < \lambda_i) = E(e(m_i)|\lambda_i, 0) \quad .$$

To see that the learning effect is negative, note that the expectation $E(e(m_i)|\lambda_i, 0) = E(e(m_i)|\lambda_i, m_i < \lambda_i)$ is increasing in λ_i. It can be proved that $P(\lambda_i|0)$ first order stochastically dominates $P(\lambda_i|1)$ since

$$P(\lambda_i|0) = Pr(\lambda \leq \lambda_i|m_i < \lambda_i) > Pr(\lambda \leq \lambda_i|m_i \geq \lambda_i) = P(\lambda_i|1)$$

So $\int_{\mathbb{R}} E(e(m_i)|\lambda_i, 0)\left[dP(\lambda_i|1) - dP(\lambda_i|0)\right] < 0$.

This "Bayesian theory" seems disconfirmed by the data. On the one hand, it is true that the percentage of respondents who believe that volunteers exert less effort than non-volunteers increases when subjects are paired up with volunteers, which is consistent with the idea that information on the higher frequency of volunteers might lower the expected value of volunteering as a signal of motivation and so effort. On the other hand, however, the third row of Figure **??** shows that, among those who believe that volunteers exert more effort than non-volunteers, expectations on the "volunteer effect" become relatively higher when the partner is a volunteer. Moreover, among non-volunteers, those partnered up with a volunteer improve their

beliefs on non-volunteers. This is inconsistent with the above model, where being partnered with a volunteer reveals that the "volunteering threshold" is lower than otherwise thought, thus decreasing expectations on non-volunteers as well as volunteers.

Finally, and most conclusively, the expected "volunteer effect" is not significantly different across treatment conditions and, if anything, is larger among volunteers matched with volunteers than among volunteers matched with non-volunteers. For these reasons the data encourages to reject the "Bayesian theory" of expectation formation in this framework.

Bibliography

[1] George A Akerlof and Rachel E Kranton. Economics and identity. *The Quarterly Journal of Economics*, 115(3):715–753, 2000.

[2] George A Akerlof and Rachel E Kranton. Identity and the economics of organizations. *Journal of Economic perspectives*, 19(1):9–32, 2005.

[3] Robert Akerlof and Richard Holden. Movers and shakers. *The Quarterly Journal of Economics*, 131(4):1849–1874, 2016.

[4] Michael G Allingham and Agnar Sandmo. Income tax evasion: A theoretical analysis. *Journal of public economics*, 1(3-4):323–338, 1972.

[5] Cameron Anderson, Dacher Keltner, and Oliver P John. Emotional convergence between people over time. *Journal of personality and social psychology*, 84(5):1054, 2003.

[6] James Andreoni. Impure altruism and donations to public goods: A theory of warm-glow giving. *The economic journal*, 100(401):464–477, 1990.

[7] James Andreoni, William G Gale, John Karl Scholz, and John Straub. Charitable contributions of time and money. *University of Wisconsin–Madison Working Paper*, 1996.

[8] Barak Ariel. Deterrence and moral persuasion effects on corporate tax compliance: findings from a randomized controlled trial. *Criminology*, 50(1):27–69, 2012.

[9] Kenneth J Arrow. *The limits of organization*. WW Norton & Company, 1974.

[10] Nava Ashraf, Oriana Bandiera, and B Kelsey Jack. No margin, no mission? a field experiment on incentives for public service delivery. *Journal of Public Economics*, 120:1–17, 2014.

[11] Nava Ashraf, Oriana Bandiera, and Scott Lee. Losing prosociality in the quest for talent? sorting, selection, and productivity in the delivery of public services. 2018.

[12] Nava Ashraf, Oriana Bandiera, and Scott S Lee. Do-gooders and go-getters. Technical report, Working paper, 2016.

[13] Kyle Bagwell and Michael H Riordan. High and declining prices signal product quality. *The American Economic Review*, pages 224–239, 1991.

[14] Oriana Bandiera, Iwan Barankay, and Imran Rasul. Social preferences and the response to incentives: Evidence from personnel data*. *The Quarterly Journal of Economics*, 120(3):917–962, 2005.

[15] Oriana Bandiera, Iwan Barankay, and Imran Rasul. The evolution of cooperative norms: Evidence from a natural field experiment. *Advances in economic analysis & policy*, 5(2), 2006.

[16] Oriana Bandiera, Iwan Barankay, and Imran Rasul. Social Incentives in the Workplace. *Review of Economic Studies*, 77(2):417–458, 2010.

[17] Roland Benabou and Jean Tirole. Intrinsic and extrinsic motivation. *The Review of Economic Studies*, 70(3):489–520, 2003.

[18] Roland Benabou and Jean Tirole. Incentives and prosocial behavior. *The American Economic Review*, 96(5):1652–1678, 2006.

[19] Johannes Berger, Claus Herbertz, and Dirk Sliwka. Incentives and cooperation in firms: Field evidence. 2011.

[20] B. Douglas Bernheim. A theory of conformity. *Journal of Political Economy*, 102(5):841–877, 1994.

[21] Tim Besley and Maitreesh Ghatak. Competition and incentives with motivated agents. *The American Economic Review*, 95(3):616–636, 2005.

[22] Marsha Blumenthal, Charles Christian, Joel Slemrod, and Matthew G Smith. Do normative appeals affect tax compliance? evidence from a controlled experiment in minnesota. *National Tax Journal*, pages 125–138, 2001.

[23] Anne Brockmeyer, Marco Hernandez, Stewart Kettle, and Spencer Smith. Casting a wider tax net: Experimental evidence from costa rica. 2018.

[24] Judy Cameron and W. David Pierce. Reinforcement, reward, and intrinsic motivation: A meta-analysis. *Review of Educational Research*, 64(3):363–423, 1994.

[25] Lucio Castro and Carlos Scartascini. Tax compliance and enforcement in the pampas evidence from a field experiment. *Journal of Economic Behavior & Organization*, 116:65–82, 2015.

[26] Michael Chirico, Robert Inman, Charles Loeffler, John MacDonald, and Holger Sieg. Deterring tax delinquency in philadelphia. Technical report, National Bureau of Economic Research, 2017.

[27] Jessica Cohen, Pascaline Dupas, et al. Free distribution or cost-sharing? evidence from a randomized malaria prevention experiment. *Quarterly journal of Economics*, 125(1):1, 2010.

[28] Frank Cowell, Susan Long, et al. *Tax evasion: An experimental approach*. Cambridge University Press, 1991.

[29] Frank A Cowell and James PF Gordon. Unwillingness to pay: Tax evasion and public good provision. *Journal of Public Economics*, 36(3):305–321, 1988.

[30] Ernesto Dal Bó and Pedro Dal Bó. "do the right thing:" the effects of moral suasion on cooperation. *Journal of Public Economics*, 117:28–38, 2014.

[31] Ernesto Dal Bó, Frederico Finan, and Martín A Rossi. Strengthening state capabilities: The role of financial incentives in the call to public service. *The Quarterly Journal of Economics*, 128(3):1169–1218, 2013.

[32] Edward L. Deci. The effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, 18:105–115, 1971.

[33] Lucia Del Carpio. Are the neighbors cheating? evidence from a social norm experiment on property taxes in peru job market paper. 2013.

[34] Roberto Dell'Anno. Tax evasion, tax morale and policy maker's effectiveness. *The Journal of Socio-Economics*, 38(6):988–997, 2009.

[35] Peter DeScioli, Kelly Asao, and Robert Kurzban. Omissions and byproducts across moral domains. *PloS one*, 7(10):e46963, 2012.

[36] Erika Deserranno. Financial incentives as signals: experimental evidence from the recruitment of village promoters in uganda. *American Economic Journal: Applied Economics*, 11(1):277–317, 2019.

[37] Nadja Dwenger, Henrik Kleven, Imran Rasul, and Johannes Rincke. Extrinsic and intrinsic motivations for tax compliance: Evidence from a field experiment in germany. *American Economic Journal: Economic Policy*, 8(3):203–32, 2016.

[38] Robert Eisenberger and Judy Cameron. Detrimental effects of reward. reality or myth? *American Psychologist*, 51(11):1153–1166, 1996.

[39] Armin Falk and Andrea Ichino. Clean evidence on peer effects. *Journal of Labor Economics*, 24(1):39–57, 2006.

[40] Armin Falk and Andrea Ichino. Clean evidence on peer effects. *Journal of Labor economics*, 24(1):39–57, 2006.

[41] Ernst Fehr and Simon Gachter. Cooperation and punishment in public good experiments. *The American Economic Review*, 90(4):980–994, 2000.

[42] Ernst Fehr and Simon Gächter. Fairness and retaliation: The economics of reciprocity. *Journal of economic perspectives*, 14(3):159–181, 2000.

[43] Ernst Fehr and Simon Gachter. Do incentive contracts crowd out voluntary cooperation? 2001.

[44] Ernst Fehr and John A. List. The hidden costs and returns of incentives - trust and trustworthyness among ceos. *Journal of the European Economic Association*, 2(5):743–771, 2004.

[45] Ernst Fehr and Klaus M. Schmidt. A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3):817–868, 1999.

[46] G. Fellner, R. Sausgruber, and C. Traxler. Testing enforcement strategies in the field: Threat, moral appeal and social information. *Journal of the European Economic Association*, 11(3):634–660, 2013.

[47] Urs Fischbacher, Simon Gächter, and Ernst Fehr. Are people conditionally cooperative? evidence from a public goods experiment. *Economics Letters*, 71(3):397 – 404, 2001.

[48] Bruno Frey and Lorenz Goette. Does pay motivate volunteers? *University of Zurich, Institute for Empirical Resarch in Economics Working Paper 7*, 1999.

[49] Bruno S Frey. How intrinsic motivation is crowded out and in. *Rationality and society*, 6(3):334–352, 1994.

[50] Alan S Gerber and Todd Rogers. Descriptive social norms and motivation to vote: Everybody's voting and so should you. *The Journal of Politics*, 71(1):178–191, 2009.

[51] Uri Gneezy and Aldo Rustichini. Pay enough or don't pay at all. *The Quarterly Journal of Economics*, 115(3):791–810, 2000.

[52] Jonathan Haidt et al. The moral emotions. *Handbook of affective sciences*, 11(2003):852–870, 2003.

[53] M Hallsworth. The omission bias. *Unpublished Paper*, 2013.

[54] Michael Hallsworth. The use of field experiments to increase tax compliance. *Oxford Review of Economic Policy*, 30(4):658–679, 2014.

[55] Michael Hallsworth, John A List, Robert D Metcalfe, and Ivo Vlaev. The making of homo honoratus: From omission to commission. Technical report, National Bureau of Economic Research, 2015.

[56] Michael Hallsworth, John A List, Robert D Metcalfe, and Ivo Vlaev. The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics*, 148:14–31, 2017.

[57] John Hasseldine, Peggy Hite, Simon James, and Marika Toumi. Persuasive communications: Tax compliance enforcement strategies for sole proprietors. *Contemporary Accounting Research*, 24(1):171–194, 2007.

[58] James J Heckman and Rodrigo Pinto. Econometric mediation analyses: Identifying the sources of treatment effects from experimentally estimated production technologies with unmeasured and mismeasured inputs. *Econometric reviews*, 34(1-2):6–31, 2015.

[59] Martin L Hoffman. Affect and moral development. *New Directions for Child and Adolescent Development*, 1982(16):83–103, 1982.

[60] Martin L Hoffman. Development of prosocial motivation: Empathy and guilt. In *The development of prosocial behavior*, pages 281–313. Elsevier, 1982.

[61] Bengt Holmstrom and Paul Milgrom. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *JL Econ. & Org.*, 7:24, 1991.

[62] David W Johnson and Roger T Johnson. *Cooperation and competition: Theory and research.* Interaction Book Company, 1989.

[63] Judd B Kessler. Announcements of support and public good provision. *American Economic Review*, 2013.

[64] Stewart Kettle, Marco Hernandez, Michael Sanders, Oliver Hauser, and Simon Ruda. Failure to captcha attention: Null results from an honesty priming experiment in guatemala. *Behavioral Sciences*, 7(2):28, 2017.

[65] Henrik Jacobsen Kleven, Martin B Knudsen, Claus Thustrup Kreiner, Søren Pedersen, and Emmanuel Saez. Unwilling or unable to cheat? evidence from a tax audit experiment in denmark. *Econometrica*, 79(3):651–692, 2011.

[66] Edward P Lazear. Pay equality and industrial politics. *Journal of political economy*, 97(3):561–580, 1989.

[67] Edward P Lazear. Labor economics and the psychology of organizations. *Journal of Economic perspectives*, 5(2):89–110, 1991.

[68] Edward P. Lazear, Ulrike Malmendier, and Roberto A. Weber. Sorting in experiments with application to social preferences. *American Economic Journal: Applied Economics*, 4(1):136–63, 2012.

[69] Julian Le Grand. *Motivation, agency, and public policy: of knights and knaves, pawns and queens.* Oxford University Press on Demand, 2003.

[70] HB Lewis. Shame and guilt in neurosis. new york (international universities press) 1971. 1971.

[71] M Lewis. Self-conscious emotions: Embarrassment, pride, shame, and guilt,[w:] m. lewis, jm haviland (red.), handbook of emotions (s. 563-573), 1993.

[72] Alexandre Mas and Enrico Moretti. Peers at work. *American Economic Review*, 99(1):112–145, 2009.

[73] Nina Mazar, On Amir, and Dan Ariely. The dishonesty of honest people: A theory of self-concept maintenance. *Journal of marketing research*, 45(6):633–644, 2008.

[74] Darryl W Miller and Lawrence J Marks. The effects of imagery-evoking radio advertising strategies on affective responses. *Psychology & Marketing*, 14(4):337–360, 1997.

[75] Gareth D Myles and Robin A Naylor. A model of tax evasion with group conformity and social customs. *European Journal of Political Economy*, 12(1):49–66, 1996.

[76] OECD. Country note—education at a glance 2013: Poland. Technical report, 2013.

[77] OECD. Country fact sheet—government at a glance 2015: Poland. Technical report, 2015.

[78] University of Oklahoma. Institute of Group Relations and Muzafer Sherif. *Intergroup conflict and cooperation: The Robbers Cave experiment*, volume 10. University Book Exchange Norman, OK, 1961.

[79] Daniel Ortega and Pablo Sanguinetti. Deterrence and reciprocity effects on tax compliance: experimental evidence from venezuela. 2013.

[80] Dina Pomeranz. No taxation without information: Deterrence and self-enforcement in the value added tax. *American Economic Review*, 105(8):2539–69, 2015.

[81] Canice Prendergast. The provision of incentives in firms. *Journal of economic literature*, 37(1):7–63, 1999.

[82] Bruce Sacerdote. Peer effects with random assignment: Results for dartmouth roommates. *The Quarterly Journal of Economics*, 116(2):681–704, 2001.

[83] William Samuelson and Richard Zeckhauser. Status quo bias in decision making. *Journal of risk and uncertainty*, 1(1):7–59, 1988.

[84] Lisa L Shu, Nina Mazar, Francesca Gino, Dan Ariely, and Max H Bazerman. Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end. *Proceedings of the National Academy of Sciences*, 109(38):15197–15200, 2012.

[85] Joel Slemrod, Marsha Blumenthal, and Charles Christian. Taxpayer response to an increased probability of audit: evidence from a controlled experiment in minnesota. *Journal of Public Economics*, 79(3):455 – 483, 2001.

[86] Robert Sugden. Reciprocity: The supply of public goods through voluntary contributions. *The Economic Journal*, 94(376):772–787, 1984.

[87] Cass R Sunstein. *Choosing not to choose: Understanding the value of choice.* Oxford University Press, USA, 2015.

[88] Henri Tajfel, Michael G Billig, Robert P Bundy, and Claude Flament. Social categorization and intergroup behaviour. *European journal of social psychology*, 1(2):149–178, 1971.

[89] Benno Torgler. Moral suasion: An alternative tax policy strategy? evidence from a controlled field experiment in switzerland. *Economics of Governance*, 5(3):235–253, 2004.

[90] Benno Torgler. *Tax compliance and tax morale: A theoretical and empirical analysis.* Edward Elgar Publishing, 2007.

[91] Benno Torgler. A field experiment in moral suasion and tax compliance focusing on underdeclaration and overdeduction. *FinanzArchiv: Public Finance Analysis*, 69(4):393–411, 2013.

[92] Amos Tversky and Daniel Kahneman. The framing of decisions and the psychology of choice. *Science*, 211(4481):453–458, 1981.

[93] Michael Wenzel. Misperceptions of social norms about tax compliance: From theory to intervention. *Journal of Economic Psychology*, 26(6):862–883, 2005.

[94] Michael Wenzel and Natalie Taylor. An experimental evaluation of tax-reporting schedules: a case of evidence-based tax administration. *Journal of Public Economics*, 88(12):2785 – 2799, 2004.

[95] Shlomo Yitzhaki. A note on income tax evasion: A theoretical analysis. *Journal of public economics*, 3:201–202, 1974.

[96] Jungmin Yoo and Minjeong Kim. The effects of online product presentation on consumer responses: A mental imagery perspective. *Journal of Business Research*, 67(11):2464–2472, 2014.