Online Appendix: "Revenue Drift, Incentives, and Effort Allocation in Social Enterprises"

Theodor Vladasel* Simon C. Parker† Randolph Sloof‡ Mirjam van Praag§

March 5, 2024

This online appendix includes additional material for our paper "Revenue Drift, Incentives, and Effort Allocation in Social Enterprises". Appendix A provides a simple theoretical model, with proofs in Appendix B. We provide the full experimental instructions including company and task descriptions, payoff explanations, social preference measures, and other demographic variables in Appendix C. Appendix D then provides a set of experimental design checks, including an analysis of good cause choice, a randomization check, and tests performed using randomization inference standard errors. Appendix E details the results of ex ante power calculations and ex post minimum detectable effect sizes. Appendix F shows that results are robust to tightening or relaxing the sample restrictions and discusses the results of our pilot experiment, Appendix G details our common factor analysis and examines robustness to alternative social preference measures, and Appendix H considers a set of alternative effort measures and regression models. Appendix I discusses the results of multiple hypothesis testing adjustments, reflecting the number of outcomes and treatments we consider, and performs heterogeneity analyses by gender, previous social sector experience, and mission choice. Appendix J describes an additional study where we vary several parameters to assess the robustness of our main results and generate insight into the role of i) even stronger incentives, ii) effects on total effort, and iii) the specificity of our results to the social enterprise setting.

Contents

* Universitat Pompeu Fabra, UPF Barcelona School of Management, and Barcelona School of Economics, Ramon Trias Fargas 25-27, Edifici Jaume I, 20.1E68, 08005 Barcelona, Spain, theodor.vladasel@upf.edu

† Ivey Business School at Western University, Canada, University of Aberdeen Business School, UK, and IZA, Germany, sparker@ivey.ca

‡ Amsterdam School of Economics at the University of Amsterdam and Tinbergen Institute, the Netherlands, r.sloof@uva.nl

§ Vrije Universiteit Amsterdam, the Netherlands, Copenhagen Business School, Denmark, CEPR, UK, IZA, Germany, and Tinbergen Institute, the Netherlands, c.m.van.praag@vu.nl

Appendix A Model

We provide a simple model that illustrates the verbal reasoning in Section 2 in a more concrete way. It also formally derives our two main hypotheses and the boundary conditions under which these hypotheses obtain.

Suppose a worker exerts effort on two different tasks: a social task that cannot be (easily) measured and a commercial task that can be measured and thus contracted upon. Effort exerted on the social task is denoted $s \geq 0$, while effort devoted to the commercial task is denoted $c \geq 0$. Our main focus is on how monetary incentives for the commercial task redirect a worker's effort towards this task, at the expense of the social task. We therefore impose the restriction that $s + c = \overline{T} \equiv 1$, with $\overline{T}$ the fixed overall amount of effort a worker exerts (normalized to one). Put differently, our focus is on how workers divide a given amount of overall effort between the two tasks. We believe this captures reality well in that social enterprises typically provide environments with generally high social motivation levels, where people do work hard – such that incentives are not really be about increasing effort, but mostly about re-directing it. Imposing the restriction is also in line with our experimental design, where the overall amount of sliders is fixed (and participants typically perform all of these). The costs of exerting effort are assumed to equal $C(s, c) = \frac{1}{2}s^2 + \frac{1}{2}c^2$. Note that although the two efforts do not directly interact in the cost function, the restriction $s + c = 1$ implies they are substitutes of each other.[1]

Workers obtain benefits from exerting effort, as well as from working for an organization of type O *per se*. Consider first the benefits from exerting effort. On top of base salary m_O, the organization pays a piece rate b_O (with $0 \leq b_O \leq 1$) per unit of commercial task effort. In contrast, workers are not extrinsically rewarded for performing the social task, but only intrinsically motivated to do so. Workers differ in the basic strength of their intrinsic motivation, i.e. in their prosociality type v, with v uniformly distributed on the unit interval (i.e. $v \sim U[0, 1]$). Besides their type, intrinsic motivation is also affected by the organizational environment, in particular the monetary incentives b_O the organization sets for the commercial task and its mission orientation $\nu_O \in [0, 1]$. If O's sole mission is to generate profits via the commercial task, then $\nu_O = 0$, while if the sole mission is to do good via the social task, then $\nu_O = 1$. Multiple-mission organizations like social enterprises have a ν_O that falls in between. We assume that (in utility terms) the worker gets $\gamma(b_O, v_O) \cdot v$ per unit of social task effort. Multiplicative factor $\gamma(b_O, v_O) \in [0, 1]$ is assumed to be a function of b_O, capturing that commercial task monetary incentives may directly crowd-out, or possibly crowd-in, a worker's basic intrinsic motivation to exert social task effort (Frey, 1994). Similarly, we assume $\gamma(b_O, v_O)$ weakly increases with v_O, reflecting the impact of social norm setting via organizational mission v_O; the more prosocial this mission, the more a worker's basic intrinsic motivation to exert social effort is strengthened.

The environment created via (b_O, ν_O) also affects how workers value the organization *per se*. Irrespective of their own effort choice, more prosocial types dislike an organization's use of financial incentives. They experience a disutility of $\rho \cdot v \cdot b_O$ if the organization sets piece rate b_O, with $\rho \geq 0$ a parameter reflecting the extent of such distaste. Our multiplicative specification implies that more prosocial types dislike a piece rate of a given size more. Finally, we assume that workers care about how well their own prosociality preferences match organizational mission. Workers experience disutility $\varphi \cdot (v - \nu_O)^2$ the further away their own ideals are from the organization's ideal. Parameter $\varphi \geq 0$ measures the extent of such mission-match motivations.[2]

[1] An alternative setup would be to leave the $s+c = 1$ restriction out, so overall effort varies with the incentives provided, but assume instead that $C(s, c; \delta) = \frac{1}{2}s^2 + \frac{1}{2}c^2 + \delta \cdot s \cdot c$, with $\delta > 0$ (and $\delta^2 < 1$). Parameter δ then measures the extent to which the two tasks are substitutes and compete for the worker's time. Results essentially would be the same as derived below; results would likely change only if $\delta < 0$ and the tasks were complements.

[2] In our general setup, an organization's mission may affect a worker's marginal incentives to exert social effort via factor $\gamma(b_O, v_O)$, as well as provide a direct non-monetary utility premium (or better, discount) via term $\varphi \cdot (v - \nu_O)^2$. This builds on a large body of existing 'mission-motivation' models where both these approaches have been used. Besley and Ghatak (2005) assume the marginal return to social effort increases with the mission-match between worker mission preferences v and organizational mission ν_O; see also Besley and Ghatak (2018) and

With these elements, the worker's optimal effort allocation between tasks follows from:

$$\max_{s,c} U(s,c) \equiv m_O - \varphi \cdot (v - \nu_O)^2 - \rho \cdot v \cdot b_O + \gamma(b_O, v_O) \cdot v \cdot s + b_O \cdot c - \frac{1}{2}s^2 - \frac{1}{2}c^2, \quad (1)$$
$$\text{s.t. } s, c \geq 0 \text{ and } s + c = 1.$$

Solving (1) yields the following optimal effort levels s^* and c^*:

$$s^* = \frac{\gamma(b_O, v_O) \cdot v - b_O + 1}{2} \quad \text{and} \quad c^* = \frac{b_O - \gamma(b_O, v_O) \cdot v + 1}{2}. \quad (2)$$

From these we immediately arrive at Proposition 1, which justifies our first hypothesis regarding the impact of monetary rewards through the incentive channel.

Proposition 1 *(Incentive channel) If* $\frac{\partial \gamma}{\partial b_O} < \frac{1}{2}$*, an increase in incentive pay* b_O *shifts all worker types' attention towards the commercial task:* $\frac{\partial c^*}{\partial b_O} = \frac{1}{2} - \frac{\partial \gamma}{\partial b_O} \cdot v > 0$ *and* $\frac{\partial s^*}{\partial b_O} = -\frac{1}{2} + \frac{\partial \gamma}{\partial b_O} \cdot v < 0$ *for all* $v \in [0,1]$.

Our formulation of Hypothesis 1 in the main text therefore assumes that stronger monetary incentives for the commercial task do not induce strong crowding-in effects with respect to the social task. In fact, in line with Frey (1997), our Section 2 verbal reasoning refers to crowding out (i.e. $\frac{\partial \gamma}{\partial b_O} \leq 0$) as the most realistic scenario in the social enterprise context.[3] In a similar vein, the verbal argument that crowding out likely becomes stronger for high-powered incentives can be captured by assuming $\frac{\partial^2 \gamma}{\partial b_O^2} \leq 0$. In the main text we also reason that mismatch between an organization's mission and its monetary incentives may weaken intrinsic motivation; assuming $\frac{\partial^2 \gamma}{\partial v_O \partial b_O} \leq 0$ would capture this. Our formal specification above is sufficiently rich to capture all verbal arguments for Hypothesis 1. Importantly, all these forces predict effort allocation to move in the same direction in response to stronger monetary incentives.[4,5]

Turning to the selection channel, let $V_O \equiv U(s^*, c^*)$ denote the worker's utility level when working for organization O, given that the worker chooses effort optimally according to (2):

$$V_O = m_O - \varphi \cdot (v - \nu_O)^2 - \rho \cdot v \cdot b_O + \frac{1}{4} \cdot \left[(b_O - \gamma(b_O, v_O) \cdot v)^2 + 2(b_O + \gamma(b_O, v_O) \cdot v) - 1 \right]. \quad (3)$$

The final term reflects the net benefits from exerting optimal effort. Now suppose there are three types of organizations, each with different objectives ('missions') and incentive contracts. In particular, for-profits (FP) only care about maximizing profits and provide steep monetary incentives: $\nu_{FP} = 0$ and $b_{FP} = 1$. Non-profits (NP) are solely interested in the social goals and do not use incentive pay at all: $\nu_{NP} = 1$ and $b_{NP} = 0$. Social enterprises (SE) fall in between: they care about the two tasks equally and may deploy monetary incentives: $\nu_{SE} = \frac{1}{2}$ and $0 \leq b_{SE} \leq 1$. Workers self-sort into the organization corresponding to $\max\{V_{FP}, V_{SE}, V_{NP}\}$.

Cassar (2019). Barigozzi et al. (2018) instead assume that workers get a fixed, effort-independent utility benefit when working for a non-profit rather than a for-profit. Delfgaauw and Dur (2010) consider both approaches and compare the case where intrinsic public sector motivation is effort-dependent with the one where it is not.

[3] Note that this direct crowding out may be to the organization's detriment if it leads to less social effort than it aims for given its mission.

[4] The optimal effort levels in (2) also illustrate that, for a given incentive level b_O, effort allocation may vary with the organization's mission ν_O. For instance, a social enterprise with $b_{SE} = 1$ induces less commercial effort than a for-profit with the same steep monetary incentives $b_{FP} = 1$. Similarly, a social enterprise with $b_{SE} = 0$ induces less social effort than a non-profit with the same (lack of) incentives $b_{NP} = 0$.

[5] When incentives are absent, i.e. $b_{SE} = 0$, the expected value of social task effort for workers who select into social enterprises is $E[s^*] = (\gamma \cdot E[v] + 1)/2$, where $E[v] = (\underline{v} + \overline{v})/2$ under the assumed uniform distribution (see Lemma 2 for how these thresholds are defined). Since in this case $\overline{v} = 3/4$ necessarily (when both SE and NP provide flat incentives, indifference follows from equidistance to mission), then $E[v] \geq 3/8$. Thus, social enterprise social effort $E[s^*] \geq (3\gamma + 8)/16 > 1/2$ when $b_{SE} = 0$, in line with our baseline prediction in the main text. Note that for $\gamma \equiv 1$ as considered below, this tightens to $E[s^*] \geq 11/16 \approx 0.69$.

A first useful observation is that different prosociality types self-select into the three types of organizations in an intuitive way. This is a direct consequence of stronger monetary incentives bringing lower incremental benefits the higher the worker's prosociality, while a more prosocial mission brings higher additional benefits for more prosocial types (i.e. $\frac{\partial^2 V_O}{\partial v \partial b_O} \leq 0$ and $\frac{\partial^2 V_O}{\partial v \partial \nu_O} \geq 0$).

Lemma 1 *There exist thresholds $\underline{v}$ and $\overline{v}$, with $0 \leq \underline{v} \leq \overline{v} \leq 1$, such that types $v \leq \underline{v}$ sort into the for-profit, those in $[\underline{v}, \overline{v}]$ sort into the social enterprise, while types $v > \overline{v}$ sort into the non-profit. Moreover, it holds that $\frac{\partial \underline{v}}{\partial b_{SE}} > 0 \Rightarrow \frac{\partial \overline{v}}{\partial b_{SE}} < 0$: if an increase in b_{SE} leads to more selfish types' out-selection, then it necessarily also leads to more pro-social types' out-selection.*

We provide the proofs of Lemma 1 and subsequent formal results in Online Appendix B. The selection effect of stronger incentives derives from how variations in b_{SE} affect workers' organizational type choice, i.e. from how thresholds $\underline{v}$ and $\overline{v}$ shift with changes in b_{SE}. Lemma 1's final part excludes the counter-intuitive case where an increase in b_{SE} makes the social enterprise less attractive to the more selfish types, but more attractive to the more prosocial types (that is, $\underline{v}$ and $\overline{v}$ cannot both shift to the right). Three possibilities arise after an increase in b_{SE}: (a) in-selection of both more selfish and more prosocial types, (b) in-selection of more selfish types, but out-selection of more prosocial types, and (c) out-selection of both more selfish and more prosocial types. Intuitively, one would expect case (a) to occur if the marginal demotivational effects of stronger incentives are weak or absent (i.e. if $-\rho$ and $\frac{\partial \gamma}{\partial b_O}$ are both close to zero) and case (c) if these are very strong (either $-\rho$ or $\frac{\partial \gamma}{\partial b_O}$ is strongly negative). Case (b) then applies if demotivational effects are substantive, but not overwhelming. We make this intuition explicit by focusing on the specific case where $\gamma(b_O, v_O) \equiv 1$.[6] Following this, we later justify Hypothesis 2 since we consider case (b) the most plausible.

Lemma 2 *Let $\gamma(b_O, v_O) \equiv 1$. Suppose $m_{FP} = m_{SE} = m_{NP}$ and $\varphi \geq \frac{\rho}{2} + \frac{7}{4}$. Thresholds $\underline{v}$ and $\overline{v}$ then satisfy $0 < \underline{v} < \overline{v} < 1$ and equal $\underline{v} = \frac{(b_{SE}+3)(1-b_{SE})+\varphi}{(2+4\rho)(1-b_{SE})+4\varphi}$ and $\overline{v} = \frac{(b_{SE}+2)b_{SE}+3\varphi}{(2+4\rho)b_{SE}+4\varphi}$, such that:*

(a) *If $\rho < \frac{\varphi(1+4b_{SE})+b_{SE}^2}{2\left(3\varphi-b_{SE}^2\right)}$: $\frac{\partial \underline{v}}{\partial b_{SE}} < 0$, $\frac{\partial \overline{v}}{\partial b_{SE}} > 0$ and $\frac{\partial \underline{v}}{\partial b_{SE}} + \frac{\partial \overline{v}}{\partial b_{SE}} < 0$;*

(b) *If $\rho \in \left[\frac{\varphi(1+4b_{SE})+b_{SE}^2}{2\left(3\varphi-b_{SE}^2\right)}, \frac{\varphi(3+4b_{SE})-\left(1-b_{SE}^2\right)}{2\left((1-b_{SE})^2+\varphi\right)}\right]$: $\frac{\partial \underline{v}}{\partial b_{SE}} < 0$, $\frac{\partial \overline{v}}{\partial b_{SE}} \leq 0$ and (thus) $\frac{\partial \underline{v}}{\partial b_{SE}} + \frac{\partial \overline{v}}{\partial b_{SE}} \leq 0$;*

(c) *If $\rho > \frac{\varphi(3+4b_{SE})-\left(1-b_{SE}^2\right)}{2\left((1-b_{SE})^2+\varphi\right)}$: $\frac{\partial \underline{v}}{\partial b_{SE}} \geq 0$ and $\frac{\partial \overline{v}}{\partial b_{SE}} \leq 0$. $\frac{\partial \underline{v}}{\partial b_{SE}} + \frac{\partial \overline{v}}{\partial b_{SE}} \geq 0$ may potentially happen only for high values of b_{SE} strictly above $\frac{1}{2}$.*

In practice, we observe that all three types of organizations co-exist, so Lemma 2 focuses on the most relevant case where all three organizations attract at least some worker types, i.e. $0 < \underline{v} < \overline{v} < 1$ for all possible incentive schemes $b_{SE} \in [0, 1]$ the social enterprise might offer. To limit the number of parameters, we assume $m_{FP} = m_{SE} = m_{NP}$, but let φ be sufficiently large that at least the prosociality types 'sufficiently close' to one of the organizations' mission

[6] We do so as it allows for explicit and simple analytical solutions, while likely losing little in terms of general insights. To see the latter, note that:

$$\frac{\partial V_O}{\partial b_O} = \frac{1}{2}(1 + b_O) - \rho \cdot v - \frac{1}{2} \cdot v \cdot \left[\gamma - \frac{\partial \gamma}{\partial b_O} \cdot (1 - b_O + \gamma \cdot v)\right] \equiv \frac{1}{2}(1 + b_O) - \rho \cdot v - \frac{1}{2} \cdot v \cdot \Gamma.$$

From this, we also have $\frac{\partial^2 V_O}{\partial v \partial b_O} = -\left(\rho + \frac{1}{2} \cdot \Gamma\right) < 0$ (assuming $\frac{\partial \gamma}{\partial b_O} < \frac{\gamma + 2\rho}{1 - b_O + \gamma \cdot v}$). The direction in which a given threshold ($\underline{v}$ or $\overline{v}$) shifts upon an increase in b_O follows from the sign of $\frac{\partial V_O}{\partial b_O}$ evaluated at this threshold. This sign is largely determined by the size of $\rho + \frac{1}{2} \cdot \Gamma$. If this term is small, $\frac{\partial V_O}{\partial b_O} > 0$ at both thresholds, if intermediate $\frac{\partial V_O}{\partial b_O} > 0$ only at the lower threshold $\underline{v}$, while if $\rho + \frac{1}{2} \cdot \Gamma$ is large $\frac{\partial V_O}{\partial b_O} < 0$ at both thresholds. Such variations in $\rho + \frac{1}{2} \cdot \Gamma$ can be qualitatively captured by variations in ρ alone. Note that for $\gamma(b_O, v_O) \equiv 1$ we have $\Gamma = 1$.

ν_O opt for that particular organization.[7] This also requires that ρ is not too large relative to φ, as otherwise workers' dislike of incentives is so intense that all types want to work for the non-profit, except for those with v very close to 0, who prefer the for-profit.

For $b_{SE} = 1$ we intuitively obtain $\underline{v} = \frac{1}{4}$. If the social enterprise offers the exact same steep incentives as the for-profit does, low prosociality types self-select over FP and SE purely on the basis of their mission-match. Because type $v = \underline{v} = \frac{1}{4}$ is equidistant to either FP or SE, all types below are closer to FP and thus prefer the for-profit, while higher types are closer to SE and thus prefer the social enterprise. Similarly, we intuitively have that $\overline{v} = \frac{3}{4}$ if $b_{SE} = 0$ and the social enterprise offers the exact same flat compensation as the non-profit does.

Lemma 2 also delineates when the three different in- and out-selection patterns occur. In the absence of a substantive demotivational component of incentives (ρ is low such that case (a) applies), an increase in b_{SE} makes the social enterprise more attractive across the board; both more selfish types and more prosocial types are then drawn into the social enterprise. The pull towards the SE is then necessarily stronger for more selfish types: $\frac{\partial \underline{v}}{\partial b_{SE}} + \frac{\partial \overline{v}}{\partial b_{SE}} \leq 0$. For ρ moderately large such that case (b) applies, stronger incentives still lead to an in-selection of more selfish types, but an out-selection of more pro-social types. The latter is the 'unintended' effect social entrepreneurs especially worry about in practice. The social enterprise's workforce composition then becomes less pro-social overall, as interval $[\underline{v}, \overline{v}]$ shifts to the left.

Finally, only when ρ is very large as in part (c), an increase in b_{SE} makes the social enterprise less attractive to workers on both sides of the spectrum and the SE workforce shrinks. The more selfish types that initially chose the SE given their dislike of incentives, then (after an increase in b_{SE}) switch to the FP with even stronger incentives because the SE provides less of an additional benefit in terms of this dislike than before. For increases of already steep incentives within the SE (i.e. only for $b_{SE} > \frac{1}{2}$) it may then even happen that the out-selection of more selfish types is a stronger force than the out-selection of the more prosocial types is, such that the SE workforce becomes more prosocial on average. This exceptional case does not seem particularly realistic.[8] Our second hypothesis regarding the selection effect is thus based on the more typical cases in Lemma 2. Proposition 2 makes this explicit and formalizes how the (average) social enterprise effort allocation is affected through the selection channel.

Proposition 2 *(Selection channel) The overall selection effect of a higher b_{SE} on commercial effort corresponds to the change in the average worker type it induces, i.e. $-\frac{\partial}{\partial b_{SE}}\left(\frac{\underline{v}+\overline{v}}{2}\right) = -\frac{1}{2}\left(\frac{\partial \underline{v}}{\partial b_{SE}} + \frac{\partial \overline{v}}{\partial b_{SE}}\right)$. Per Lemma 2, this effect is positive if either ρ is not very large such that case (c) does not apply, or the initial incentive level b_{SE} is not very steep ($b_{SE} \leq \frac{1}{2}$ is sufficient).*

Hypothesis 2 predicting a positive selection effect follows from Proposition 2, assuming the stated boundary conditions are met. This is done in the main text's verbal reasoning by presuming that the most plausible case is the in-selection of more selfish types and the out-selection of more prosocial types after an increase in b_{SE}, i.e. Lemma 2's case (b). In this case, the selection effect is unambiguously positive.

[7] This is solely for simplicity. We could easily allow for base salary differences across organizations, such that e.g. a higher fixed pay is offered the less use is made of incentive pay, i.e. $m_{FP} < m_{SE} < m_{NP}$. We just assume that φ is sufficiently large for proximity to the organization's mission to outweigh these differences in base salaries for at least some types.

[8] Indeed, in the experiment we find that the social enterprise is chosen more often for higher values of incentive pay, i.e. the SE workforce does *not* shrink with b_{SE}.

Appendix B Proofs

Proof of Lemma 1. Moving across our three types of organizations (i.e. moving from NP to either SE or FP, or from SE to FP), an increase in b_O is always accompanied by a decrease in ν_O. From the V_O expression in (3) we obtain $\frac{\partial^2 V_O}{\partial v \partial b_O} = -\rho - \frac{1}{2}\cdot\left[\gamma - \frac{\partial \gamma}{\partial b_O}\cdot(1 - b_O + \gamma\cdot v)\right] < 0$, assuming that $\frac{\partial \gamma}{\partial b_O} < \frac{\gamma+2\rho}{1-b_O+\gamma\cdot v}$. Hence V_O increases in v to a larger extent (or decreases in v to a smaller extent) the smaller piece rate b_O is. Also note from V_O that an increase in ν_O makes organization O more attractive for higher pro-sociality types: $\frac{\partial^2 V_O}{\partial v \partial \nu_O} = 2\varphi + [2(1-b_O) + 4\gamma v]\cdot\frac{\partial \gamma}{\partial \nu_O} \geq 0$. Therefore, if some type $\widehat{v}$ is indifferent between two organizations $A = (m_A, b_{Low}, \nu_{High})$ and $B = (m_B, b_{High}, \nu_{Low})$, all types $v > \widehat{v}$ necessarily prefer organization A with lower piece rate b_{Low} and higher mission ν_{High}. Similarly, all types $v < \widehat{v}$ necessarily prefer organization B.

The final statement follows from $\frac{\partial^2 V_O}{\partial v \partial b_O} < 0$ and $\underline{v} \leq \overline{v}$; note that $\frac{\partial \underline{v}}{\partial b_{SE}} > 0$ holds only if $\frac{\partial V_{SE}}{\partial b_{SE}} < 0$ at $v = \underline{v}$, such that an increase in b_{SE} makes the formerly indifferent (between FP and SE) type strictly worse off under the SE. This implies $\frac{\partial V_{SE}}{\partial b_{SE}} < 0$ at $v = \overline{v}$ as well and therefore $\frac{\partial \overline{v}}{\partial b_{SE}} < 0$. ■

Proof of Lemma 2. Let $w_O \equiv m_O - \varphi\cdot(v-\nu_O)^2 = m_O - \varphi v^2 + 2\varphi v \nu_O - \varphi \nu_O^2$. With $\gamma(b_O, v_O) \equiv 1$ and $m_{FP} = m_{SE} = m_{NP}$ (and $\nu_{FP} = 0$, $\nu_{SE} = \frac{1}{2}$ and $\nu_{NP} = 1$) we then have:

$$w_{SE} - w_{FP} = \varphi\cdot\left(v - \frac{1}{4}\right); \quad w_{NP} - w_{SE} = \varphi\cdot\left(v - \frac{3}{4}\right); \quad w_{NP} - w_{FP} = \varphi\cdot(2v-1).$$

Comparing the worker's value of working for FP relative to NP based on (3), we obtain:

$$\begin{aligned} V_{FP} - V_{NP} &= -(w_{NP} - w_{FP}) + \left[\frac{1 + 2(1-v)}{4} - \rho\cdot v\right] \\ &= -\varphi\cdot(2v-1) + \left[\frac{1+2(1-v)}{4} - \rho\cdot v\right]. \end{aligned}$$

The r.h.s. is strictly decreasing in v and equals zero for indifferent type $\widehat{v} = \left(\frac{3+4\varphi}{2+4\rho+8\varphi}\right)$. It holds that $\widehat{v} \leq 1$ for $\varphi \geq \frac{1-4\rho}{4}$ (thus, under the assumption made on φ). Types below $\widehat{v}$ then strictly prefer FP over NP; the reverse is true for types above $\widehat{v}$. Similarly, comparing SE with NP gives:

$$\begin{aligned} V_{SE} - V_{NP} &= -(w_{NP} - w_{SE}) + \left[\frac{b_{SE}^2 + 2b_{SE}(1-v)}{4} - \rho\cdot v\cdot b_{SE}\right] \\ &= -\varphi\cdot\left(v - \frac{3}{4}\right) + \left[\frac{b_{SE}^2 + 2b_{SE}(1-v)}{4} - \rho\cdot v\cdot b_{SE}\right]. \end{aligned}$$

Again, the r.h.s. decreases in v and equals zero for indifferent type $\overline{v} = \left(\frac{(b_{SE}+2)b_{SE}+3\varphi}{(2+4\rho)b_{SE}+4\varphi}\right)$. It holds that $\overline{v} \leq 1$ for $\varphi \geq b_{SE}(b_{SE} - 4\rho)$; with $b_{SE}(b_{SE}-4\rho) \leq 1$, this holds under the assumption made on φ. For $\varphi > 0$ it also holds that $\overline{v} > 0$ necessarily. Finally, comparing FP with SE gives:

$$\begin{aligned} V_{FP} - V_{SE} &= -(w_{SE} - w_{FP}) + (1 - b_{SE})\left[\frac{1 + b_{SE} + 2(1-v)}{4} - \rho\cdot v\right] \\ &= -\varphi\cdot\left(v - \frac{1}{4}\right) + (1-b_{SE})\left[\frac{1+b_{SE}+2(1-v)}{4} - \rho\cdot v\right]. \end{aligned}$$

The r.h.s. equals zero for indifferent type $\underline{v} = \left(\frac{(b_{SE}+3)(1-b_{SE})+\varphi}{(2+4\rho)(1-b_{SE})+4\varphi}\right)$. It holds that $\underline{v} \leq 1$ for $\varphi \geq \frac{(1-b_{SE})(1+b_{SE}-4\rho)}{3}$; with $\frac{(1-b_{SE})(1+b_{SE}-4\rho)}{3} \leq \frac{1}{4}$, this holds under the assumption made on φ.

For $\varphi > 0$ it also holds that $\overline{v} > 0$ necessarily.

If $\widehat{v} < \underline{v}$, then for all types within $(\widehat{v}, \underline{v})$ we have $NP > FP > SE$ and thus necessarily $\overline{v} \leq \widehat{v}$ must hold. Similarly, If $\widehat{v} > \overline{v}$, then for all types within $(\overline{v}, \widehat{v})$ we have $FP > NP > SE$ and thus necessarily $\widehat{v} \leq \underline{v}$ holds. Hence, we either have $\underline{v} \leq \widehat{v} \leq \overline{v}$ or $\overline{v} \leq \widehat{v} \leq \underline{v}$ (i.e. $\widehat{v}$ always lies in the middle). In the latter case, the social enterprise does not attract any worker types. Now $\underline{v} \leq \overline{v}$ holds iff:

$$\frac{(b_{SE}+3)(1-b_{SE})+\varphi}{(2+4\rho)(1-b_{SE})+4\varphi} \leq \frac{(b_{SE}+2)b_{SE}+3\varphi}{(2+4\rho)b_{SE}+4\varphi}$$

$$\Longleftrightarrow$$

$$(2+4\rho)\left[(1-b_{SE})b_{SE}+\varphi(4b_{SE}-3)\right] \leq 4\varphi\left[2b_{SE}(b_{SE}+2)-3+2\varphi\right].$$

The l.h.s. increases in b_{SE} for $\varphi \geq \frac{2b_{SE}-1}{4}$. The r.h.s. also increases in b_{SE}. The inequality is thus certainly satisfied if the l.h.s. evaluated at its maximum ($b_{SE}=1$) falls short of the minimum of the r.h.s. (i.e. at $b_{SE}=0$): $\varphi(2+4\rho) \leq 4\varphi[-3+2\varphi]$. Rewriting gives $\varphi \geq \frac{\rho}{2}+\frac{7}{4}$.

Turning to how the thresholds vary with b_{SE}, from the expression for $\overline{v}$ we obtain:

$$\frac{\partial \overline{v}}{\partial b_{SE}} = \frac{-(2+4\rho)(3\varphi-b_{SE}^2)+8\varphi(1+b_{SE})}{\left[(2+4\rho)b_{SE}+4\varphi\right]^2} = \frac{2\varphi(1+4b_{SE})+2b_{SE}^2-4\rho(3\varphi-b_{SE}^2)}{\left[(2+4\rho)b_{SE}+4\varphi\right]^2}.$$

This gives $\frac{\partial \overline{v}}{\partial b_{SE}} > (<) 0$ for $\rho > (<) \frac{\varphi(1+4b_{SE})+b_{SE}^2}{2\left(3\varphi-b_{SE}^2\right)}$. The r.h.s. is increasing in b_{SE} and decreasing in φ. With $\varphi > \frac{7}{4}$, $\rho > \frac{35}{34}$ is sufficient for $\frac{\partial \overline{v}}{\partial b_{SE}} < 0$. Similarly,

$$\begin{aligned}\frac{\partial \underline{v}}{\partial b_{SE}} &= \frac{(2+4\rho)\left((1-b_{SE})^2+\varphi\right)-8\varphi(1+b_{SE})}{\left[(2+4\rho)(1-b_{SE})+4\varphi\right]^2} \\ &= \frac{-2\varphi(3+4b_{SE})+2(1-b_{SE}^2)+4\rho\left((1-b_{SE})^2+\varphi\right)}{\left[(2+4\rho)(1-b_{SE})+4\varphi\right]^2}\end{aligned}$$

This gives $\frac{\partial \underline{v}}{\partial b_{SE}} > (<) 0$ if $\rho > (<) \frac{\varphi(3+4b_{SE})-\left(1-b_{SE}^2\right)}{2\left((1-b_{SE})^2+\varphi\right)}$. The r.h.s. is increasing in b_{SE} and equals $\frac{7}{2}$ for $b_{SE}=1$. Hence $\rho > \frac{7}{2}$ is sufficient for $\frac{\partial \underline{v}}{\partial b_{SE}} > 0$. Note that $\frac{\varphi(3+4b_{SE})-\left(1-b_{SE}^2\right)}{2\left((1-b_{SE})^2+\varphi\right)} \geq \frac{\varphi(1+4b_{SE})+b_{SE}^2}{2\left(3\varphi-b_{SE}^2\right)}$ certainly for $\varphi \geq 1$. This yields all elements (a) to (c) of Lemma 2, except for the sign of $\frac{\partial \underline{v}}{\partial b_{SE}} + \frac{\partial \overline{v}}{\partial b_{SE}}$ in parts (a) and (c).

We use two routes to sign $\frac{\partial \underline{v}}{\partial b_{SE}} + \frac{\partial \overline{v}}{\partial b_{SE}}$. The first makes use of the similar structure of $\frac{\partial \underline{v}}{\partial b_{SE}}$ and $\frac{\partial \overline{v}}{\partial b_{SE}}$. Let $A \equiv (2+4\rho)\left((1-b_{SE})^2+\varphi\right)$, $B \equiv 8\varphi(1+b_{SE})$, and$C \equiv (2+4\rho)(3\varphi-b_{SE}^2)$. Note that $A, B, C > 0$ and $A - C < 0$ for $\varphi > \frac{1}{2}$, and thus under the assumption made on φ. Moreover, define $h(x) \equiv \left[(2+4\rho)x+4\varphi\right]^2$, with $h(x) > 0$ strictly increasing in x. With these shorthand notations, we have that $\frac{\partial \underline{v}}{\partial b_{SE}} = \frac{A-B}{h(1-b_{SE})}$ and $\frac{\partial \overline{v}}{\partial b_{SE}} = \frac{-C+B}{h(b_{SE})}$. Straightforward manipulation yields $\frac{\partial \underline{v}}{\partial b_{SE}} + \frac{h(b_{SE})}{h(1-b_{SE})} \cdot \frac{\partial \overline{v}}{\partial b_{SE}} = \frac{A-C}{h(1-b_{SE})} < 0$. Note that $\frac{\partial \underline{v}}{\partial b_{SE}} + \frac{\partial \overline{v}}{\partial b_{SE}} \leq \frac{\partial \underline{v}}{\partial b_{SE}} + \frac{h(b_{SE})}{h(1-b_{SE})} \cdot \frac{\partial \overline{v}}{\partial b_{SE}} < 0$ whenever $\frac{\partial \overline{v}}{\partial b_{SE}} \leq \frac{h(b_{SE})}{h(1-b_{SE})} \cdot \frac{\partial \overline{v}}{\partial b_{SE}}$. For $\frac{\partial \overline{v}}{\partial b_{SE}} > 0$ (cf. case (a)) this occurs when $\frac{h(b_{SE})}{h(1-b_{SE})} \geq 1$, i.e. $b_{SE} \geq \frac{1}{2}$. For $\frac{\partial \overline{v}}{\partial b_{SE}} < 0$ (cf. case (c)) this occurs when $\frac{h(b_{SE})}{h(1-b_{SE})} \leq 1$, i.e. $b_{SE} \leq \frac{1}{2}$.

The range $b_{SE} < \frac{1}{2}$ under case (a) is covered by the second route, which is to work out the inequality $\frac{\partial \underline{v}}{\partial b_{SE}} + \frac{\partial \overline{v}}{\partial b_{SE}} < 0$. Some algebra yields:

$$\frac{(2+4\rho)\left((1-b_{SE})^2+\varphi\right)-8\varphi(1+b_{SE})}{\left[(2+4\rho)(1-b_{SE})+4\varphi\right]^2} + \left[\frac{-(2+4\rho)(3\varphi-b_{SE}^2)+8\varphi(1+b_{SE})}{\left[(2+4\rho)b_{SE}+4\varphi\right]^2}\right] < 0 \quad \Longleftrightarrow$$

$$(2+4\rho)^2\left[2\left(1-b_{SE}\right)^2 b_{SE}^2-\varphi\left(3-6b_{SE}+2b_{SE}^2\right)\right]$$
$$+\left(2+4\rho\right)8\varphi\left[1-3b_{SE}^2-\varphi\left(3-4b_{SE}\right)\right]+16\varphi^2\left[5-6b_{SE}\left(1-b_{SE}\right)-2\varphi\right]<0.$$

With $\varphi>\frac{3}{2}$, the first and second term are necessarily negative for all $b_{SE}\leq\frac{1}{2}$. The l.h.s. is then strictly decreasing in ρ (given $b_{SE}\leq\frac{1}{2}$); therefore, it is sufficient that the inequality holds for $\rho=0$. After rewriting, it reduces to:

$$2\left(1-b_{SE}\right)^2 b_{SE}^2+\varphi\left(-14b_{SE}^2+6b_{SE}+1\right)+8\varphi^2\left[-b_{SE}+3b_{SE}^2\right]<8\varphi^2\left(\varphi-1\right).$$

Evaluating each of the three terms on the l.h.s. at their individual maximum ($b_{SE}=\frac{1}{2}$ for the first, $b_{SE}=\frac{3}{14}$ for the second, and $b_{SE}=\frac{1}{2}$ for the third), the inequality certainly holds if:

$$\frac{1}{8}+\varphi\left(\frac{9}{14}+1\right)+8\varphi^2\left[\frac{1}{4}\right]<8\varphi^2\left(\varphi-1\right).$$

This reduces to $\frac{1}{8}+\varphi\left(\frac{9}{14}+1\right)<8\varphi^2\left(\varphi-\frac{5}{4}\right)$. Under the assumption on φ made, the r.h.s. increases in φ to a larger exent than the l.h.s. does. For $\varphi=\frac{7}{4}$, the inequality is satisfied as $3\leq\frac{49}{4}$. Hence $\frac{\partial\underline{v}}{\partial b_{SE}}+\frac{\partial\overline{v}}{\partial b_{SE}}<0$ for all $b_{SE}\leq\frac{1}{2}$. ■

Proof of Proposition 2. For a general distribution function F on $[0,1]$ (with density f), the overall impact of an increase in b_{SE} on average attention to the commercial task in the SE equals $\frac{\partial\left(\int_{\underline{v}}^{\overline{v}}c^*\cdot\frac{f(v)}{F(\overline{v})-F(\underline{v})}dv\right)}{\partial b_{SE}}$. Using Leibniz' rule we obtain:

$$\begin{aligned}\frac{\partial\left(\int_{\underline{v}}^{\overline{v}}c^*\cdot\frac{f(v)}{F(\overline{v})-F(\underline{v})}dv\right)}{\partial b_{SE}}=&\frac{1}{F\left(\overline{v}\right)-F\left(\underline{v}\right)}\cdot\left\{\int_{\underline{v}}^{\overline{v}}\frac{\partial c^*}{\partial b_{SE}}f\left(v\right)dv+c^*\left(\overline{v}\right)\cdot f\left(\overline{v}\right)\cdot\frac{\partial\overline{v}}{\partial b_{SE}}-c^*\left(\underline{v}\right)\cdot f\left(\underline{v}\right)\cdot\frac{\partial\underline{v}}{\partial b_{SE}}\right\}\\&-\frac{1}{\left(F\left(\overline{v}\right)-F\left(\underline{v}\right)\right)^2}\cdot\left\{f\left(\overline{v}\right)\cdot\frac{\partial\overline{v}}{\partial b_{SE}}-f\left(\underline{v}\right)\cdot\frac{\partial\underline{v}}{\partial b_{SE}}\right\}\cdot\int_{\underline{v}}^{\overline{v}}c^*\left(v\right)f\left(v\right)dv\\&=\underbrace{\int_{\underline{v}}^{\overline{v}}\frac{\partial c^*}{\partial b_{SE}}\cdot\frac{f\left(v\right)}{F\left(\overline{v}\right)-F\left(\underline{v}\right)}dv}_{\text{A: incentive channel}}\\&\underbrace{-\frac{f\left(\overline{v}\right)}{F\left(\overline{v}\right)-F\left(\underline{v}\right)}\cdot\frac{\partial\overline{v}}{\partial b_{SE}}\cdot\left[\int_{\underline{v}}^{\overline{v}}c^*\left(v\right)\cdot\frac{f\left(v\right)}{F\left(\overline{v}\right)-F\left(\underline{v}\right)}dv-c^*\left(\overline{v}\right)\right]}_{\text{B: out-selection of prosocial types}}\\&\underbrace{-\frac{f\left(\underline{v}\right)}{F\left(\overline{v}\right)-F\left(\underline{v}\right)}\cdot\frac{\partial\underline{v}}{\partial b_{SE}}\cdot\left[c^*\left(\underline{v}\right)-\int_{\underline{v}}^{\overline{v}}c^*\left(v\right)\cdot\frac{f\left(v\right)}{F\left(\overline{v}\right)-F\left(\underline{v}\right)}dv\right]}_{\text{C: in-selection of extrinsically-motivated types}}.\end{aligned}$$

For $F\sim U[0,1]$, we have $\frac{f(\underline{v})}{F(\overline{v})-F(\underline{v})}=\frac{1}{\overline{v}-\underline{v}}$ and, given the linearity of $c^*\left(v\right)$ in v from equation (2), $\int_{\underline{v}}^{\overline{v}}c^*\left(v\right)\cdot\frac{f(v)}{F(\overline{v})-F(\underline{v})}dv=c^*\left(\frac{\underline{v}+\overline{v}}{2}\right)=\frac{c^*(\underline{v})+c^*(\overline{v})}{2}$. Term B thus reduces to $\frac{1}{\overline{v}-\underline{v}}\cdot\frac{\partial\overline{v}}{\partial b_{SE}}\cdot\left[\frac{c^*(\underline{v})-c^*(\overline{v})}{2}\right]$ and term C to $\frac{1}{\overline{v}-\underline{v}}\cdot\frac{\partial\underline{v}}{\partial b_{SE}}\cdot\left[\frac{c^*(\underline{v})-c^*(\overline{v})}{2}\right]$. For $\gamma\equiv1$, we have $\frac{c^*(\underline{v})-c^*(\overline{v})}{2}=\frac{\overline{v}-\underline{v}}{2}$ per equation (2), yielding that the overall selection effect equals $-B-C=-\frac{1}{2}\left(\frac{\partial\underline{v}}{\partial b_{SE}}+\frac{\partial\overline{v}}{\partial b_{SE}}\right)$. ■

Appendix C Instructions

Welcome Thank you for taking part in this study designed to learn about how people make decisions.[9] The choices we ask you to make are based on methods and techniques from economics, business administration, and psychology. The experiment is expected to take around 20 minutes. At the top of your screen, you will be able to see what percentage of the questionnaire you have already completed.

Personal data will be kept confidential. Your answers will only be used for this research. The published results will not refer to a person by name and will not describe individual choices. We will not disclose information to third parties. Aside from the participation fee, this experiment allows you to earn additional money. Out of all respondents who completed the questionnaire, we will randomly select 40 respondents for payment; as we expect around 800 participants in this experiment, the chance of being chosen for payment is about 1/20. Depending on the choices made, those chosen for payment can earn up to £80. Your cooperation is greatly appreciated.

The task This experiment will consist of four parts. In Parts 2 and 3, you will perform a slider task. This consists of a screen with 15 horizontal sliders, labeled from 0 to 100. As shown below, each slider is initially positioned at 50 and can be moved towards 0 or 100. These labels carry no inherent value, and only provide an axis for the slider. Each slider has a number to the right of it showing its current position. You can use your mouse in any way you like to move each slider. You can readjust the position of each slider as many times as you wish. You will have a chance to practice moving sliders shortly.

The task involves making a choice between placing the slider at 25 or 75, each corresponding to an action and a payoff that will be made clear at the beginning of each slider round. Placing the slider at 25 will always generate a payoff to you. Placing the slider at 75 will always generate a payoff to a good cause (which you will select below). Each slider will carry a payoff between £0 and £1, stated explicitly whenever you perform the task. Note that a slider will be taken into consideration for your payoff only if positioned at *exactly* 25 or 75, as accuracy is valued.

Figure C.1: Slider example.

Payment You will receive a fixed fee of £3 for completing this experiment. In addition, you can earn money in each of the rounds where you will perform the slider task, with the exception of the practice round. Finally, you can earn money in some of the questions asked towards the end of the experiment. It will always be clearly indicated if and how (much) money can be earned. When the experiment has concluded, a number of participants will be randomly selected as winners and will receive payment. We expect around 800 participants, and only 40 will be selected for payment: we therefore expect around 1 in 20 participants to be selected for payment. Throughout the experiment, you will also have the chance to earn money for both yourself and a good cause. You will have a chance to select a good cause below. If you are selected for payment, any payoff you have generated for the good cause will be transferred to that particular good cause.

[9] A preview can be accessed here.

Part 1 – Good cause and practice

This part of the experiment allows you to choose your preferred good cause and become familiar with the task. Please proceed to the next screen.[10]

Please choose one of the three organizations below corresponding to your preferred good cause:

- **Workforce reintegration**: The Big Issue Foundation seeks to promote the social and financial inclusion of its vendors by identifying and motivating individuals to engage with the services that will help them move forward and deal with their homelessness and health issues and achieve their own goals.
- **Fair and equitable trade**: Fairtrade Foundation seeks to connect disadvantaged producers and consumers, promote fairer trading conditions through standardization and certification and empower producers to combat poverty, strengthen their position, and take more control over their lives.
- **Water quality and environment**: Water Aid seeks to deliver clear water, improved sanitation, and proper hygiene to developing countries through a combination of technical solutions and hygiene education. They aim to ensure the effectiveness of their projects by using carbon-neutral, sustainable methods that preserve the environment.

You now have a chance to practice moving sliders. Please remember that a slider is considered correctly placed only if placed at exactly 25 or exactly 75. The numbers only represent positions that correspond to actions providing an own payoff and a payoff to the previously selected good cause. These 2 sliders are given for you to become familiar with the task. You will not be paid for this practice round. Please keep in mind that in the actual task you will position sets of 15 sliders. When you are sufficiently familiar with this task, please proceed to the next screen.

Comprehension check Before proceeding to the actual task, please answer the following questions.

If each slider placed at 25 produces an own payoff of 5 points, and each slider placed at 75 produces a good cause payoff of 20 points, what is the good cause payoff when 5 sliders are placed at 75?

- 100
- 25
- 40

If each slider placed at 25 produces an own payoff of 10 points, and each slider placed at 75 produces a good cause payoff of 20 points, what is the good cause payoff when 2 sliders are placed at 75?

- 100
- 25
- 40

If each slider placed at 25 produces an own payoff of 5 points, and each slider placed at 75 produces a good cause payoff of 20 points, what is your payoff when 5 sliders are placed at 25?

- 100
- 25
- 40

This is the end of Part 1. Please proceed to the next screen.

[10] Randomization is employed with regards to the order of: i) good causes, ii) attention check questions (and options), iii) contracts in Part 2, and iv) contracts in the choice question in Part 3.

Part 2 – All contracts

You will now be performing the slider task under a set of different contracts. Throughout the experiment you will behave as an employee of a set of companies. These companies provide similar services, but have different objectives, as explained at the beginning of each scenario. **Please read the company descriptions carefully**. Moving the slider will allow you to make choices as an employee of those companies. The actions described within each contract will correspond to potential actions of employees of such companies, and generate either a payoff to you or the good cause you selected earlier. The text will explain clearly how the payoffs are generated. Please proceed to the next screen.

Workforce reintegration (if selected)

For-profit Imagine you are working for a company providing garbage collection services on the market. **The company only cares about generating revenue through the sale of services.**

- By placing the slider exactly at position 25 you can generate revenue for the company through the sale of goods and services; each slider you position at 25 will give you a payoff of £1.
- By placing the slider exactly at position 75 you can contribute to the professional development of its employees; each slider you position at 75 will give a payoff of £1 to the good cause.[11]

Please place the 15 sliders below as you see fit.

Nonprofit Imagine you are working for a company aimed at reintegrating long term unemployed people into the workforce by hiring them to provide garbage collection services. **The company only cares about workforce reintegration through the professional development of the long term unemployed.**

- By placing the slider exactly at position 25 you can generate revenue for the company through the sale of services; each slider you position at 25 will give you a payoff of £0.
- By placing the slider exactly at position 75 you can contribute to the professional development of its employees; each slider you position at 75 will give a payoff of £1 to the good cause.

Please place the 15 sliders below as you see fit.

Social enterprise Imagine you are working for a company aimed at reintegrating long term unemployed people into the workforce by hiring them to provide garbage collection services that are then sold on the market. **It is in the best interest of the organization that both ensuring the professional development of the long term unemployed and generating revenue through the sale of services receive attention from employees.**

- By placing the slider exactly at position 25 you can generate revenue for the company through the sale of services; each slider you position at 25 will give you a payoff of £0/£0.25/£0.50/£1.
- By placing the slider exactly at position 75 you can contribute to the professional development of its employees; each slider you position at 75 will give a payoff of £1 to the good cause.

Please place the 15 sliders below as you see fit.

[11] The availability of the social task in for-profit contracts approximates the possibility of CSR activities.

Fair and equitable trade (if selected)

For-profit Imagine you are working for a company investing in new businesses. **The company only cares about generating returns for its investors by selecting the most promising ventures.**

- By placing the slider exactly at position 25 you can generate returns for the company by investing in the most promising ventures; each slider you position at 25 will give you a payoff of £1.
- By placing the slider exactly at position 75 you can make finance accessible to fair trade businesses in developing countries; each slider you position at 75 will give a payoff of £1 to the good cause.

Please place the 15 sliders below as you see fit.

Nonprofit Imagine you are working for a company dedicated to investing in new businesses. **The company only cares about alleviating poverty by supporting fair trade businesses in developing countries.**

- By placing the slider exactly at position 25 you can generate returns for the company by investing in the most promising ventures; each slider you position at 25 will give you a payoff of £0.
- By placing the slider exactly at position 75 you can make finance accessible to fair trade businesses in developing countries; each slider you position at 75 will give a payoff of £1 to the good cause.

Please place the 15 sliders below as you see fit.

Social enterprise Imagine you are working for a company investing in new businesses. **It is in the best interests of the organization that both offering fair trade businesses in developing countries access to loans and credit facilities and ensuring a positive rate of return on investments receive attention.**

- By placing the slider exactly at position 25 you can generate returns for the company by investing in the most promising ventures; each slider you position at 25 will give you a payoff of £0/£0.25/£0.50/£1.
- By placing the slider exactly at position 75 you can make finance accessible to fair trade businesses; each slider you position at 75 will give a payoff of £1 to the good cause.

Please place the 15 sliders below as you see fit.

Water quality and environment (if selected)

For-profit Imagine you are working for a company providing water services to a variety of other organizations on the market. **The company only cares about generating revenue by expanding market access.**

- By placing the slider exactly at position 25 you can generate revenue for the company by expanding market access; each slider you position at 25 will give you a payoff of £1.
- By placing the slider exactly at position 75 you can reduce the carbon emissions resulting from product packaging and delivery; each slider you position at 75 will give a payoff of £1 to the good cause.

Please place the 15 sliders below as you see fit.

Nonprofit Imagine you are working for a company providing environmentally sustainable water services to a variety of other organizations. **The company only cares about having an environmentally friendly product, with minimal carbon emissions and fully recyclable packaging.**

- By placing the slider exactly at position 25 you can generate revenue for the company by expanding market access; each slider you position at 25 will give you a payoff of £0.
- By placing the slider exactly at position 75 you can reduce the carbon emissions resulting from product packaging and delivery; each slider you position at 75 will give a payoff of £1 to the good cause.

Please place the 15 sliders below as you see fit.

Social enterprise Imagine you are working for a company providing environmentally sustainable water services to a variety of other organizations on the market. **It is in the best interests of the organization that both ensuring that production and delivery are done with minimal environmental impact and increasing revenues by expanding market access receive attention.**

- By placing the slider exactly at position 25 you can generate revenue for the company by expanding market access; each slider you position at 25 will give you a payoff of £0/£0.25/£0.50/£1.
- By placing the slider exactly at position 75 you can reduce the carbon emissions resulting from product packaging and delivery; each slider you position at 75 will give a payoff of £1 to the good cause.

Please place the 15 sliders below as you see fit.

This is the end of Part 2. Please proceed to the next screen.

Part 3 – Preferred contract

You will now perform the slider task once more. However, this time you can choose your preferred contract from the ones in Part 2. Please proceed to the next screen.

Workforce reintegration (if selected)

Which contract would you like to perform the slider task in?

- The company provides garbage collection services on the market and cares only about generating revenue. Placing the slider at exactly 25 produces an own payoff of £1. Placing the slider at exactly 75 produces a payoff of £1 for the good cause.
- The company reintegrates the unemployed into the workforce by hiring them to provide garbage collection services and cares only about the professional development of its employees. Placing the slider at exactly 25 produces an own payoff of £0. Placing the slider at exactly 75 produces a payoff of £1 for the good cause.
- The company reintegrates unemployed people into the workforce by hiring them to provide garbage collection services that are sold on the market. It is in the best interests of the organization that both generating revenue and aiding the professional development of its employees receive attention. Placing the slider at exactly 25 produces an own payoff of £0/£0.25/£0.50/£1. Placing the slider at exactly 75 produces a payoff of £1 for the good cause.

Fair and equitable trade (if selected)

Which contract would you like to perform the slider task in?

- The company invests in the most promising new businesses and cares only about generating returns for investors. Placing the slider at exactly 25 produces an own payoff of £1. Placing the slider at exactly 75 produces a payoff of £1 for the good cause.
- The company provides access to finance for fair trade businesses in developing countries and cares only about poverty alleviation. Placing the slider at exactly 25 produces an own payoff of £0. Placing the slider at exactly 75 produces a payoff of £1 for the good cause.
- The company provides access to finance for fair trade businesses in developing countries with a positive rate of return on investments. It is in the best interests of the company that both generating positive returns and poverty alleviation receive attention. Placing the slider at exactly 25 produces an own payoff of £0/£0.25/£0.50/£1. Placing the slider at exactly 75 produces a payoff of £1 for the good cause.

Water quality and environment (if selected)

Which contract would you like to perform the slider task in?

- The company provides water services to other organizations on the market and cares only about generating revenue. Placing the slider at exactly 25 produces an own payoff of £1. Placing the slider at exactly 75 produces a payoff of £1 for the good cause.
- The company provides water services to other organizations and cares only about environmental sustainability. Placing the slider at exactly 25 produces an own payoff of £0. Placing the slider at exactly 75 produces a payoff of £1 for the good cause.
- The company provides water services to other organizations on the market. It is in the best interest of the organization that both generating revenue and environmental sustainability receive attention. Placing the slider at exactly 25 produces an own payoff of £0/£0.25/£0.50/£1. Placing the slider at exactly 75 produces a payoff of £1 for the good cause.

[Subjects then perform slider task in chosen contract.]
This is the end of Part 3. Please proceed to the next screen.

Part 4 – Questionnaire

Before you complete the experiment, please answer the following questions.

What is your gender?

- Male
- Female

What is your age?

- Under 25
- 25-34
- 35-44
- 45-54
- 55 or older

What is the highest level of schooling you have completed?

- High school diploma
- Bachelor degree

- Master degree
- Other

Are you currently as student?

- Yes
- No

What is your educational background?

- Economics and business
- Arts, architecture, and design
- Science, technology, engineering, and mathematics
- Law, social sciences, and humanities
- Medicine, health, and care
- Other

What was your gross income (across all sources of income) in 2018? *We understand this information is sensitive. Therefore, if you want, you can keep it private. However, this information may help us understand differences in economic decision-making.*

- Less than £10,000
- Between £10,000 and £25,000
- Between £25,000 and £50,000
- Between £50,000 and £75,000
- More than £75,000
- I prefer not to answer this question

For this question, you will be randomly and anonymously paired with another participant. **This question allows you to earn some money**. You are endowed with £10 and have to decide how much of the £10 you would like to share with the other participant. Please enter the amount you would like to give to the other participant below (you will keep the remainder of the £10 for yourself). *Values between 0 and 10 are allowed, up to two decimals (e.g. 9.99 or 0.01).* **If selected for payment, you will receive the amount you chose to keep and the randomly selected participant will receive the amount you chose to give. At the same time, you will also be randomly paired with another participant deciding how to share £10 with you. If selected for payment, you will also receive this amount. Note that the person you give to and the person that gives to you will not be the same person.** [*Altruism*]

- ...

Imagine a similar situation to the one just described (i.e. sharing £10), with three differences. First, in this case, you are the recipient. Second, you can choose to refuse the amount received if you consider it inappropriate. However, if you refuse the amount, neither you or the other person would receive any payoff. **Note that this question will not earn you money**. What would be the minimum amount offered that you would accept? *Values between 0 and 10 are allowed, up to two decimals (e.g. 9.99 or 0.01).* [*Inequality aversion*]

- ...

For each of the statements below, please select the option that best describes you. *The options are: strongly disagree, somewhat disagree, neither agree nor disagree, somewhat agree, and strongly agree.* [*Compassion*]

- I am rarely moved by the plight of the underprivileged. (reverse coded)

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

Imagine you won £1,000 in a lottery. Considering your current situation, how much would you donate to a good cause? *Values between 0 and 1,000 are allowed, up to two decimals (e.g. 999.99 or 0.01).* [*Hypothetical altruism*]

- ...

How do you assess your willingness to share with others without expecting anything in return when it comes to a good cause? *Please use a scale from 0 to 10, where 0 means you are "completely unwilling to share" and 10 means you are "very willing to share". You can use values in between to indicate where you fall on the scale.* [*Willingness to share*]

- ...

Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks? *Please use a scale from 0 to 10, where 0 means you are "completely unwilling to take risks" and 10 means you are "fully prepared to take risks". You can use values in between to indicate where you fall on the scale.* [*Risk preferences*]

- ...

How willing are you to give up something that is beneficial for you today in order to benefit more from that in the future? *Please use a scale from 0 to 10, where 0 means you are "completely unwilling to give up something today" and 10 means you are "fully prepared to give up something today". You can use values in between to indicate where you fall on the scale.*[*Time preferences*]

- ...

How would you label an organization with both economic and social value creation goals?

- ...

How would you perceive a social enterprise that introduces performance bonuses?

- ...

What was the own payoff per slider offered by the social enterprise contract?

- £0/£0.25/£0.50/£1 [manipulation check]

In the past, have you:

- Been employed by a non-profit? Yes/no
- Been employed by a social enterprise? Yes/no

How often do you:

- Volunteer? Rarely/often
- Donate to social organizations? Rarely/often
- Work professionally with social organizations? Rarely/often [*Prosocial behavior*]

Appendix D Experimental Checks

Sample and summary statistics This appendix presents a set of basic checks on the data generated through our experiment, whose design we summarize in Table 1. We impose several restrictions on our main sample in order to ensure the highest quality of data. First, we consider sliders placed at 23-27 and 73-77 as correct, indicating a clear intention to exert commercial or social effort; we then require that subjects have placed at most 10 of the 60 sliders they perform overall outside of these ranges. Second, we require subjects to pass at least one of the attention and manipulation checks. Third, we require our gender variables and the one offered by Prolific to be in agreement. Fourth, we require subjects not to complete the experiment in less than 10 minutes or more than 40 minutes (potential signals of lack of attention). These criteria leave us with a sample of 708 subjects, although Table F.1 shows our findings are robust to tightening or relaxing these restrictions. Table 2 provides summary statistics for our sample. Half of the participants are women, 16.5% are students, 40% have a bachelor degree, and 13% have a master degree. Respondents span the income and age ranges, although a large share are below age 44 and below £50,000 annual income. Average *Compassion* is 29.19 out of 40 (very similar to the average of 28.19 in the original study by Perry, 1996) and subjects share on average £4.2 in the dictator game (with a large fraction sharing exactly £5). Subjects report being willing to share without expecting anything in return (mean 6.4 on a 0-10 scale), and 24.4% and 10.5% report previous non-profit or social enterprise employment. Table D.3 further shows that most social preference measures are positively correlated. Subjects are moderately willing to take risks (mean 5.1 on a 0-10 scale), but are willing to give something up today in order to benefit in the future (mean 6.5 on a 0-10 scale). Finally, subjects completed the experiment in around 18.5 minutes on average, although substantial variation exists.

Good cause (mission) choice To ensure the saliency of the social task, we allowed subjects to choose their preferred good cause from the options: workforce reintegration, fair and equitable trade, and water quality and environment. This choice of good cause informs the company descriptions (mission) that subjects face. Subjects' choices are shown in Table D.1, by treatment. Table D.4 performs a multinomial logit estimation of the choice of mission on demographics, social preferences, and dummies for the treatment subjects were allocated to (i.e. £0.25, £0.50, and £1 incentive levels, against a £0 baseline). The results suggest that social preferences and demographics are largely uncorrelated with the choice of good cause. Subjects in the £0.25 and £0.50 treatments were less likely to select a workforce reintegration or fair trade mission, preferring an environmental mission instead, and the treatment dummies are jointly significant ($p = 0.023$). However, a χ^2 test cannot reject the independence of mission and treatment ($p = 0.111$). The latter is consistent with the structure of the experiment, as subjects were not aware of the treatment they were randomly allocated to (i.e., the social enterprise bonus) when they chose the good cause. To alleviate any concerns regarding the endogenous nature of the mission choice, our robustness checks using regression analyses include mission choice dummies, essentially comparing within groups of individuals choosing the same good cause. Moreover, the results are virtually the same across the different social missions (Online Appendix I).

Randomization check We assess whether our randomization procedure has been successful by estimating a set of regressions of various demographics and social preferences on treatment dummies in Table D.5. The £0.25 treatment has a slightly larger share of individuals with income between £25,000 and £50,000, and subjects took longer to practice the slider task in the £0.50 and £1 treatments. These significant coefficients are within the bounds of the number of significant effects appearing by chance and become insignificant when performing multiple hypothesis testing adjustments (Romano and Wolf, 2005; List et al., 2019). For other demographics and social preferences the dummies are jointly insignificant (all $p > 0.25$), and

produce a poor fit of the data (all $R^2 < 0.01$). This is true not only for the main analysis sample ($N = 708$), but also when we use all the available observations ($N = 796$). Overall, this analysis suggests that our randomization has been successful.

Randomization inference Recent work by Young (2019) suggests that statistical inference in experiments is often based on potentially inappropriate asymptotic theory, occasionally resulting in standard errors that are too small and an over-rejection of null hypotheses. Instead, he proposes the use of randomization inference, where treatment is randomly re-allocated among experimental observations, with 1,000 permutations. The analysis is performed each time, producing an empirical distribution of the coefficients of interest, against which the obtained coefficient can be compared, producing an empirical p-value equivalent (Heß, 2017). The results in Table D.2 show the results of randomization inference analysis performed for social enterprise *Social effort* without sorting in columns (1)-(4), social enterprise *Social effort* with sorting in columns (5)-(8), and social enterprise employees' *Compassion* and *Social motivation* in columns (9)-(12). Odd columns present the experimental p-values and even columns present the equivalent randomization inference p-values, obtained from regressing these outcomes on treatment dummies. These regressions are similar to the pairwise analyses in Tables 3 and 4, but consider all treatments jointly. Social effort results are identical across methods, while the differences are very small for social preferences, reinforcing our finding that high-powered incentives lead less prosocial employees to select into social enterprises.

Table D.1: **Choice of Good Cause**

Treatment	Workforce reintegration	Fair and equitable trade	Water quality and environment	N
£0	50	36	84	170
£0.25	41	24	113	178
£0.50	39	36	110	185
£1	47	36	92	175
Total	177	132	399	708

Number of participants in each treatment that selected the given good cause.

Table D.2: **Experimental and Randomization Inference p-values**

	Social effort w/o sorting				Social effort w/ sorting				Motivation w/ sorting, strata			
	W/o strata		W/ strata		W/o strata		W/ strata		Compassion		Social motiv.	
	EXP (1)	RI (2)	EXP (3)	RI (4)	EXP (5)	RI (6)	EXP (7)	RI (8)	EXP (9)	RI (10)	EXP (11)	RI (12)
£0.25	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.475	0.477	0.441	0.461
£0.50	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.695	0.707	0.558	0.583
£1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.051	0.055	0.003	0.004
N	708		708		341		341		341		341	

In each column, the dependent variable (social effort, compassion, social motivation) is regressed on treatment dummies, with the £0 treatment as baseline. EXP = experimental p-value; RI = randomization inference p-value; sorting denotes the condition where subjects can choose their preferred contract; strata denotes that both experimental and randomization inference analyses account for stratification by gender. Results using robust standard errors or adding controls are similar.

Table D.3: **Pairwise correlations**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
(1) Female	1.000															
(2) Student	0.014	1.000														
(3) Education	0.058	0.066	1.000													
(4) Income	-0.004	0.237	0.153	1.000												
(5) Age	-0.040	-0.433	-0.085	-0.283	1.000											
(6) Compassion	0.150	0.088	0.068	0.010	0.021	1.000										
(7) Altruism	0.046	0.064	-0.038	0.034	-0.060	0.136	1.000									
(8) Inequality aversion	0.017	0.084	0.028	0.031	-0.070	-0.135	0.074	1.000								
(9) Hypothetical altruism	0.019	0.075	-0.014	0.031	0.008	0.202	0.265	0.031	1.000							
(10) Willingness to share	0.023	0.055	-0.002	-0.000	0.009	0.399	0.172	0.005	0.341	1.000						
(11) Prosocial behavior	0.129	0.127	0.036	0.071	-0.055	0.233	0.141	-0.002	0.257	0.304	1.000					
(12) Risk taking	-0.159	-0.077	-0.015	0.117	0.016	0.021	0.051	0.032	0.088	0.148	0.110	1.000				
(13) Time discounting	-0.063	0.036	0.017	0.055	-0.054	0.119	0.051	-0.086	0.153	0.321	0.140	0.183	1.000			
(14) Practice time	0.009	-0.014	-0.036	-0.049	-0.010	-0.050	0.041	-0.009	0.014	0.036	0.002	0.002	-0.001	1.000		
(15) Comprehension time	0.006	-0.021	0.097	-0.018	-0.021	-0.047	0.017	0.101	0.026	0.031	-0.000	0.029	0.007	0.155	1.000	
(16) Questions time	-0.033	0.021	0.046	-0.021	-0.017	-0.034	-0.017	-0.034	0.028	0.017	-0.017	0.006	-0.004	0.089	0.095	1.000

All pairwise correlations larger than 0.074 (in absolute terms) are significant at the 5% level.

Table D.4: **Choice of Good Cause: Multinomial Logit**

	Workforce reintegration		Fair and equitable trade	
	Coefficient	(s.e.)	Coefficient	(s.e.)
Compassion	0.053*	(0.023)	0.040	(0.024)
Altruism	-0.037	(0.047)	-0.046	(0.050)
Inequality aversion	-0.010	(0.044)	0.102**	(0.046)
Hypothetical altruism	0.000	(0.001)	0.000	(0.001)
Willingness to share	-0.056	(0.053)	-0.036	(0.054)
Prosocial behavior	0.071	(0.089)	0.054	(0.099)
Risk taking	0.029	(0.040)	0.034	(0.044)
Time preferences	0.027	(0.047)	0.073	(0.052)
Treatment = £0.25	-0.628**	(0.269)	-0.830***	(0.315)
Treatment = £0.50	-0.627**	(0.274)	-0.391	(0.292)
Treatment = £1	-0.180	(0.268)	-0.110	(0.295)
Female	0.311	(0.206)	-0.031	(0.227)
Student	-0.843**	(0.363)	-0.220	(0.340)
High school diploma	0.249	(0.322)	-0.004	(0.348)
Bachelor degree	-0.494	(0.329)	-0.426	(0.356)
Master degree	-0.243	(0.396)	-0.238	(0.439)
< £10,000	1.108*	(0.566)	0.847	(0.566)
£10,000 – £25,000	0.956	(0.543)	0.725	(0.546)
£25,000 – £50,000	0.930	(0.551)	0.579	(0.560)
£50,000 – £75,000	0.907	(0.617)	-0.043	(0.695)
> £75,000	0.740	(0.773)	-0.226	(0.947)
Age 25-34	-0.302	(0.323)	-0.771**	(0.340)
Age 35-44	-0.114	(0.350)	-0.129	(0.359)
Age 45-54	0.413	(0.376)	0.067	(0.404)
Age > 55	-0.094	(0.465)	-1.015*	(0.595)
Constant	-2.523***	(1.035)	-2.599***	(1.003)
N	708			
LR χ^2 (p-value)	85.25 (0.018)			
Pseudo-R^2	0.061			
Social preferences: χ^2 (p-value)	17.18 (0.374)			
Treatment levels: χ^2 (p-value)	11.34 (0.023)			

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors in parentheses. The baseline mission is the water quality and environment mission. For categorical variables the baselines are high school education, income < £10,000, age 18-25, and the £0 treatment. We do not report dummies for field of education for brevity, though none are significant. Although the mission was chosen in advance of subjects being aware of the bonus offered by the social enterprise contract (the treatment), a simple χ^2 test rejects the independence of mission and treatment with $p = 0.111$.

Table D.5: **Randomization Check**

	Treatment					
Characteristic	(1) £0.25	(2) £0.50	(3) £1	(4) p-value	(5) N	(6) R^2
Compassion	-0.309	0.061	-0.508	0.663	708	0.002
	(0.524)	(0.538)	(0.506)			
Altruism	0.027	-0.102	-0.007	0.946	708	0.001
	(0.248)	(0.241)	(0.237)			
Inequality aversion	0.007	0.000	-0.155	0.861	708	0.001
	(0.251)	(0.254)	(0.236)			
Hypothetical altruism	18.673	24.416	17.190	0.496	708	0.003
	(17.460)	(16.812)	(15.998)			
Willingness to share	-0.243	0.033	-0.272	0.476	708	0.003
	(0.262)	(0.257)	(0.247)			
Prosocial behavior	-0.155	0.033	-0.066	0.587	708	0.002
	(0.124)	(0.136)	(0.126)			
Social motivation (factor)	-0.055	0.029	-0.055	0.658	708	0.002
	(0.082)	(0.084)	(0.077)			
Risk taking	0.158	-0.009	-0.077	0.815	708	0.001
	(0.265)	(0.275)	(0.267)			
Time preferences	0.378	0.226	0.127	0.427	708	0.004
	(0.234)	(0.233)	(0.225)			
Age	0.073	-0.008	-0.163	0.269	708	0.005
	(0.124)	(0.123)	(0.124)			
Female	-0.011	-0.008	0.014	0.965	708	0.000
	(0.054)	(0.053)	(0.054)			
Student	-0.007	0.003	0.030	0.818	708	0.001
	(0.039)	(0.039)	(0.041)			
Bachelor degree	-0.052	-0.001	-0.023	0.717	708	0.002
	(0.052)	(0.053)	(0.053)			
Master degree	0.028	0.007	0.002	0.862	708	0.001
	(0.036)	(0.035)	(0.035)			
Low income	-0.048	-0.030	-0.028	0.843	708	0.001
	(0.053)	(0.053)	(0.054)			
Medium income	0.084*	0.077	0.050	0.291	708	0.005
	(0.049)	(0.048)	(0.048)			
High income	0.030	0.020	0.032	0.726	708	0.002
	(0.032)	(0.031)	(0.032)			
Practice time	2.677	5.863***	6.629*	0.028	708	0.007
	(2.408)	(2.097)	(3.881)			
Comprehension time	-4.894	-1.919	-5.440	0.499	708	0.004
	(4.136)	(4.293)	(4.110)			
Questions time	2.530	-0.294	1.246	0.675	708	0.002
	(2.767)	(2.041)	(2.532)			

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses. Each row presents results from a separate model, regressing the given characteristic on treatment dummies, with p-values from tests of joint significance. The omitted education categories are completed high school and other degrees; the omitted income category comprises those who prefer not to answer. Estimating a series of seemingly unrelated regressions produces similar results. Multiple hypothesis testing adjustments (Romano and Wolf, 2005; List et al., 2019) eliminate the significant coefficients.

Appendix E Power Calculations

To ensure we are able to detect meaningful changes in social task effort, we performed a series of *ex ante* power calculations. In other words, given the sample sizes we target, the expected means and standard deviations, and the required significance level, what is the statistical power to detect a given effect? Conversely, what is the smallest effect we can detect while still maintaining statistical power above the conventional 80%? The purpose of this section, therefore, is to ensure that our experiment is able to maximize power and minimize the effect sizes it can detect.

Following List et al. (2011), for independent groups with means μ_a and μ_b, standard deviations $\sigma_a = \sigma_b = \sigma$, sample sizes N_a and N_b, significance level α, and detectable effect size δ, statistical power $1 - \beta$ for a two-sided test is calculated to satisfy:

$$\delta = (t_{\alpha/2} + t_\beta)\sqrt{\frac{\sigma_a^2}{N_a} + \frac{\sigma_b^2}{N_b}} \tag{E.1}$$

Equation E.1 shows that the effect size δ we can detect increases with the required significance level (i.e. we can detect larger effects at 5% than at 1% significance) and the standard deviations of the outcomes (i.e. the lower the underlying heterogeneity, the smaller the effect we can detect), but decreases with sample size (i.e. the more observations, the smaller the effect we can detect). The formula also shows that δ and t_β (and, as a result, $1 - \beta$) are positively correlated, which implies that small effect sizes can only be detected when there is more statistical power. While statistical power rises with sample size, budget constraints limit this avenue, highlighting the trade-off between power and effect size. The results below provide a set of assumptions regarding sample sizes, means, standard deviations, and significance, in order to assess the relationship between effect size and power in our experiment.

In Panel A of Table E.1, we consider comparisons of social enterprise (SE) social effort across treatments, with sample size 200 per group, fixing one sample mean to 7.5 as our expectation of a fully balanced effort allocation, and varying the other to achieve various δ levels and standardized effect sizes $0.2 < \delta/\sigma < 1$.[12] Based on a pilot experiment, we set $\sigma \in \{1, 1.5, 2, 2.5\}$. Results in column (8) suggest that we have substantial power (below the conventional 80% only in the most conservative settings) to detect small effect sizes in two-sided tests. For similar comparisons across treatments when sorting is permitted, we expect social enterprise sample sizes around 120-150, such that power is close to that in Panel A.

In Panels B and C, we consider comparisons between the largest group (social enterprise) and smallest group (for-profit/nonprofit) within the £0.50 and £1 treatments, as suggested by the pilot experiment.[13] The larger sample size we expect for the social enterprise group is in line with the higher expected variance in this group (List et al., 2011), although the ratio of variances is perhaps smaller. Small variations in sample size or variance (between groups) do not affect the main conclusions, namely that unless standard deviations are very large ($\sigma > 1.5$), our tests significantly detect a 1-unit change in effort allocation with power $1 - \beta > 80\%$.

For completeness, in column (10) we show the minimum effect size δ_{min} for a given standard deviation in two-sided tests with 80% power. In these two-sided tests, we are virtually always able to detect changes of $\delta/\sigma \geq 0.6$. Note that so far we have used two-sided tests in our power calculations in order to be conservative. As our hypotheses are mostly one-sided, we calculate power for such tests in column (9): as expected, these tests are even more powerful.

Some of the comparisons we perform (for example, between contracts without sorting) represent dependent samples; in the case of such within-subject comparisons, power is expected to be at least as high (List et al., 2011). We confirm this result in Figure E.1, where we calculate the

[12] There is no *ex ante* reason to expect different variance in SE social effort across treatments when sorting is not allowed; therefore, we opted for equal samples across treatments (List et al., 2011). F-tests based on Table 3 data show that, *ex post*, the assumption of equal variance of social effort is valid.

[13] The latter groups are smaller, but their mean difference is expected to be large, so power is retained.

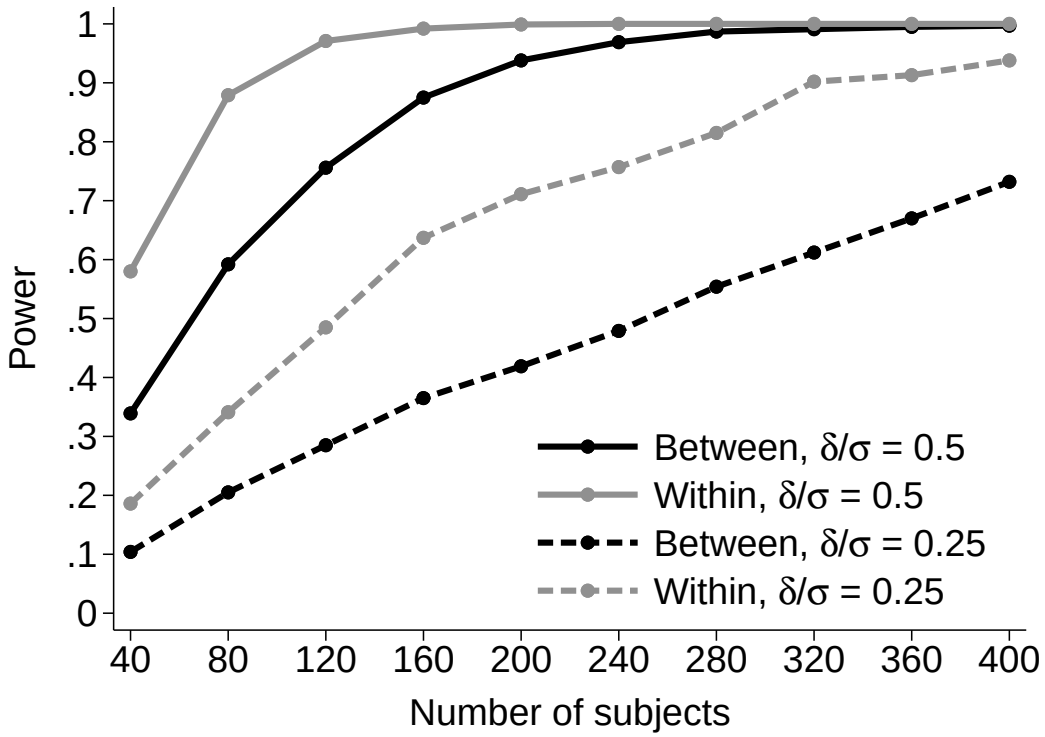

Figure E.1: Power levels for two-sided mean comparisons between- and within-individuals ($\alpha = 0.05, \sigma_a = \sigma_b = \sigma = 2, \delta = 1$ or $0.5, \delta/\sigma = 0.5$ or $0.25, H_0$: $\mu_a = \mu_b, H_a$: $\mu_a \neq \mu_b$).

power achieved for between- and within- comparisons for $\sigma = 2$ and $\delta = 1$ (i.e, $\delta/\sigma = 0.5$) and, even more conservatively, $\delta = 0.5$ (i.e, $\delta/\sigma = 0.25$), following the simulation-based approach proposed by Bellemare et al. (2016). While we do not have enough power to detect $\delta/\sigma = 0.25$, we obtain $1 - \beta > 80\%$ for $\delta/\sigma = 0.5$ whenever our groups have at least 80 subjects each.

Figure E.2 confirms the power calculations in Table E.1, showing the required sample size for detecting a given effect size δ with 80% power, when $\sigma \in \{1, 1.5, 2, 2.5\}$. Only in the most conservative settings (with high variance and small effect sizes) do we require samples larger than the ones we obtain; we are almost always able to detect 1-unit changes in effort.

Finally, since we are interested in testing a number of hypotheses, we must adjust ex ante for multiple hypothesis testing (List et al., 2019). As a conservative approach, we use a Bonferroni correction, requiring $\alpha = 0.05/k$, where k is the number of hypotheses. For instance, assuming $k = 5$, then the necessary significance level becomes $\alpha = 0.01$. In this case, the minimum effect sizes relative to the standard deviation, δ_{min}/σ, we are able to detect with 80% power are 0.342, 0.639, and 0.751 in panels A, B, and, respectively, C. Figure E.3 shows the required sample sizes for 80% power two-sided tests with significance $\alpha = 0.01$: our sample sizes are once again able to detect 1-unit changes in effort under all but the most conservative settings.

Across a range of assumptions regarding sample sizes, means, standard deviation, and significance, our experiment is *ex ante* able to detect small changes in social effort – i.e. around half a standard deviations, regardless of the actual standard deviation – even under the most conservative specifications. *Ex post*, it is important to note that while the standard deviations resulting from our experiment were higher than the ones we used for power calculations, the materialized differences were also larger, such that power was maintained throughout.[14]

[14] While List et al. (2011) warn against performing *ex post* power calculations, we can still estimate *ex post* minimum detectable effects, or MDEs (Ioannidis et al., 2017). Regressing social enterprise social effort on treatment dummies without sorting, we obtain standard errors of around 0.39, which correspond to MDEs of 1.1 units of social effort. These numbers are smaller than the estimated effects, again suggesting our study is well powered.

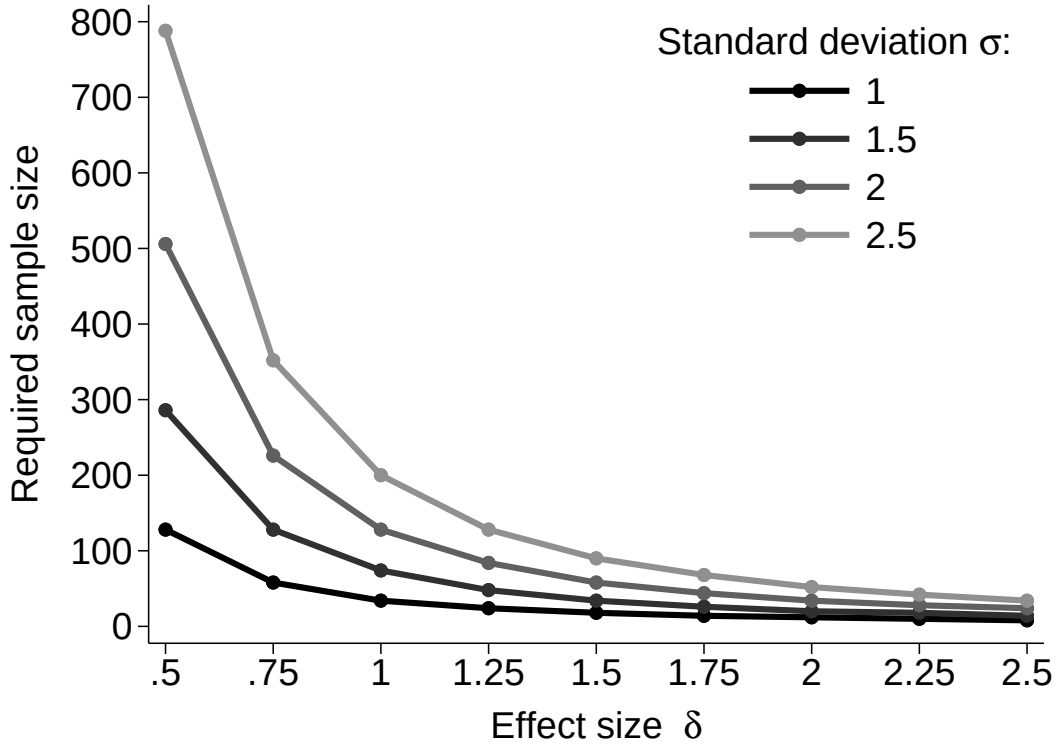

Figure E.2: Required sample size for detecting effect size δ with 80% power in two-sided mean comparison $(1-\beta=0.8, \alpha=0.05, \sigma_a=\sigma_b=\sigma, H_0\text{: } \mu_a=\mu_b, H_a\text{: } \mu_a\neq\mu_b)$.

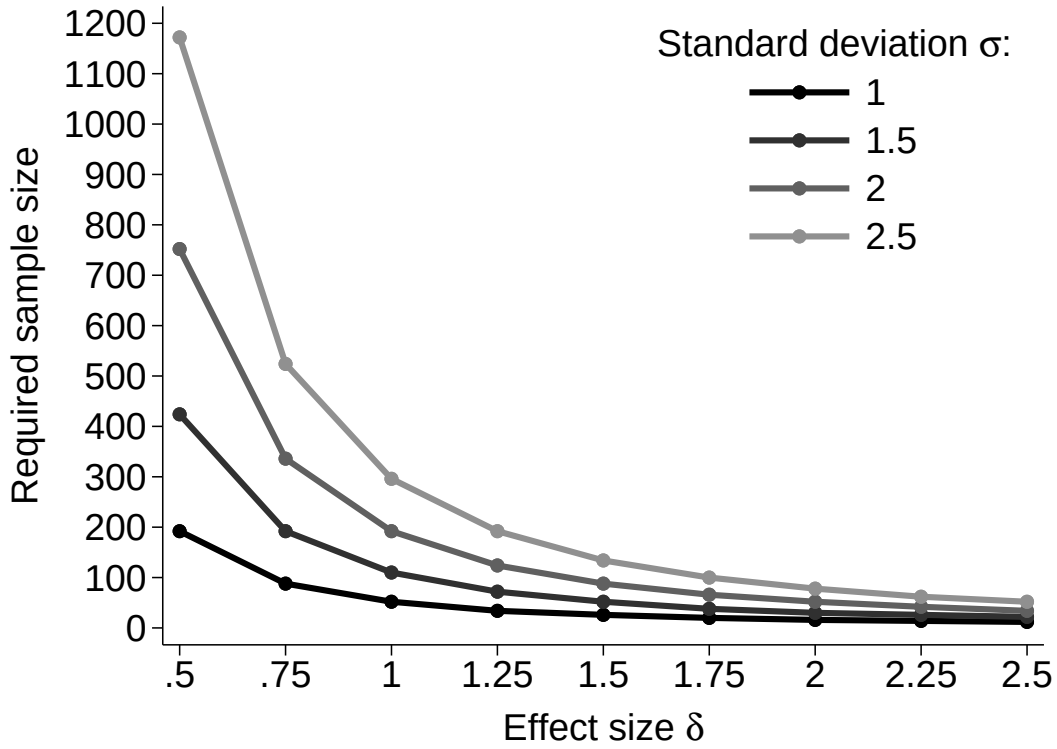

Figure E.3: Required sample size for detecting effect size δ with 80% power in two-sided mean comparison $(1-\beta=0.8, \alpha=0.01, \sigma_a=\sigma_b=\sigma, H_0\text{: } \mu_a=\mu_b, H_a\text{: } \mu_a\neq\mu_b)$.

Table E.1: **Power Calculations**

	N_a (1)	N_b (2)	μ_a (3)	μ_b (4)	σ (5)	δ/σ (6)	α (7)	2-sided $1-\beta$ (8)	1-sided $1-\beta$ (9)	$\delta^{80\%}_{min}$ (10)
A.	200	200	7.5	8	2.5	0.20	0.05	0.514	0.638	0.700
Without	200	200	7.5	8.5	2.5	0.40	0.05	0.979	0.991	
sorting	200	200	7.5	9	2.5	0.60	0.05	1.000	1.000	
	200	200	7.5	8	2	0.25	0.05	0.703	0.803	0.560
	200	200	7.5	8.5	2	0.50	0.05	0.999	1.000	
	200	200	7.5	9	2	0.75	0.05	1.000	1.000	
	200	200	7.5	8	1.5	0.33	0.05	0.914	0.954	0.420
	200	200	7.5	8.5	1.5	0.66	0.05	1.000	1.000	
	200	200	7.5	9	1.5	1.00	0.05	1.000	1.000	
	200	200	7.5	7.75	1	0.25	0.05	0.703	0.803	0.280
	200	200	7.5	8	1	0.50	0.05	0.999	1.000	
	200	200	7.5	8.5	1	1.00	0.05	1.000	1.000	
B.	100	40	7.5	8	2.5	0.20	0.05	0.186	0.281	1.310
With	100	40	7.5	8.5	2.5	0.40	0.05	0.565	0.685	
sorting	100	40	7.5	9	2.5	0.60	0.05	0.890	0.939	
(£0.50)	100	40	7.5	8	2	0.25	0.05	0.264	0.376	1.048
	100	40	7.5	8.5	2	0.50	0.05	0.756	0.845	
	100	40	7.5	9	2	0.75	0.05	0.978	0.991	
	100	40	7.5	8	1.5	0.33	0.05	0.425	0.551	0.786
	100	40	7.5	8.5	1.5	0.66	0.05	0.943	0.971	
	100	40	7.5	9	1.5	1.00	0.05	0.999	1.000	
	100	40	7.5	7.75	1	0.25	0.05	0.264	0.376	0.524
	100	40	7.5	8	1	0.50	0.05	0.756	0.845	
	100	40	7.5	8.5	1	1.00	0.05	0.999	1.000	
C.	120	25	7.5	8	2.5	0.20	0.05	0.148	0.230	1.540
With	120	25	7.5	8.5	2.5	0.40	0.05	0.439	0.566	
sorting	120	25	7.5	9	2.5	0.60	0.05	0.774	0.858	
(£1)	120	25	7.5	8	2	0.25	0.05	0.204	0.304	1.232
	120	25	7.5	8.5	2	0.50	0.05	0.618	0.732	
	120	25	7.5	9	2	0.75	0.05	0.923	0.960	
	120	25	7.5	8	1.5	0.33	0.05	0.325	0.446	0.924
	120	25	7.5	8.5	1.5	0.66	0.05	0.854	0.915	
	120	25	7.5	9	1.5	1.00	0.05	0.995	0.998	
	120	25	7.5	7.75	1	0.25	0.05	0.204	0.304	0.616
	120	25	7.5	8	1	0.50	0.05	0.618	0.732	
	120	25	7.5	8.5	1	1.00	0.05	0.995	0.998	

Power calculations for mean comparisons. In panel A, we compare any pair of treatments, with equal variance and sample size; in panels B and C, we consider comparisons between the expected largest and smallest groups within each treatment. With 80% power, the equivalent standardized minimum effect sizes $\delta^{80\%}_{min}/\sigma$ in column (10) are 0.280, 0.524, and 0.616 in panels A, B, and, respectively, C. In other words, whatever the *actual* experimental standard deviation, we are able to capture rather small changes in the mean.

Appendix F Robustness Check: Different Samples

Attention and manipulation checks To examine the extent to which subjects pay attention to the experiment, we included an attention check in our compassion sub-scale, asking subjects to select a particular item (i.e. 'Somewhat disagree'). In addition, we included a manipulation check, asking subjects to recall the bonus offered by the social enterprise contract. 95.35% of the 796 subjects passed the attention check by clicking on the required option, but only 55.90% passed the manipulation check, correctly recalling the social enterprise bonus. Rather than being due to poor understanding, this is most likely due to subjects not correctly assigning the social enterprise nomenclature to a particular contract, as the contracts subjects encountered in Parts 2 and 3 only included the company description and not a particular label. In other words, while the SE was described as a company for which both commercial and social tasks are important, the 'social enterprise' label was never actually used prior to the manipulation check. The unfortunate choice of wording in the manipulation check, coupled with the possibility of exerting effort on commercial and social tasks across all contracts, may have thus created confusion and led to wrong answers on the manipulation check. This question was also among the last asked in a rather long experiment, such that fatigue could have set in. Nonetheless, passing the manipulation check was independent of passing the attention check (χ^2 test, $p = 0.915$), suggesting that subjects did not systematically fail to pay attention. Moreover, passing or failing the attention check is independent of treatment, such that attrition for this reason is random (and generally, those who pass either check are not statistically different from those who fail on meaningful dimensions). Nevertheless, we exclude the 2.01% of subjects who failed *both* of these checks, although we have verified that including these subjects in our analyses does not affect our results. Table F.1 shows social enterprise social effort for different samples, relaxing and tightening restrictions around i) slider placement, ii) attention and manipulation checks, iii) gender, and iv) time taken to complete the experiment. Results are very similar across panels, including those where we require subjects to pass the attention check (Panels C and D), to pass at least one of the two checks (Panels A and E), and where we do not impose a restriction around attention and manipulation checks (Panels B, F, and G).

Definition of gender To stratify our randomization by gender, we used the pre-screening feature on the Prolific platform. This feature uses questions the platform previously asked its participants with regards to gender, allowing us to target our experiment at different subgroups. More specifically, we ran two identical experiments, restricting potential subjects to men in one and women in the other. To ensure subjects are not aware of this aspect of our experimental design and as a data quality check, we also ask subjects for their gender in Part 4. The pre-existing platform variable is consistent with the questionnaire answer, with an agreement rate of 99.26%. Throughout the analysis, we restrict the sample to observations where the two gender variables agree. Our results are robust to relaxing this restriction, as can be seen in Table F.1 by comparing Panels A, C, and E on the one hand, and Panels B, D, F, and G on the other.

Slider task placement In the experiment, we required subjects to position sliders exactly at 25 and 75 in order to produce a unit of real effort.[15] Indeed, 95.04% of the total 47,760

[15] The original task in Gill and Prowse (2012) is designed to measure total effort and consists of 48 sliders to be placed at exactly 50 in the space of two minutes. We use 15 sliders per contract to reduce the likelihood of subjects becoming bored with the task. In addition, our use of positions 25 and 75 as focal points is purely a matter of labeling, which we make clear to our subjects. A pilot experiment confirmed that these labels did not affect the decisions made by subjects and that the task is neither trivial, as it is not immediately obvious where precisely positions 25 and 75 are found, nor prohibitively difficult. As the effort required to move the slider in each direction is identical, differences in individuals' cost of taking the commercial or social action are only driven by social motivation differences. Alternatively, subjects could have moved the slider to their preferred distribution between commercial and social effort on a 0-100 scale; however, this effort allocation measure is similar to the dictator game we employ to measure altruism. We believe that allowing individuals to allocate effort in a binary

sliders were positioned correctly, and this does not differ by company type. However, despite an intention to exert commercial or social effort, there may be minor errors in positioning the slider. For instance, the slider could be positioned at 23, 24, 26, 27, 73, 74, 75, or 76, and these represent 0.86% of sliders. In our main results, we count minor deviations as units of effort under the assumption that they closely match an intention to place the slider precisely, but our results are unchanged when we only use precisely placed sliders, as Panels A and B of Table F.1 show. The remaining 4.1% of sliders are placed at other numbers, and in some observations more than half the sliders are inadequately placed. These subjects moved sliders more or less randomly and we drop them from the analysis; their inclusion attenuates our results only slightly, see Panels E, F, and G in Table F.1. Finally, a small number of participants placed sliders exclusively at 0 or 100, which indicate the direction of effort intended, but are clear deviations, such that they are not included in our main sample; recoding these observations (as 25 and 75) to count as units of effort leaves our results virtually identical.

Duration outliers There were several outliers with regards to the duration of the experiment, i.e. 2.1% of subjects took less than 10 minutes and 1.5% of subjects took more than 40 minutes. For the former, a short completion time may signal low attention paid to the task, reducing the quality of the data we obtain. The most likely reasons for the latter are that the session was left running while the subject was away temporarily or that a connection timed-out temporarily; either way, subjects may have paid less attention to the study. Our main sample excludes these observations, but the results are robust to including them, as can be seen in Table F.1 by comparing Panels A, C, and E on the one hand, and Panels B, D, F, and G on the other.

Answer consistency Under sorting, subjects choose their preferred contract and perform the slider task again. Consequently, subjects perform one contract twice, raising concerns about answer consistency. Reassuringly, the correlations between social effort levels with and without sorting are 0.681, 0.456, and 0.703 for individuals choosing the for-profit, nonprofit, and respectively, social enterprise contract. Consistency in repeated contracts does not vary with treatment level ($\chi^2 = 0.634$, $p > 0.5$). Figure F.1 shows a scatter plot of social effort with and without sorting for individuals choosing the social enterprise contract, weighted by number of observations. Most data points lie on the diagonal, suggesting no or minor deviations in repeated contracts. Overall, concerns about consistency do not threaten the validity of our results.

Pilot experiment Prior to completing the experiment we analyze in this paper, we conducted a pilot with 183 subjects, designed to guide our experimental design and power calculations. The main difference between them lies in the social enterprise contract description. Whereas we now write that "it is in the best interests of the company that both tasks receive attention", the pilot informed subjects that "the company cares equally about both tasks". We deemed this phrasing to provide too strong an anchor on a fully balanced effort allocation (a 50/50 split) and unrealistic to a certain extent. We preferred to give a more ambiguous description instead, allowing subjects to allocate their effort according to their perception of company needs. In the pilot, subjects were only required to move 10 sliders per contract, which we changed to 15 sliders per contract in order to remove any perceived similarity to the £10 dictator game. The pilot did not include a £0.25 treatment and was not stratified by gender. Nonetheless, the pilot results in Figure F.2 display a similar pattern as those we present in Figure 1: adverse social task specialization arises in the absence of financial incentives, while bonuses induce a more balanced effort allocation without reducing social motivation levels, regardless of incentive steepness (although with an elevated mission drift risk in the £1 treatment).

manner across 15 sliders carries less risk of introducing a purely mechanical relationship between social preferences and effort allocations, avoiding common method bias. Furthermore, we use 15 sliders – rather than 10, as we did in the pilot experiment – in order to limit any scale similarity between the slider task and the dictator game.

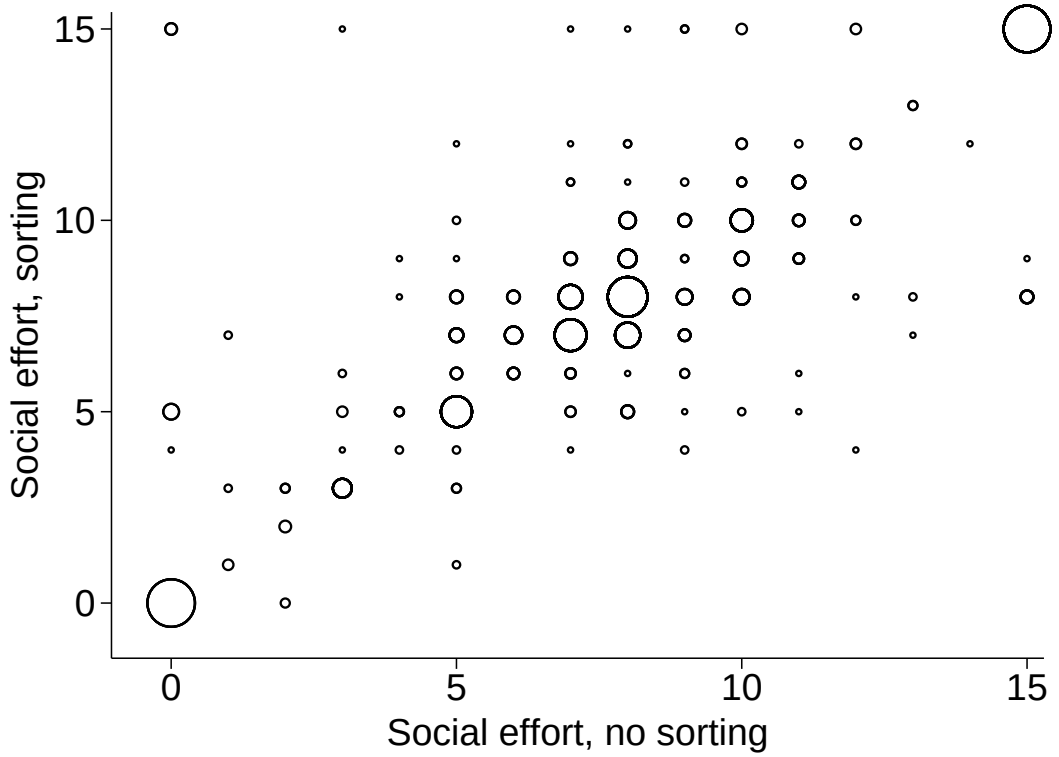

Figure F.1: Answer consistency across social enterprise contracts with and without sorting, for subjects who performed the social enterprise contract twice.

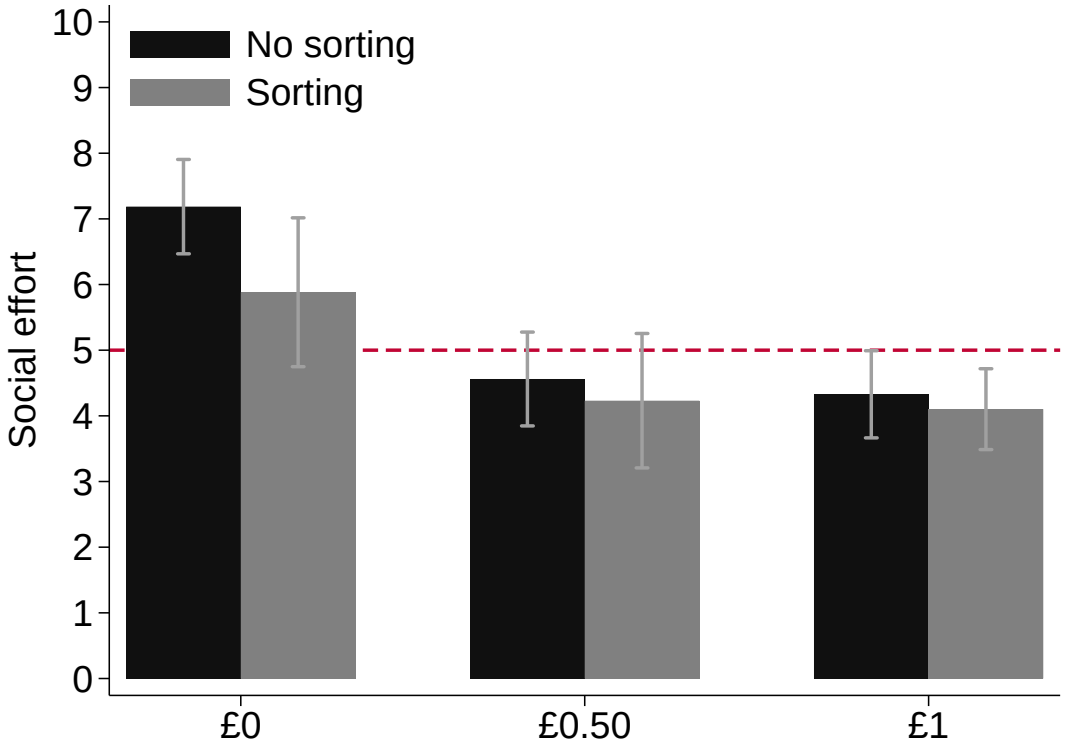

Figure F.2: Pilot data social effort in social enterprises by treatment, with 95% confidence intervals; the dashed line represents a fully balanced effort allocation.

Table F.1: **Social Enterprise Social Effort: Other Samples**

	No sorting				Sorting			
	(1) £0	(2) £0.25	(3) £0.50	(4) £1	(5) £0	(6) £0.25	(7) £0.50	£1 (8)
A. $NS = 0, F = F_p, 10 \leq t \leq 40$, pass either check ($N = 594$, $N_{SE} = 291$)								
Units	11.000	7.099	7.088	7.338	10.350	7.390	7.951	7.504
	(3.599)	(3.672)	(3.762)	(3.741)	(3.416)	(3.140)	(3.054)	(2.572)
Share	0.733	0.473	0.472	0.489	0.690	0.492	0.530	0.500
	(0.239)	(0.244)	(0.250)	(0.249)	(0.227)	(0.209)	(0.203)	(0.171)
B. $NS = 0$ ($N = 623$, $N_{SE} = 307$)								
Units	10.913	7.238	7.141	7.463	10.302	7.405	7.940	7.459
	(3.676)	(3.703)	(3.776)	(3.774)	(3.447)	(3.025)	(3.019)	(2.620)
Share	0.727	0.482	0.476	0.497	0.686	0.493	0.529	0.497
	(0.245)	(0.246)	(0.251)	(0.251)	(0.229)	(0.201)	(0.201)	(0.174)
C. $NS \leq 10, F = F_p, 10 \leq t \leq 40$, pass attention check ($N = 686$, $N_{SE} = 332$)								
Units	10.928	7.088	7.165	7.323	10.687	7.362	8.000	7.590
	(3.612)	(3.760)	(3.587)	(3.526)	(3.421)	(3.180)	(3.018)	(2.557)
Share	0.728	0.474	0.478	0.490	0.712	0.491	0.535	0.506
	(0.240)	(0.249)	(0.239)	(0.234)	(0.228)	(0.210)	(0.200)	(0.170)
D. $IS \leq 10$, pass attention check ($N = 717$, $N_{SE} = 346$)								
Units	10.948	7.173	7.215	7.325	10.775	7.378	7.989	7.500
	(3.649)	(3.757)	(3.592)	(3.633)	(3.441)	(3.073)	(2.987)	(2.682)
Share	0.729	0.479	0.481	0.491	0.718	0.492	0.534	0.500
	(0.243)	(0.249)	(0.239)	(0.242)	(0.229)	(0.203)	(0.198)	(0.178)
E. $IS \leq 30, F = F_p, 10 \leq t \leq 40$, pass either check ($N = 722$, $N_{SE} = 349$)								
Units	10.794	7.088	7.112	7.222	10.274	7.426	7.978	7.500
	(3.714)	(3.743)	(3.659)	(3.599)	(3.800)	(3.333)	(2.992)	(2.632)
Share	0.727	0.480	0.474	0.484	0.706	0.496	0.539	0.503
	(0.239)	(0.250)	(0.243)	(0.239)	(0.225)	(0.222)	(0.204)	(0.170)
F. $IS \leq 30$ ($N = 767$, $N_{SE} = 372$)								
Units	10.700	7.149	7.158	7.272	10.321	7.566	7.959	7.407
	(3.827)	(3.775)	(3.635)	(3.664)	(3.785)	(3.298)	(2.949)	(2.735)
Share	0.724	0.488	0.478	0.487	0.709	0.505	0.538	0.497
	(0.242)	(0.254)	(0.242)	(0.244)	(0.226)	(0.219)	(0.201)	(0.177)
G. Full sample ($N = 796$, $N_{SE} = 388$)								
Units	10.239	6.984	6.901	7.040	9.419	7.388	7.774	7.208
	(4.296)	(3.868)	(3.790)	(3.802)	(4.550)	(3.457)	(3.101)	(2.935)
Share	0.717	0.493	0.479	0.489	0.717	0.505	0.547	0.500
	(0.252)	(0.258)	(0.246)	(0.245)	(0.247)	(0.247)	(0.209)	(0.182)

These sampling criteria, though not comprehensive, cover a range of choices regarding slider placement, duration, gender variables, and attention checks. We display social effort as units and shares due to the larger number of imprecisely placed sliders not counted towards the total in some panels (esp. E, F, and G), which could distort the interpretation of effort allocation. Standard deviations in parentheses. N, N_{SE} = total and social enterprise sample size; NS = incorrect sliders, including imprecise (e.g. 23 is incorrect); IS = incorrect sliders, excluding imprecise (e.g. 23 is correct); F, F_p = gender variables from our experiment and Prolific; t = experiment time.

Appendix G Robustness Check: Social Preferences

Composite social motivation measure Since social motivation may entail different aspects – altruism, compassion, reciprocity, etc. –, none of the individual measures of social preferences may perfectly capture this complex concept. To extract the maximum information from the various measures we collect, we perform a common factor analysis, with an orthogonal varimax rotation. We find that our six social preference variables load onto a single factor with Eigenvalue larger than 1 accounting for 80.78% of variance, which we label *Social motivation* (see Table G.1). *Inequality aversion* loads negatively on this factor and *Altruism* has a smaller loading than our other measures, suggesting that this game-theoretic measurement may be an imperfect proxy for social motivation (see also Figure G.1 for variable loadings on the first two factors). Due to its broader nature, we use *Social motivation* throughout the experiment, together with *Compassion*. Note that including *Risk preferences* and *Time preferences*, potentially correlated with social preferences, in the principal factor analysis produce similar results, as does using the individuals variables underlying *Prosocial behavior*.

Factor analysis robustness Our measure of *Social motivation* is extracted through a common factor analysis, with an orthogonal varimax rotation, retaining the factor with Eigenvalue larger than 1. We consider this the simplest and most transparent way of performing factor analysis for the purpose of summarizing and understanding the latent variable underlying our social preference measures. However, this approach assumes that i) the underlying factors are uncorrelated and ii) only factors with Eigenvalue larger than 1 are relevant (Conway and Huffcutt, 2003). To check whether other techniques for extracting the principal factor(s) affect our analysis, we compare the results of common factor, principal-component factor, iterated principal factor, and maximum-likelihood factor extraction techniques with orthogonal (varimax) and non-orthogonal (oblimin and oblimax) rotations. Most of these approaches produce a single factor with Eigenvalue larger than 1 and the drop in Eigenvalues is always largest when going from Factor 1 to Factor 2: as a result, the extraction of a single latent factor is validated across different methods. *Compassion*, *Altruism*, *Hypothetical altruism*, *Willingness to share*, and *Prosocial behavior* load strongly onto this factor in each case. Table G.2 displays the correlations between the principal factors obtained with the different approaches. These correlations range from 0.75 to 1.00, indicating highly consistent factors across different factor extraction and rotation methods (we find similar results performing this analysis with a promax rotation). Using common factor analysis with an orthogonal varimax rotation seems justified in this case (Goretzko et al., 2021). Moreover, our sample size of 708 respondents should be large enough to produce undistorted results (Conway and Huffcutt, 2003; Goretzko et al., 2021).

Revealed preference social motivation Social task effort in the for-profit contract without sorting may also provide a measure of social motivation, because individuals renounce personal pay-offs in order to exert social task effort. This *revealed preference* measure is positively correlated with our other social preference measures, loads positively on the *Social motivation* factor, and produces similar results as the other measures (available upon request). However, due to the random order of Part 2 contracts, the social enterprise bonus is revealed to some subjects before they perform the for-profit contract; it could thus be contaminated by the treatment in a way that is correlated with subsequent choices, such that *Compassion* and *Social motivation* provide cleaner measures.

Changes in social motivation In Table 4, we provide a series of tests for equality of means, variances, and distributions in *Compassion* and *Social motivation* (plotted in Figure G.2). We also estimate linear regressions of these social preference measures for individuals who select into social enterprises on the treatment dummies. The results in columns (1) and (2) of Table G.3

show some motivation crowd-out in the £1 treatment. To examine distributional changes, we create dummies for whether individuals are in the bottom or top 25% of individuals in a given measure, and estimate linear probability models for their presence in social enterprises. In the £1 treatment, we find an increase (decrease) in the number of individuals at the bottom (top) of the distribution of *Social Motivation*. Our measure of compassion registers no distributional shifts across treatments. Only the selection of low *Social motivation* individuals into social enterprises in the £1 treatment survives multiple hypotheses test adjustments ($p = 0.011$, without controls) (List et al., 2019). Another way to analyze such shifts is to perform quantile regressions of social preference variables on treatment dummies. Table G.4 suggests that the *Social motivation* of individuals who select into the social enterprise contract is reduced across the distribution in the £1 treatment, although this is only weakly significant; *Compassion* is unaffected.

Alternative social motivation measures While incentivized measures are preferable to hypothetical ones, it is important to show how sensitive our results are to using different constructs. In addition, social preference games in the lab (e.g., dictator) may not accurately capture social motivation in the field (Levitt and List, 2007; Galizzi and Navarro-Martinez, 2019). We complement such measures with psychological scales and hypothetical questions to alleviate this external validity concern (and the main analysis focuses on *Compassion* and a composite *Social motivation* factor). The results using these alternative measures are shown in columns (1)-(5) of Table G.5. Increasing incentives are correlated with lower levels of social preferences in the SE contract, in particular altruism and willingness to share when the bonus is £1. However, the List et al. (2019) multiple hypothesis testing adjustment renders all coefficients statistically insignificant at conventional levels ($p > 0.1$), suggesting that strong incentives do not attract significantly less motivated workers. Interestingly, column (2) suggests that incentives' potential to widen the distribution of individual payoffs does not deter inequality-averse individuals; workers do not seem to perceive incentives as 'unfair' from a redistribution perspective, or at least do not anticipate this consequence. As columns (6) and (7) suggest, higher social enterprise incentive levels do not attract individuals with a higher risk propensity or more myopic individuals.[16]

Social preferences by contract choice One argument for why adverse specialization occurs in social enterprises relates to the highly socially motivated individuals who join this organizational form. Regardless of the social enterprise bonus, we expect that other-regarding preferences are lowest for self-selected for-profit workers and highest for self-selected nonprofit workers; social enterprise workers are in between, with some differences across treatments. To see this, Table G.6 presents a regression analysis counterpart to the comparisons in Table 4, considering subjects make a single choice between the three contracts: social enterprise motivation is different from for-profit motivation but not nonprofit motivation outside of the £1 treatment. To examine this possibility, we regress our social preference measures on dummies for Part 3 contract choices, controlling for treatment and choice of good cause (i.e. mission fixed effects). Table G.7 shows that our expectation is met for social enterprise and for-profit worker comparisons, with the exception of *Inequality aversion*, *Risk preferences*, and *Time preferences*. It does not appear that more inequality averse or less risk tolerant individuals join social enterprises, although social enterprise workers put more weight on the future relative to for-profit workers. While for-profit and nonprofit workers are highly different in their social preferences, social enterprise and nonprofit workers are remarkably similar, with a statistical difference observed only for *Hypothetical altruism* (otherwise $p > 0.1$). This supports our argument that individuals selecting into social enterprises are highly socially motivated, which may result in adverse specialization when no monetary rewards are provided.

[16] We have also checked that stronger incentives do not attract individuals with different levels of education or income. They appear to attract individuals who took less time to complete the comprehension check (as a potential proxy for cognitive ability), but multiple hypothesis testing adjustments eliminate the significant coefficients.

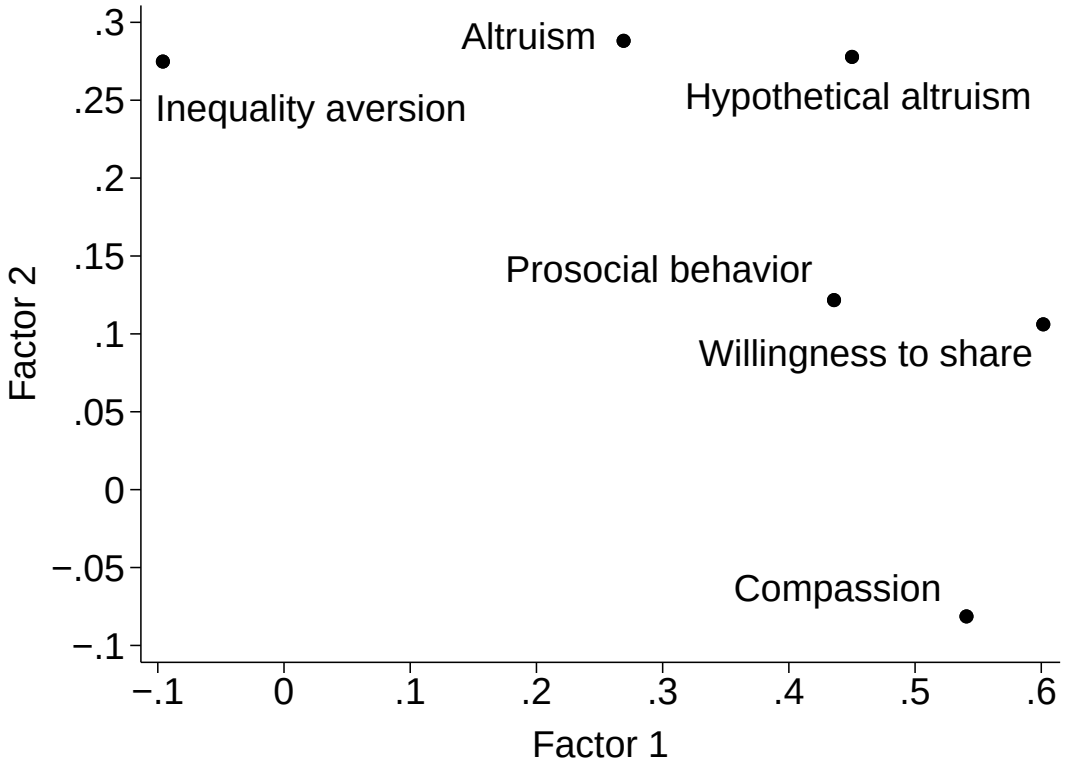

Figure G.1: Social preference loadings on the first two factors obtained from a common factor analysis with orthogonal varimax rotation.

Table G.1: **Composite Social Preferences: Factor Loadings**

Variable	Factor 1	Factor 2	Uniqueness
Compassion	0.5407	-0.0814	0.7010
Altruism	0.2691	0.2881	0.8446
Inequality aversion	-0.0960	0.2748	0.9153
Hypothetical altruism	0.4501	0.2778	0.7203
Willingness to share	0.6016	0.1061	0.6268
Prosocial behavior	0.4358	0.1217	0.7953
Eigenvalue	1.1953	0.2015	
Variance explained	80.78%	19.22%	
Label	*Social motivation*		

Factor loadings for a common factor analysis with an orthogonal varimax rotation; results are similar with oblique (non-orthogonal) rotations (see Table G.2). As social context may interact with risk and time preferences, we have also checked that including these variables in our measure of motivation does not impact the results. Reassuringly, the results are qualitatively similar. Risk and time preferences load more on Factor 2, so the Factor 1 has a slightly smaller, yet still dominant, explanatory power. Results are also similar when we include the revealed social preferences from the for-profit contract or use the 5 items that comprise *Prosocial behavior* individually.

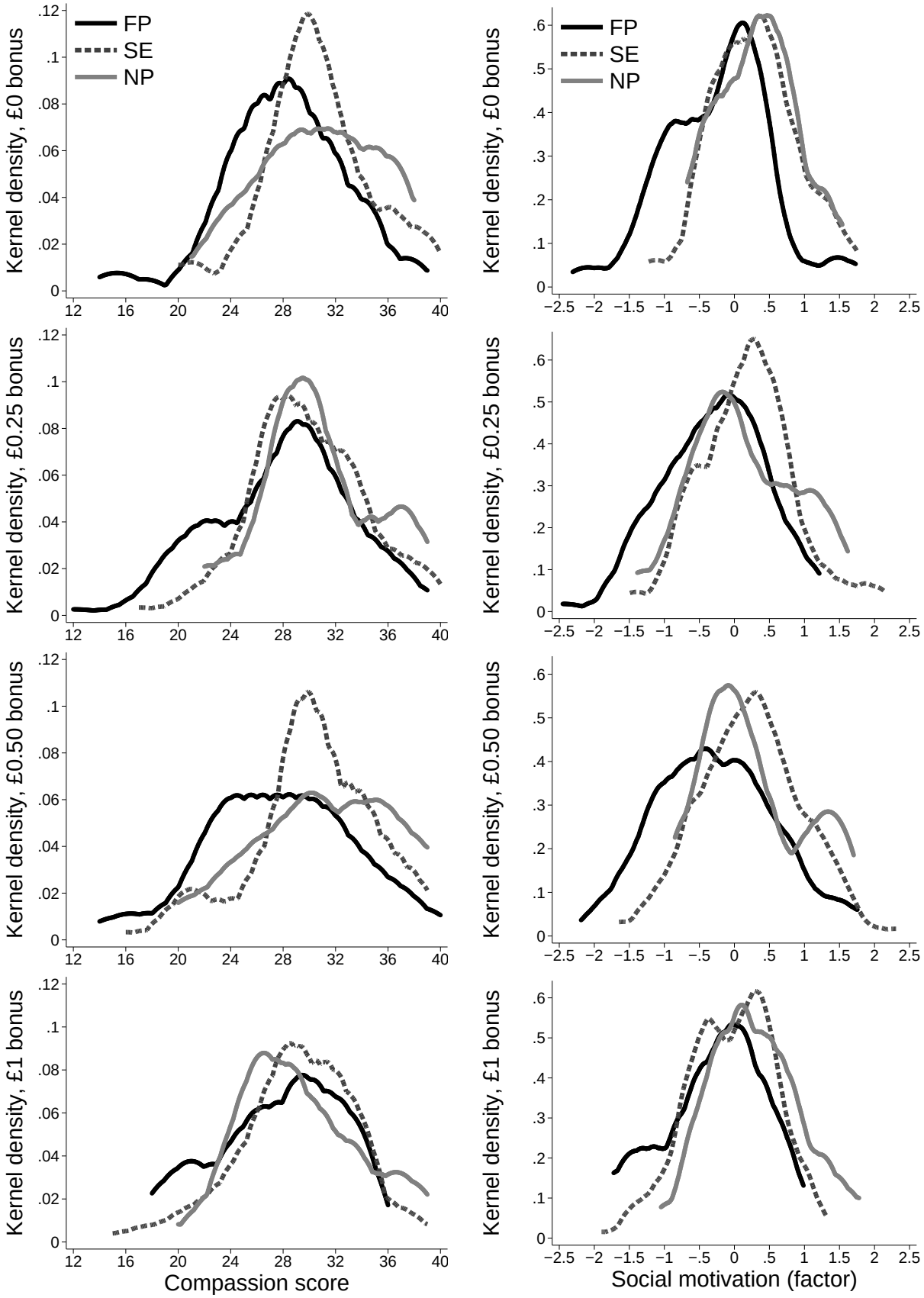

Figure G.2: Distribution of *Compassion* and *Social motivation* under sorting, by treatment and contract choice (FP = for-profit, NP = nonprofit, SE = social enterprise).

Table G.2: **Correlations of Factors Obtained with Different Methods**

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1)	CF, varimax	1.00											
(2)	PF, varimax	0.98	1.00										
(3)	IF, varimax	0.92	0.87	1.00									
(4)	ML, varimax	0.83	0.76	0.92	1.00								
(5)	CF, oblimin	0.99	0.99	0.89	0.77	1.00							
(6)	PF, oblimin	0.98	0.99	0.86	0.75	0.99	1.00						
(7)	IF, oblimin	0.97	0.94	0.98	0.88	0.95	0.94	1.00					
(8)	ML, oblimin	0.85	0.78	0.93	0.99	0.79	0.78	0.89	1.00				
(9)	CF, oblimax	0.99	0.99	0.89	0.77	1.00	0.99	0.95	0.79	1.00			
(10)	PF, oblimax	0.98	0.99	0.86	0.75	0.99	1.00	0.94	0.78	0.99	1.00		
(11)	IF, oblimax	0.97	0.94	0.98	0.88	0.95	0.94	1.00	0.89	0.95	0.94	1.00	
(12)	ML, oblimax	0.85	0.78	0.93	0.99	0.79	0.78	0.89	1.00	0.79	0.78	0.89	1.00

CF = common factor; PF = principal-component factor; IF = iterated principal factor; ML = maximum-likelihood factor. All correlations produce p-values < 0.001.

Table G.3: **Social Preferences, Conditional on Social Enterprise Sorting**

	Mean		Bottom 25%		Top 25%	
	(1)	(2)	(3)	(4)	(5)	(6)
A. Compassion						
£0.25	-0.122	-0.074	0.022	-0.005	-0.005	0.007
	(0.162)	(0.170)	(0.067)	(0.068)	(0.082)	(0.085)
£0.50	-0.064	0.030	0.027	0.001	0.064	0.106
	(0.158)	(0.153)	(0.064)	(0.062)	(0.080)	(0.080)
£1	-0.304**	-0.184	0.113*	0.087	-0.001	0.044
	(0.148)	(0.155)	(0.064)	(0.065)	(0.075)	(0.078)
R^2	0.016	0.143	0.013	0.112	0.004	0.110
B. Social motivation						
£0.25	-0.097	-0.073	0.144**	0.138**	-0.059	-0.039
	(0.123)	(0.123)	(0.059)	(0.062)	(0.089)	(0.089)
£0.50	-0.070	-0.025	0.098*	0.095*	-0.047	-0.012
	(0.119)	(0.117)	(0.051)	(0.054)	(0.086)	(0.085)
£1	-0.339***	-0.274***	0.187***	0.177***	-0.204***	-0.170**
	(0.107)	(0.112)	(0.052)	(0.058)	(0.078)	(0.081)
R^2	0.039	0.137	0.026	0.100	0.032	0.134
Controls	No	Yes	No	Yes	No	Yes

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. $N = 341$. Robust standard errors in parentheses. Columns present regressions of *Compassion* (standardized) and *Social motivation* (factor) on treatment dummies; the baseline category is the £0 treatment. Controls include age, gender, studentship, education, income, risk and time preferences, and choice of good cause.

Table G.4: **Social Preferences in Social Enterprises: Quantile Regressions**

	(1) 10^{th} pct.	(2) 25^{th} pct.	(3) 50^{th} pct.	(4) 75^{th} pct.	(5) 90^{th} pct.
A. Compassion					
£0.25	0.295	-0.006	-0.256	0.080	0.115
	(0.347)	(0.261)	(0.226)	(0.252)	(0.420)
£0.50	0.201	0.179	0.039	0.181	-0.101
	(0.302)	(0.232)	(0.212)	(0.222)	(0.323)
£1	-0.115	0.030	-0.260	-0.121	-0.374
	(0.301)	(0.226)	(0.229)	(0.190)	(0.318)
B. Social motivation					
£0.25	-0.106	-0.151	-0.060	-0.051	-0.045
	(0.205)	(0.181)	(0.156)	(0.172)	(0.242)
£0.50	-0.239	-0.067	0.036	0.043	-0.003
	(0.207)	(0.182)	(0.161)	(0.182)	(0.205)
£1	-0.317*	-0.297*	-0.260	-0.259	-0.359*
	(0.186)	(0.177)	(0.163)	(0.176)	(0.209)

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. $N = 341$. Bootstrapped standard errors in parentheses (1,000 replications). Columns present quantile regressions of *Compassion* (standardized) and *Social motivation* (factor) on treatment dummies for the 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} quantiles; the baseline category is the £0 treatment. Controls include age, gender, income, studentship, education, risk and time preferences, and choice of good cause.

Table G.5: **Social Preferences by Treatment, Conditional on Social Enterprise Sorting**

	Altruism (1)	Inequality aversion (2)	Hypothetical Altruism (3)	Willing to share (4)	Prosocial behavior (5)	Risk (6)	Time (7)
£0.25	-0.389	-0.316	14.214	-0.150	-0.291	-0.136	-0.076
	(0.308)	(0.384)	(29.871)	(0.375)	(0.208)	(0.437)	(0.378)
£0.50	-0.497*	-0.149	2.896	-0.039	-0.189	0.167	-0.109
	(0.294)	(0.366)	(28.500)	(0.358)	(0.198)	(0.417)	(0.361)
£1	-0.762***	-0.199	-43.925	-0.810**	-0.295	-0.043	-0.555
	(0.281)	(0.350)	(27.262)	(0.343)	(0.189)	(0.399)	(0.345)
Test of joint significance p-value:							
	0.051	0.870	0.051	0.014	0.426	0.862	0.217
R^2	0.023	0.002	0.023	0.031	0.008	0.002	0.013

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. $N = 341$. Standard errors in parentheses. Results from regressions of social preference measures on treatment dummies. The baseline category comprises the £0 treatment. Controlling for age, gender, income, studentship, education, risk and time preferences, and choice of good cause does not alter the qualitative picture. Multiple hypothesis testing adjustments render all coefficients statistically insignificant at conventional levels ($p > 0.1$).

Table G.6: **Motivation across Contracts, with Sorting**

	Compassion (standardized)				Social motivation (factor)			
	(1) £0	(2) £0.25	(3) £0.50	(4) £1	(5) £0	(6) £0.25	(7) £0.50	(8) £1
For-profit	-0.472***	-0.418***	-0.523***	-0.289	-0.543***	-0.494***	-0.509***	-0.248
	(0.163)	(0.159)	(0.164)	(0.209)	(0.124)	(0.119)	(0.128)	(0.159)
Nonprofit	0.002	0.109	0.186	0.076	0.020	-0.078	0.063	0.353**
	(0.211)	(0.223)	(0.262)	(0.214)	(0.143)	(0.188)	(0.179)	(0.150)
R^2	0.071	0.052	0.107	0.018	0.137	0.099	0.124	0.061
N	170	178	185	175	170	178	185	175

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors in parentheses. We regress *Compassion* or *Social Motivation* on subjects' choice of contract, using the social enterprise motivation means in Panel B of Table 4 as a baseline.

Table G.7: **Social Preferences by Contract Choice**

	Compassion (1)	Altruism (2)	Inequality aversion (3)	Hypothetical Altruism (4)	Willing to share (5)	Prosocial behavior (6)	Social motivation (7)	Risk (8)	Time (9)
For-profit	28.873***	3.716***	2.695***	96.148***	6.015***	1.307**	-0.136*	5.086***	6.209***
	(0.496)	(0.248)	(0.236)	(13.888)	(0.264)	(0.137)	(0.081)	(0.264)	(0.216)
Social enterprise	30.977***	4.822***	2.496***	143.172***	7.169***	1.534***	0.307***	5.176***	6.643***
	(0.522)	(0.237)	(0.247)	(18.353)	(0.262)	(0.149)	(0.083)	(0.280)	(0.230)
Nonprofit	31.514***	4.913***	2.685***	192.515***	7.371***	1.527***	0.427***	5.044***	6.607***
	(0.647)	(0.251)	(0.313)	(22.131)	(0.314)	(0.153)	(0.094)	(0.341)	(0.305)
Test of equality p-values:									
For-profit vs social enterprise	0.000	0.000	0.312	0.001	0.000	0.031	0.000	0.675	0.020
For-profit vs nonprofit	0.000	0.000	0.971	0.000	0.000	0.086	0.000	0.894	0.157
Nonprofit vs social enterprise	0.338	0.663	0.499	0.019	0.420	0.950	0.142	0.666	0.888
R^2	0.973	0.801	0.588	0.442	0.888	0.546	0.093	0.812	0.900

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. $N = 708$. Robust standard errors in parentheses. Results from regressions of social preferences on contract choice (suppressing the constant), with fixed effects for choice of treatment and mission.

Appendix H Robustness Check: Effort Measures

Share of social effort Our main analysis focuses on sliders moved to the position equivalent to social effort, but our sampling restrictions include some observations where not all 15 sliders in a contract were placed correctly. This may create problems in interpreting results using social effort if total effort is not adjusted accordingly. For example, for one individual 6 units of social effort out of 15 correctly placed sliders result in 40% social effort, while for another 6 units of social effort out of 10 correctly placed sliders result in 60% social effort.[17] To address this, we assess the effects of treatment on the share of effort exerted in the social task, i.e. social effort as a fraction of total effort. The results for this dependent variable in Table H.1 completely mirror those in Table 3 for social effort units. Throughout this section, we show results for both dependent variables in order to ensure robustness (and our multiple hypothesis testing adjustments take this into account). Considering the different sampling restrictions with regards to slider placement in Table F.1, using units of social effort becomes more problematic when including subjects with more incorrectly placed sliders, which may attenuate our adverse specialization results for the £0 bonus social enterprise. Compare, for instance, column (1) in Table F.1, where we progressively relax slider placement restrictions. The units of social effort go down from 11 in Panel A, where we restrict the sample to subjects who only placed sliders at 25 and 75, to 10.24 in Panel G, where all subjects are included, and sliders placed at 23-27 and 73-77 are considered correct. While units of social effort decrease, the share of social effort only varies between 0.733 and 0.717, suggesting that this measure captures effort allocation well regardless of slider placement restrictions. That results across both variables are very similar is encouraging, and we focus our analysis on units of social effort due to its higher transparency.

Fixed effects models Without sorting, all subjects perform the for-profit, nonprofit, and social enterprise contracts. Comparisons across contracts must then adjust for the paired nature of the test. More specifically, the results in Table 3 suggest the social enterprise contract is always different from the nonprofit and for-profit contracts in simple and paired t-tests. An alternative way to account for non-independence is to estimate individual fixed effects models. We regress social effort (as units or shares) on dummies for social enterprise and nonprofit contracts, using Part 2 observations only and the for-profit contract as a baseline. The results in Table H.2 confirm our results: social effort is higher in the social enterprise and nonprofit contracts relative to the for-profit contract; the former are different from each other, with $p < 0.0001$, although the gap is much smaller in the £0 treatment.

Adverse specialization The top panel of Figure 2 shows that 30%-40% of social enterprise workers in the £0 treatment only exert social effort. This bimodal distribution stands in contrast to a distribution centered around the SE social effort average, suggesting adverse specialization is driven by a subgroup of workers, rather than by higher social effort across the board. What drives this behavior? Our theoretical framework implies that a high level of social motivation should increase the likelihood that workers exert only social effort. To examine this, we regress a dummy for maximum social effort on *Compassion* and *Social motivation* in Table H.3. With or without sorting, there does not appear to a linear association between *Social motivation* and maximum social effort.[18] More compassionate workers are more likely to exert maximum effort, especially in the top decile of the distribution; the direction is the same for *Social motivation*,

[17] Note that we do not find differences in social enterprise total effort across treatments (as subjects could choose to move less that 15 sliders), regardless of the sample we use. Our additional study in Appendix J provides further evidence that total effort is not an important adjustment margin.

[18] Although the negative effects of *Social motivation* on adverse specialization in Panel B are not statistically significant, the quadratic results suggest that both the least and the most motivated individuals may exert maximum social effort. The former may do so as a response to performing a contract they would not otherwise have chosen and may feel compelled to exert substantial social effort (Lazear et al., 2012).

although these results are not significant (potentially due to small sample size). However, these individuals have self-selected into the social enterprise contract and are more motivated than those who self-selected into the for-profit contract, as per Table G.6. Figure H.1 displays a scatter plot of social enterprise *Social effort* in the £0 treatment and *Social motivation*. Those who self-select into the social enterprise (the gray dots) are more likely to have higher motivation, as there are visibly fewer observations with *Social motivation* < 0. They are also more likely to exert maximum social effort, as the concentration of gray dots in the upper right-hand side suggests. Overall, adverse specialization in the absence of monetary incentives appears to be driven especially by individuals with very high levels of compassion. In additional checks, we verified that these individuals experience the largest changes in effort allocation once incentives are introduced.

Measures of imbalance In our main analysis, we focus on units of *Social effort* as the most straightforward measure of effort allocation, and compare social enterprise social effort with a fully balanced effort allocation (i.e. 7.5 units) and with for-profit and nonprofit social effort (and their average). Moreover, results are similar when we perform comparisons using the share of social effort, with a fully balanced effort allocation as a reference (i.e. a 50% share). Because our theory is centered on the notion of balance, we can also capture effort allocation with more direct measures of (im)balance. These measures have a straightforward reference point (i.e. full balance implies a value of zero) and account for incorrectly placed sliders. The difference between social and commercial effort provides a metric of how dominant the social task is relative to the commercial task, although this variable can become negative if commercial effort dominates; this variable allows for deviations from full balance to cancel each other out and can be considered a flexible measure of overall imbalance. Conversely, deviations from full balance can be considered as distortions regardless of their direction; thus, total imbalance can be conceptualized as the absolute value of the difference between social and commercial effort. Table H.4 presents the results from using both of these variables, for which a fully balanced effort allocation produces a value of zero. The £0 treatment shows a significant level of imbalance, while all other treatments are associated with significantly more balanced effort allocations (similar across incentive levels). In addition, as Panel C shows, subjects in the top decile of *Compassion* have higher levels of imbalance, consistent with a relationship between social preferences and adverse specialization.

Tobit models Table 5 shows the results of linear regressions of the different measures of social effort and balance on treatment dummies, a dummy for the sorting condition, and their interaction. While this represents the simplest and most transparent estimation method, all three dependent variables exhibit a certain degree of censoring. Social effort can only range between 0 and 15, as can the measure of absolute balance, while the share of social effort ranges from 0% to 100%, with around 16% of observations being censored in each case. Therefore, Tobit regressions are a more appropriate estimation technique. The Tobit results we show in Table H.5 are fully parallel to the ones obtained with linear regression, suggesting censoring in the dependent variables is not a concern.

Absence and presence of bonus As our results suggest, social enterprise social effort does not differ significantly between the £0.25, £0.50, and £1 treatments. To examine their joint impact in an analysis of extensive versus intensive margin effects, we aggregate these three treatments into a single *Bonus* dummy, whereas the £0 treatment corresponds to an social enterprise that uses no bonus. Table H.6 replicates the analysis in Table 5 with this simple dummy for the presence or absence of incentives. Confirming our previous findings, allowing for sorting does not matter for the relationship between incentives and effort allocation.

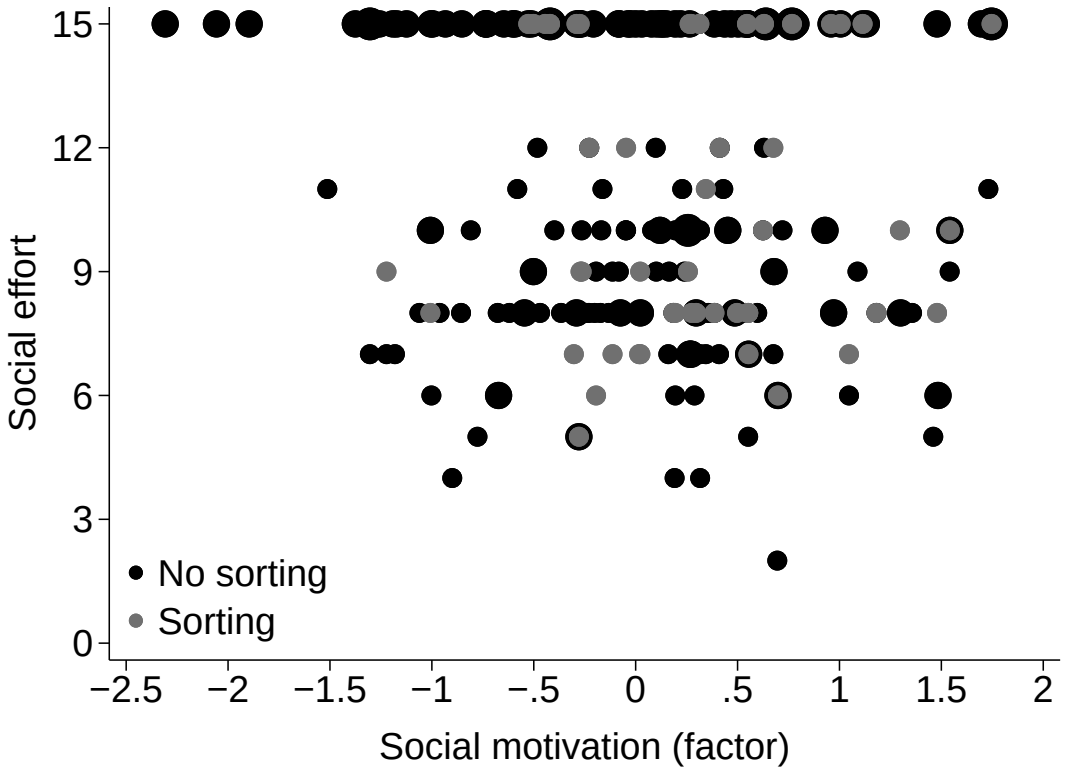

Figure H.1: Social effort in the social enterprise contract, £0 treatment.

Table H.1: **Social Effort Share, by Contract and Treatment**

	No sorting				Sorting			
	£0	£0.25	£0.50	£1	£0	£0.25	£0.50	£1
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
For-profit	0.328	0.344	0.349	0.376	0.266	0.276	0.273	0.325
	(0.023)	(0.022)	(0.022)	(0.023)	(0.028)	(0.027)	(0.029)	(0.062)
Social enterprise	0.729	0.477	0.475	0.492	0.709	0.504	0.535	0.507
	(0.018)	(0.019)	(0.018)	(0.018)	(0.032)	(0.025)	(0.021)	(0.015)
Nonprofit	0.887	0.854	0.869	0.854	0.890	0.917	0.911	0.897
	(0.015)	(0.016)	(0.017)	(0.018)	(0.030)	(0.034)	(0.056)	(0.046)
Social enterprise effort t-tests of equality p-values:								
vs £0		0.000	0.000	0.000		0.000	0.000	0.000
vs £0.25			0.922	0.553			0.340	0.925
vs £0.50				0.479				0.256
Social enterprise effort t-tests of equality with nonprofit and for-profit average, p-values:								
	+0.000	−0.000	−0.000	−0.000	+0.001	−0.006	−0.128	−0.012

Standard errors in parentheses. Within each column the for-profit, nonprofit, and social enterprise social effort levels are different from each other ($p < 0.0001$). We employ matched pair t-tests for the no sorting condition, acknowledging that all individuals performed the slider task in all contract types. In the bottom row, '+' means the social enterprise is closer to the nonprofit than the for-profit, and '−' means the social enterprise is closer to the for-profit than the nonprofit.

Table H.2: **Social Effort without Sorting: Fixed Effects Models**

	(1) £0	(2) £0.25	(3) £0.50	(4) £1
A. Units of social effort				
Social enterprise	6.018***	1.978***	1.881***	1.703***
	(0.426)	(0.331)	(0.319)	(0.336)
Nonprofit	8.388***	7.567***	7.757***	7.057***
	(0.455)	(0.433)	(0.425)	(0.443)
B. Share of social effort				
Social enterprise	0.401***	0.133***	0.126***	0.116***
	(0.028)	(0.022)	(0.021)	(0.022)
Nonprofit	0.559***	0.510***	0.521***	0.478***
	(0.030)	(0.028)	(0.028)	(0.029)
Observations	510	534	555	525
Subjects	170	178	185	175

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses. The baseline category consists of social effort in the for-profit contract. Social effort is different in the social enterprise and nonprofit contracts across all treatments, with $p < 0.0001$.

Table H.3: **Adverse Specialization and Motivation**

	No sorting			Sorting		
	(1)	(2)	(3)	(4)	(5)	(6)
A. Compassion						
Compassion	0.027	0.037		0.235**	0.114*	
	(0.047)	(0.040)		(0.085)	(0.080)	
Compassion^2		0.095***			0.150***	
		(0.025)			(0.050)	
Compassion $\geq 90^{\text{th}}$ pct.			0.363***			0.946***
			(0.134)			(0.148)
R^2	0.093	0.154	0.136	0.579	0.657	0.740
N	170	170	170	49	49	49
B. Social motivation						
Social motivation	-0.186	-0.076		0.241	0.145	
	(0.057)	(0.056)		(0.176)	(0.183)	
Social motivation^2		0.085*			0.146	
		(0.047)			(0.163)	
Social motivation $\geq 90^{\text{th}}$ pct.			-0.021			0.573
			(0.139)			(0.372)
R^2	0.106	0.122	0.091	0.531	0.546	0.566
N	170	170	170	49	49	49

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors in parentheses. Results from linear regressions of dummies for exerting only social effort in the social enterprise with a £0 bonus on subjects' *Compassion* or *Social Motivation* (and their square terms, or a dummy for the top decile of the distribution). Controls include age, gender, income, studentship, education, risk and time preferences, and choice of good cause. Results are also qualitatively similar for the other social preference measures.

Table H.4: **Measures of imbalance**

	Absolute imbalance $\lvert S-C\rvert$		Absolute imbalance $S-C$	
	No sorting	Sorting	No sorting	Sorting
	(1)	(2)	(3)	(4)
A. Differences in balance across treatments				
£0 (benchmark)	7.894	6.959	6.871	6.265
	(0.349)	(0.623)	(0.553)	(0.844)
£0.25	-2.197***	-2.192***	-7.522***	-6.128***
	(0.551)	(0.805)	(0.774)	(1.091)
£0.50	-2.570***	-2.523***	-7.632***	-5.212***
	(0.546)	(0.768)	(0.767)	(1.040)
£1	-2.717***	-3.495***	-7.111***	-6.065***
	(0.553)	(0.735)	(0.777)	(0.995)
R^2	0.043	0.064	0.163	0.111
N	708	341	708	341
B. t-tests of equality p-values				
£0.25 vs £0.50	0.490	0.627	0.884	0.321
£0.25 vs £1	0.343	0.043	0.592	0.942
£0.50 vs £1	0.786	0.103	0.493	0.291
C. Compassion				
$\geq 90^{\text{th}}$ pct.	4.590***	9.210***	5.805***	10.885***
	(1.449)	(1.986)	(1.710)	(2.432)
R^2	0.153	0.701	0.175	0.677
N	170	49	170	49
D. Social motivation				
$\geq 90^{\text{th}}$ pct.	-0.202	5.293	-0.461	6.849
	(1.706)	(3.885)	(2.007)	(4.522)
R^2	0.107	0.597	0.123	0.569
N	170	49	170	49

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors in parentheses. Results from linear regressions of balance measures on treatment dummies (Panel A) and social preference measures (Panels C and D). The significant difference between the £0.25 and £1 treatments under the $\lvert S-C\rvert$ balance measure is eliminated when adjusting for multiple hypothesis testing.

Table H.5: **Tobit models: Intensive and Extensive Margin Effects**

	Social effort units		Social effort share		Absolute balance	
	(1)	(2)	(3)	(4)	(5)	(6)
£0.25	-4.683***	-4.822***	-0.310***	-0.319***	-3.074***	-3.284***
	(0.509)	(0.509)	(0.034)	(0.034)	(0.749)	(0.742)
£0.50	-4.825***	-4.936***	-0.321***	-0.328***	-3.395***	-3.672***
	(0.508)	(0.511)	(0.034)	(0.034)	(0.750)	(0.745)
£1	-4.519***	-4.594***	-0.298***	-0.303***	-3.627***	-3.803***
	(0.504)	(0.508)	(0.034)	(0.034)	(0.748)	(0.752)
Sorting	-0.491	-0.412	-0.033	-0.028	-1.210	-1.096
	(0.645)	(0.631)	(0.043)	(0.042)	(1.069)	(1.039)
Sorting × £0.25	0.962	0.876	0.064	0.059	0.169	0.346
	(0.757)	(0.748)	(0.050)	(0.050)	(1.207)	(1.171)
Sorting × £0.50	1.572**	1.485**	0.106**	0.100**	0.102	0.023
	(0.750)	(0.736)	(0.050)	(0.049)	(1.165)	(1.134)
Sorting × £1	0.791	0.704	0.050	0.045	-0.652	-0.734
	(0.709)	(0.695)	(0.047)	(0.046)	(1.136)	(1.103)
Constant	11.832***	12.434***	0.789***	0.829***	9.074***	9.156***
	(0.402)	(0.760)	(0.027)	(0.051)	(0.640)	(1.130)
Controls	No	Yes	No	Yes	No	Yes
N	1,049	1,049	1,049	1,049	1,049	1,049
Left-censored	46 (4.38%)		46 (4.38%)		3 (0.03%)	
Right-censored	128 (12.20%)		128 (12.20%)		174 (16.59%)	
Pseudo-R^2	0.034	0.039	0.240	0.277	0.013	0.022

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered at the subject level in parentheses. The baseline category is the £0 treatment when sorting is not possible. See Table 5 for details.

Table H.6: **Aggregating Treatments: Intensive and Extensive Margin Effects**

	Social effort units		Social effort share		Absolute balance	
	(1)	(2)	(3)	(4)	(5)	(6)
Bonus	-3.744***	-3.850***	-0.248***	-0.255***	-2.494***	-2.694***
	(0.318)	(0.322)	(0.021)	(0.021)	(0.510)	(0.513)
Sorting	-0.303	-0.238	-0.020	-0.016	-0.935	-0.875
	(0.470)	(0.463)	(0.031)	(0.031)	(0.837)	(0.822)
Sorting × Bonus	0.824	0.756	0.054	0.050	-0.362	-0.333
	(0.501)	(0.495)	(0.033)	(0.033)	(0.868)	(0.851)
Constant	10.935***	11.544***	0.729***	0.769***	7.894***	7.860***
	(0.277)	(0.630)	(0.018)	(0.042)	(0.466)	(0.923)
Controls	No	Yes	No	Yes	No	Yes
N	1,049	1,049	1,049	1,049	1,049	1,049
R^2	0.151	0.175	0.150	0.173	0.062	0.108

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered at the subject level in parentheses. The baseline category is the £0 treatment when sorting is not possible (i.e. no bonus); *Bonus* captures all other treatments with a positive incentive. See Table 5 for details.

Appendix I Multiple Hypothesis Testing and Heterogeneity

Multiple hypothesis testing Throughout the analysis we compare social effort across several treatments and outcomes, thus increasing the rate of false positive discoveries (Romano and Wolf, 2005). To alleviate this concern, we follow the procedure described by List et al. (2019) to account for multiple hypothesis testing (MHT) in conducting pair-wise comparisons between the four treatments (£0, £0.25, £0.05, £1) and two outcomes (social enterprise social effort as units and shares). Table I.1 shows unadjusted p-values, List et al. (2019) multiplicity-adjusted p-values, and p-values from the application of conservative Bonferroni and Holm corrections. We perform the tests separately for the conditions with and without sorting. In comparing the no bonus group (£0) with the bonus groups (£0.25, £0.05, £1), significance is not affected: even with the strongest penalties for MHT, the effort allocation is more balanced when a bonus is present. These results hold for MHT adjustments accounting for comparisons by gender: with or without sorting, the effort allocation is more balanced when the bonus is positive.

Gender differences Women are often found to have stronger other-regarding preferences and to be more likely to engage with social, rather than commercial ventures (Croson and Gneezy, 2009; Dimitriadis et al., 2017). In our data women exhibit higher compassion, higher previous prosocial behavior, lower risk tolerance, and higher *Social motivation*, which survive MHT adjustments (Romano and Wolf, 2005). This may imply that i) women exert more social effort and potentially exhibit stronger adverse specialization, and ii) the introduction and strength of social enterprise monetary incentives may lead to different sorting patterns and effort allocation for men and women. For these reasons, our randomization was stratified by gender, allowing us to perform comparisons across groups without loss of precision.[19] When we regress social effort on treatment dummies, gender, and their interactions in Table I.2, women's social effort is less crowded out by incentives, and significantly so in the £0.50 treatment; however, the differences in the share of effort devoted to the social task are not significant when sorting is allowed. Furthermore, MHT adjustments suggest that gender differences in the effects of treatment on social effort are not significantly different for men and women, as also seen in Figure I.1. In Table I.3 we regress our motivation measures on gender, treatment dummies, and their interaction. We find that women's motivation is crowded out to a smaller extent, but not significantly so.

Previous social organization experience Individuals with previous social sector experience – working for or with non-profits or social enterprises – may differ from other individuals in two ways. Their work may have rendered them more socially motivated (Hockerts, 2017) or may have accustomed them to an institutional logic where revenue generation and commercial practices are the exception rather than the norm (Pache and Santos, 2010), so incentives may elicit different reactions from this subgroup. We create a dummy variable for individuals who have worked i) in a non-profit, ii) in a social enterprise, or iii) with a social organization and compare results across groups with and without such experience (results are similar if we also include volunteering and donations). Results for the subsamples of individuals with and without previous experience in the social sector are similar in both the sorting and non-sorting conditions. Individuals with a social sector background exert slightly less social effort, such that their effort allocation in the £0 treatment is slightly more balanced, although adverse specialization is still present. One speculative interpretation may be that, in contrast to the above expectation, over time social sector employees become attuned to organizations' financial issues and exert more effort on the commercial task to compensate for this perceived deficiency. Nonetheless, the differences between those with a social sector background and those without remain small.

[19] Bruhn and McKenzie (2009) recommend controlling for strata dummies when assessing treatment effects in regression analyses. Our regressions with and without controls (in Table 5, for example) show that controlling for gender – our stratifying variable – does not affect our overall results.

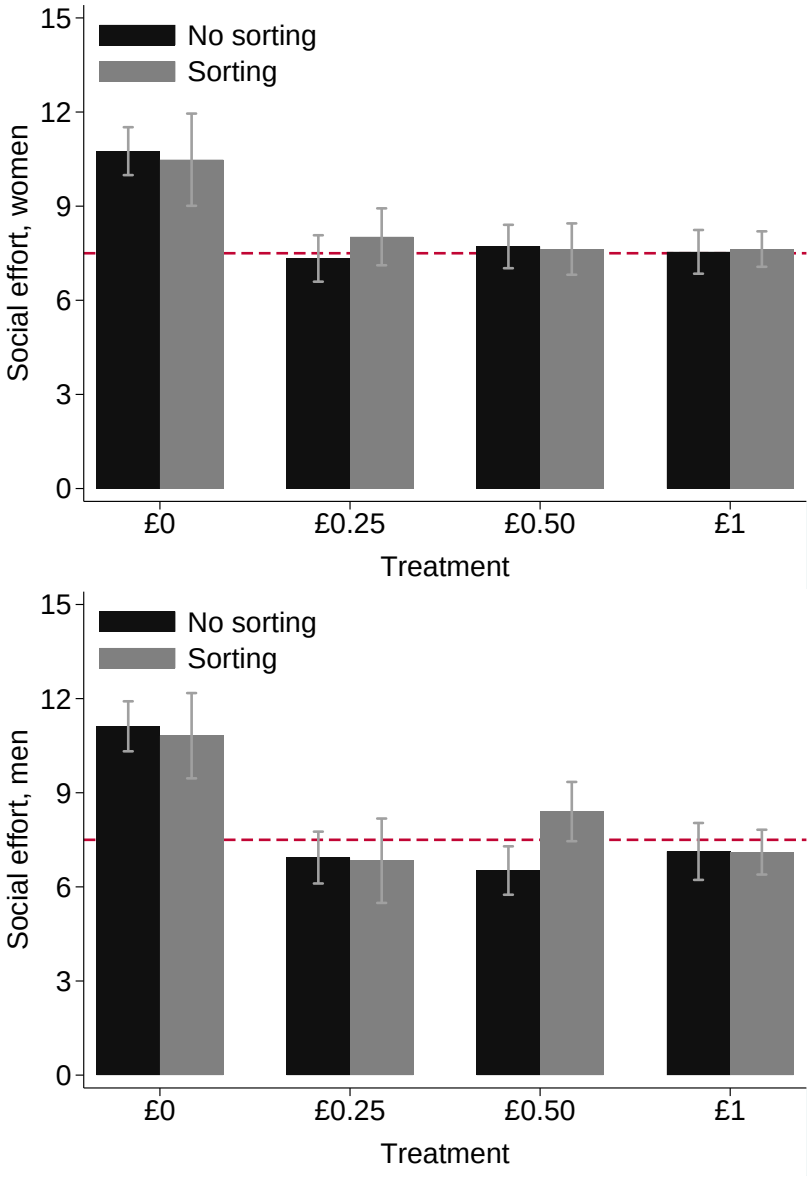

Figure I.1: Social effort in social enterprises by treatment for women (top) and men (bottom), with 95% confidence intervals; the dashed line represents a fully balanced effort allocation.

Mission heterogeneity Because social preferences may be weakly correlated with good cause choice (Table D.4), we use good cause choice dummies in our regression analyses, effectively performing within-mission analyses. However, this approach does not necessarily imply that the effects do not differ by mission, another potentially important source of heterogeneity. We therefore analyze social enterprise *Social effort* separately for each mission, summarizing the results in Figure I.2. Despite small samples in the sorting condition, the results are very similar to our pooled sample, with evidence of adverse specialization in the £0 treatment and effective balanced in the £0.25, £0.50, and £1 treatments, especially when individuals are allowed to select their preferred contract. Differences across chosen good causes are therefore limited and do not add much insight beyond our main conclusions. The uniform effects of monetary incentives across on social effort across these three representative social enterprise missions (which comprise more than 60% of issues tackled by social enterprises, Mair et al., 2012), also hints at the validity of our results for other types of missions.

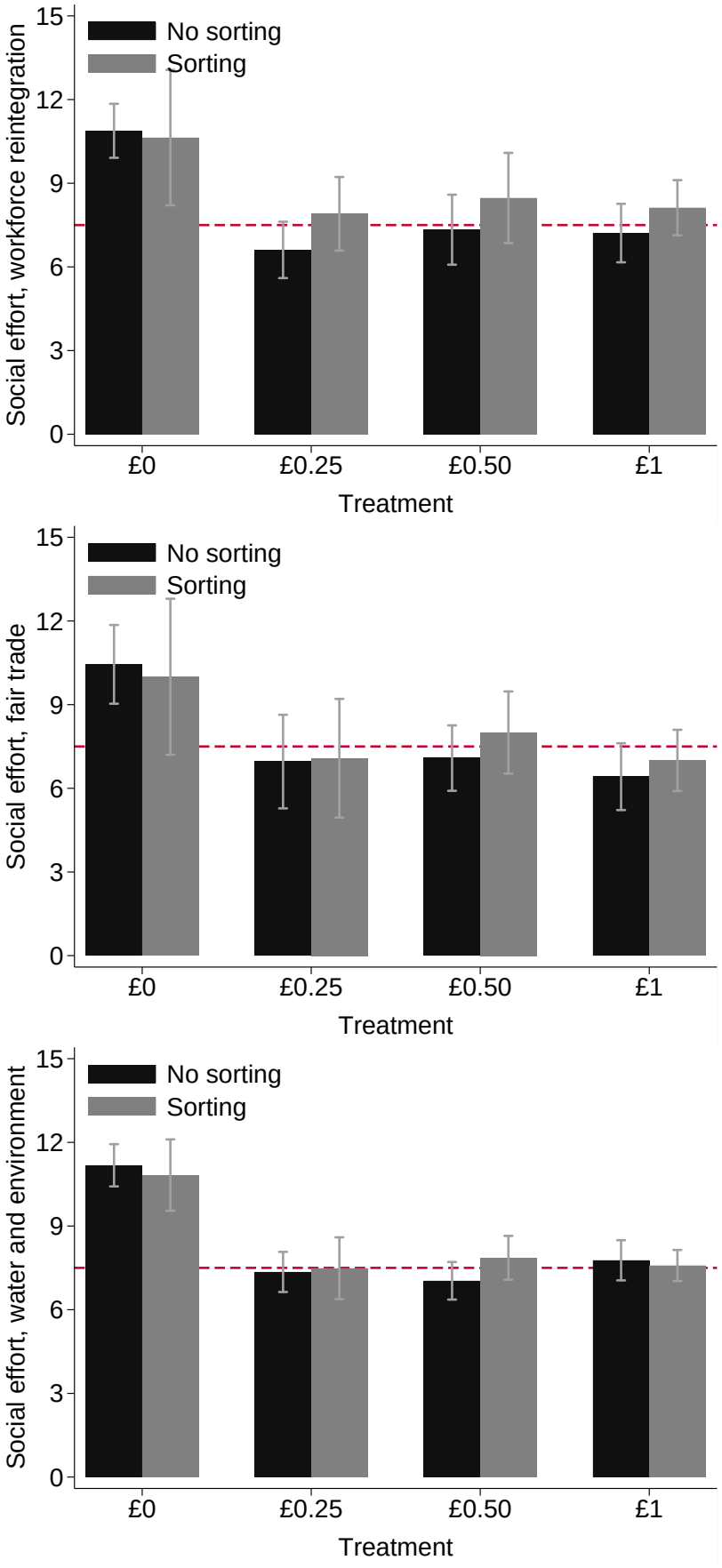

Figure I.2: Social effort in social enterprises by treatment and mission, with 95% confidence intervals; the dashed line represents a fully balanced effort allocation.

Table I.1: **Multiple Hypothesis Testing Adjustments**

Comparison			p-value			
Group 1	Group 2	Difference	Unadjusted	Adjusted	Bonferroni	Holm
(1)	(2)	(3)	(4)	(5)	(6)	(7)
A1. Units of social enterprise social effort, no sorting						
£0	£0.25	3.806	0.0003	0.0003	0.0040	0.0040
£0	£0.50	3.827	0.0003	0.0003	0.0040	0.0033
£0	£1	3.592	0.0003	0.0003	0.0040	0.0027
£0.25	£0.50	0.021	0.9553	0.9553	1.0000	0.9553
£0.25	£1	0.213	0.5843	0.8233	1.0000	1.0000
£0.50	£1	0.234	0.5503	0.8367	1.0000	1.0000
A2. Share of social enterprise social effort, no sorting						
£0	£0.25	0.251	0.0003	0.0003	0.0040	0.0037
£0	£0.50	0.254	0.0003	0.0003	0.0040	0.0030
£0	£1	0.236	0.0003	0.0003	0.0040	0.0023
£0.25	£0.50	0.002	0.9210	0.9340	1.0000	1.0000
£0.25	£1	0.015	0.5590	0.8037	1.0000	1.0000
£0.50	£1	0.017	0.4933	0.7920	1.0000	1.0000
B1. Units of social enterprise social effort, sorting						
£0	£0.25	3.098	0.0003	0.0003	0.0040	0.0040
£0	£0.50	2.632	0.0003	0.0003	0.0040	0.0030
£0	£1	3.032	0.0003	0.0003	0.0040	0.0023
£0.25	£0.50	0.465	0.3390	0.5513	1.0000	1.0000
£0.25	£1	0.065	0.8733	0.8873	1.0000	1.0000
£0.50	£1	0.400	0.3087	0.5563	1.0000	1.0000
B2. Share of social enterprise social effort, sorting						
£0	£0.25	0.204	0.0003	0.0003	0.0040	0.0027
£0	£0.50	0.173	0.0003	0.0003	0.0040	0.0037
£0	£1	0.202	0.0003	0.0003	0.0040	0.0030
£0.25	£0.50	0.031	0.3400	0.5100	1.0000	1.0000
£0.25	£1	0.003	0.9210	0.9210	1.0000	0.9210
£0.50	£1	0.028	0.2833	0.5887	1.0000	1.0000

Results from pairwise comparisons of treatment groups using the multiple hypothesis testing p-value adjustments proposed by List et al. (2019), performed separately for the conditions with or without sorting. Each test considers two outcomes (social effort as units and share) and four treatments (£0, £0.25, £0.05, £1), and produces an estimate for the unadjusted p-value, the List et al. (2019) multiplicity-adjusted p-value, and p-values from the application of conservative Bonferroni and Holm corrections.

Table I.2: **Gender Differences in Social Enterprise Effort Allocation**

	No sorting				Sorting			
	Units		Share		Units		Share	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
£0.25	-4.184***	-4.296***	-0.276***	-0.284***	-3.985***	-4.539***	-0.263***	-0.300***
	(0.577)	(0.593)	(0.038)	(0.039)	(0.919)	(0.913)	(0.061)	(0.061)
£0.50	-4.596***	-4.651***	-0.306***	-0.309***	-2.418***	-2.924***	-0.161***	-0.195***
	(0.558)	(0.562)	(0.037)	(0.037)	(0.798)	(0.814)	(0.053)	(0.054)
£1	-3.988***	-3.984***	-0.264***	-0.263***	-3.711***	-4.114***	-0.247***	-0.274***
	(0.571)	(0.582)	(0.038)	(0.039)	(0.738)	(0.758)	(0.049)	(0.051)
Female	-0.365	-0.370	-0.024	-0.025	-0.337	-0.517	-0.022	-0.034
	(0.555)	(0.587)	(0.037)	(0.039)	(0.960)	(0.956)	(0.064)	(0.064)
Female × £0.25	0.764	0.814	0.049	0.053	1.527	1.816	0.101	0.119
	(0.787)	(0.806)	(0.052)	(0.054)	(1.246)	(1.211)	(0.083)	(0.081)
Female × £0.50	1.558**	1.572**	0.105**	0.106**	-0.431	-0.374	-0.025	-0.022
	(0.762)	(0.772)	(0.051)	(0.051)	(1.143)	(1.113)	(0.076)	(0.074)
Female × £1	0.780	0.741	0.054	0.051	1.230	1.176	0.082	0.078
	(0.772)	(0.787)	(0.051)	(0.052)	(1.063)	(1.054)	(0.071)	(0.070)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
N	708	708	708	708	341	341	341	341
R^2	0.173	0.186	0.171	0.185	0.131	0.240	0.129	0.237

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. $N = 341$. Robust standard errors in parentheses. The baseline category comprises the £0 treatment for men. Multiple hypothesis testing adjustments render all interaction coefficients statistically insignificant at conventional levels ($p > 0.1$).

Table I.3: **Gender Differences in Social Preferences**

	Compassion (standardized)		Social Motivation	
	(1)	(2)	(3)	(4)
£0.25	-0.334	-0.256	-0.189	-0.108
	(0.255)	(0.247)	(0.198)	(0.200)
£0.50	-0.276	-0.208	-0.134	-0.107
	(0.253)	(0.232)	(0.194)	(0.188)
£1	-0.549**	-0.423*	-0.464***	-0.383**
	(0.236)	(0.226)	(0.170)	(0.171)
Female	-0.064	-0.080	-0.035	-0.015
	(0.251)	(0.245)	(0.185)	(0.184)
Female × £0.25	0.365	0.327	0.159	0.066
	(0.329)	(0.328)	(0.252)	(0.252)
Female × £0.50	0.403	0.436	0.121	0.151
	(0.321)	(0.304)	(0.242)	(0.236)
Female × £1	0.444	0.433	0.226	0.199
	(0.301)	(0.293)	(0.217)	(0.214)
Controls	No	Yes	No	Yes
R^2	0.047	0.149	0.049	0.140

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. $N = 341$. Robust standard errors in parentheses. The baseline category is the £0 treatment for men.

Appendix J Additional study

Purpose To ensure robustness with regards to our experimental design choices – especially in the effort required to move sliders – and investigate additional results on incentive strength, we ran an additional study. In particular, this study answers the following questions: 1) Are the incentive channel results robust to alternative experimental parameters, i.e. more sliders and a time constraint? 2) Do larger incentives distort the effort allocation towards commercial action? and 3) Do incentives affect the total effort level in social enterprises?

Design We retain the core features of our main experiment (i.e. company descriptions, slider task, compassion), abstract from features that do not affect our results (i.e. sorting, multiple contracts within-person, £0.50/slider bonus), and add others (incentives across company types, time limit, 24 sliders maximum, £2/slider bonus). We implement a straightforward 3x4 between-subject design, allocating both organizational form and incentives randomly *across* participants: each individual is exposed to only one organizational form-incentive treatment. This approach allows us to also examine how incentive effects differ by organizational form.

As summarized in Table J.1, we no longer consider the £0.50/slider bonus, but introduce a £2/slider bonus in order to check for distortions in the direction of commercial effort. The core slider task remains the same, but we now follow Gill and Prowse (2012) more closely, imposing a 1-minute time limit and increasing the possible number of sliders to 24. This approach allows us to observe variation in both effort allocation and total effort, since placing sliders is not trivial. At the end of the experiment, we collect demographics (as before), measure individual compassion (as before), and also include an attention check (as before, when we measure compassion) and two manipulation checks (one regarding the bonus, one regarding the organizational form).

We collect data on Prolific Academic using the same protocol as in our main experiment (we exclude prior participants from this additional study, but retain stratified randomization using Prolific's gender variable). We target 72 subjects per cell to ensure adequate statistical power, for a total of 864 subjects. The largest possible individual payoff is £48 (in the £2 treatment) and the average expected bonus is £30 (combining own and good cause payoffs), paid to around 1 in 10 subjects (i.e. 86 subjects), beyond the £2 participation fee that everyone receives for a study of roughly 8 minutes. A preview of the study can be accessed here.

After imposing restrictions on not taking less than 3 or more than 16 minutes to complete the study (3.8% of subjects) and placing at least 1 slider (2.3% of subjects), we obtain 811 uniformly distributed observations, with no evidence of differential attrition across treatments (χ^2 test, $p \geq 0.998$ in raw data and final sample). For this sample, good cause choices in Table J.2 are similar to those in our main study in Table D.1 and subject characteristics effectively pass a randomization check. For each covariate in Table J.3, treatment indicators are jointly insignificant (all $p > 0.150$ except for age 25-34 and comprehension time), have virtually no explanatory power (all $R < 0.025$), and a joint test across all covariates is insignificant ($p = 0.821$); covariates similarly pass the randomization test in the raw data. As in our main study we consider a slider 'correct' when placed at 23-27 and 73-77, so total effort is the sum of correct sliders; all remaining sliders – differently placed and untouched – are considered 'incorrect'.

Replication Although we alter several important parameters of our experiment, columns (5)-(7) in Panels A and B of Table J.4 suggest our results are robust: a majority of effort is dedicated to the social task (78.8%) when the social enterprise offers no commercial incentive, while both the £0.25 and £1 treatments induce a more balanced effort allocation (43.7%, $p = 0.000$ and 40.1%, $p = 0.000$, respectively). While nominally higher and not statistically different from full balance ($p = 0.108$), the share of social effort in the £0.25 treatment is not different from that in the £1 treatment ($p = 0.494$). This additional study constitutes a successful replication of our main experimental result that incentives can shift effort towards commercial tasks – especially

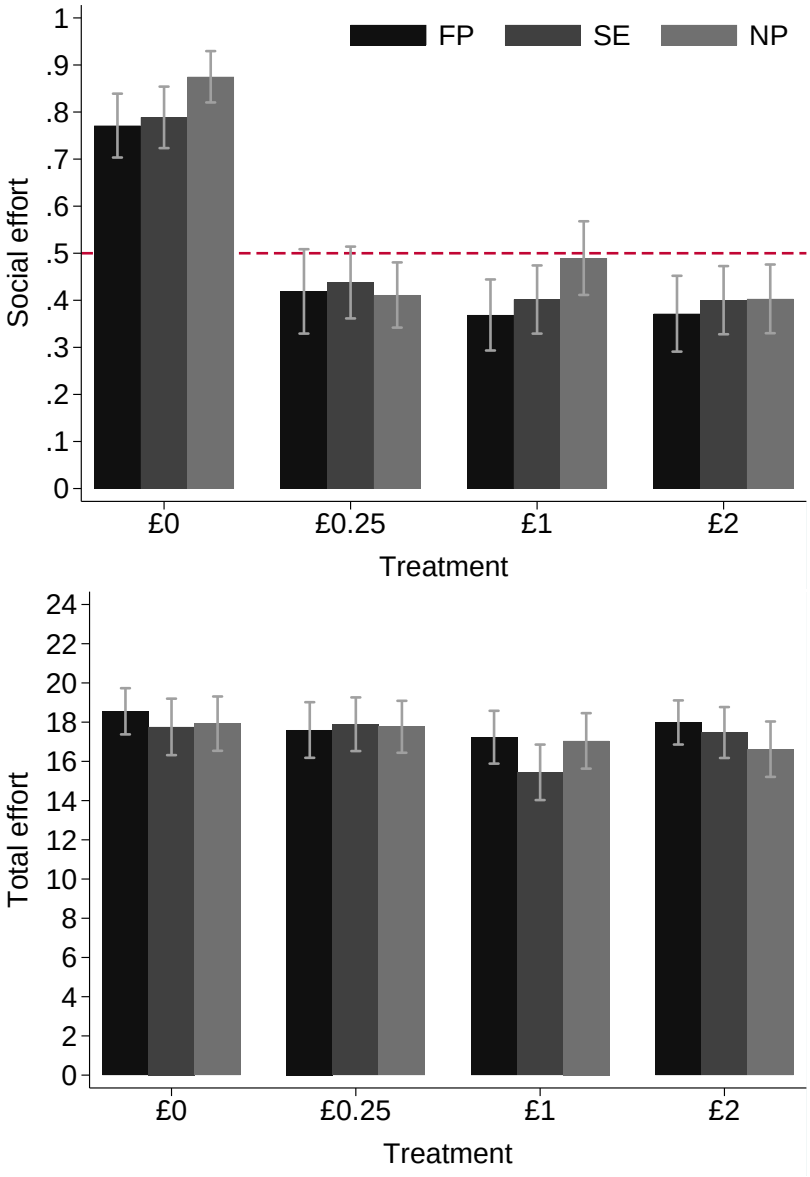

Figure J.1: Social effort as share of total effort (top) and total effort (bottom), with 95% confidence intervals; the dashed line represents a fully balanced effort allocation.

when low-powered, even when subjects allocate effort under a tight time constraint and decide on total effort, as also summarized in the top panel of Figure J.1.

Incentive strength Relative to our main study, we added a treatment with even stronger commercial task incentives in social enterprises, £2/slider. As visible in Figure J.1's top panel and in column (8) of Table J.4, Panels A and B, this higher-powered incentive does not distort the effort allocation further away from social tasks, which now receive 40% of total effort. This social effort share is different from that obtained in the £0 treatment ($p = 0.000$), but not those obtained in the £0.25 and £1 treatments ($p = 0.477$ and $p = 0.978$). This result provides further evidence that the presence of incentives is more important than their strength, thus performing an important coordination role in multitasking settings.

Total effort In our original experiment, we required subjects to place 15 sliders and, although they could choose how many sliders to place correctly (in a way that would allow them to decide total effort), most chose to place all the sliders and we detected no total effort differences across

treatments. To examine this possible margin of adjustment more closely, we designed our second study to allow subjects to choose how many of the 24 possible sliders to place in 1 minute; the sliders/time ratio is thus identical to that in Gill and Prowse (2012). On average, subjects spend 58.5 seconds on the slider task, with 81.9% using the entire time available; we find no differences across treatments in either of these measures (joint significance test, $p = 0.926$ and $p > 0.649$, respectively). Out of a total of 14,140 possible sliders, 72.6% are placed correctly.

Across treatments, subjects placed an average of 17.4 sliders correctly; only 20% of subjects placed all the sliders correctly, so we observe substantial variation in total effort. In Table J.4, Panels D and E, columns (5)-(8) show that social enterprise total effort does not vary significantly across the £0, £0.25, and £2 treatments (all pairwise $p > 0.5$), but it is lower in the £1 treatment ($0.014 \leq p \leq 0.023$), as also visible in the bottom panel of Figure J.1. Note, however, that this is the only treatment (i.e. organizational form by incentive level) where we observe lower total effort and that the differences lose statistical significance when we adjust for multiple hypothesis testing ($0.059 \leq p \leq 0.095$ for social enterprise comparisons only; $p > 0.15$ for comparisons of all organizational forms) or when we add covariates to the regression in Table J.5. Moreover, when we study the effect of incentives on social effort through OLS regressions in column (7) of Table J.5, controlling for total effort does not affect our estimates, suggesting that this does not represent an important margin of adjustment. Monetary rewards do not appear to influence total effort in the social enterprise context, but instead affect how that effort is allocated, supporting our emphasis on the *nature* of effort.

Organizational form Table J.4 allows us to compare the effect of incentives on total effort and its allocation across organizational forms: for-profits in columns (1)-(4), social enterprises in columns (5)-(8), and nonprofits in columns (9)-(12). Total effort is similar across firm types at all bonus levels (with minor differences between social enterprises and other organizations in the £1 treatment, see Panel F). In the absence of commercial task incentives, social effort is around 78% in both for-profits and social enterprises ($p = 0.711$) but is significantly higher in nonprofits (87%, $p = 0.045$ relative to social enterprises in Panel C and $p = 0.018$ relative to for-profits). Once we introduce monetary rewards, however, behavior is similar across all three organizational forms at all bonus levels: between 36.8% and 48.9% of total effort is allocated to social tasks (the only significant difference occurs between for-profits and nonprofits in the £1 treatment, $p = 0.028$). These results are displayed in Figure J.1.

We investigate differences across organizational forms more formally in Panel A of Table J.5, regressing our effort measures on incentive intensity and its interactions with firm type indicators. In column (1) we abstract from organizational form: monetary rewards induce lower total effort in the £1 treatment relative to the £0 treatment, but not in the £0.25 or £2 treatments. For reasons discussed above and given its apparent non-linearity, it is unlikely this is a meaningful and representative effect of incentives on total effort; this regression's low R^2 also suggests that incentives have little explanatory power for the number of correctly placed sliders. When we account for organizational form (without or with controls) in columns (2) and (3), the lower total effort in the £0.25 treatment becomes weaker and we uncover no differences across firm types. These results match those in Table J.4 and support our contention that total effort does not represent an important margin of adjustment.

In column (4), monetary rewards reduce social effort across all organizational forms: relative to an 81% baseline in the £0 treatment, social effort is almost halved with incentives present. However, this effect does not differ across organizational forms when we add the interaction terms without controls in column (5), with the exception of nonprofits in the £0 bonus treatment; this difference disappears once we add controls in column (6). Results are similar when we control for the total number of correctly placed sliders in column (7), highlighting that effort allocation is largely independent of total effort in this study. In Panel B we draw similar conclusions when we rerun the analysis combining the £0.25, £1, and £2 treatments.

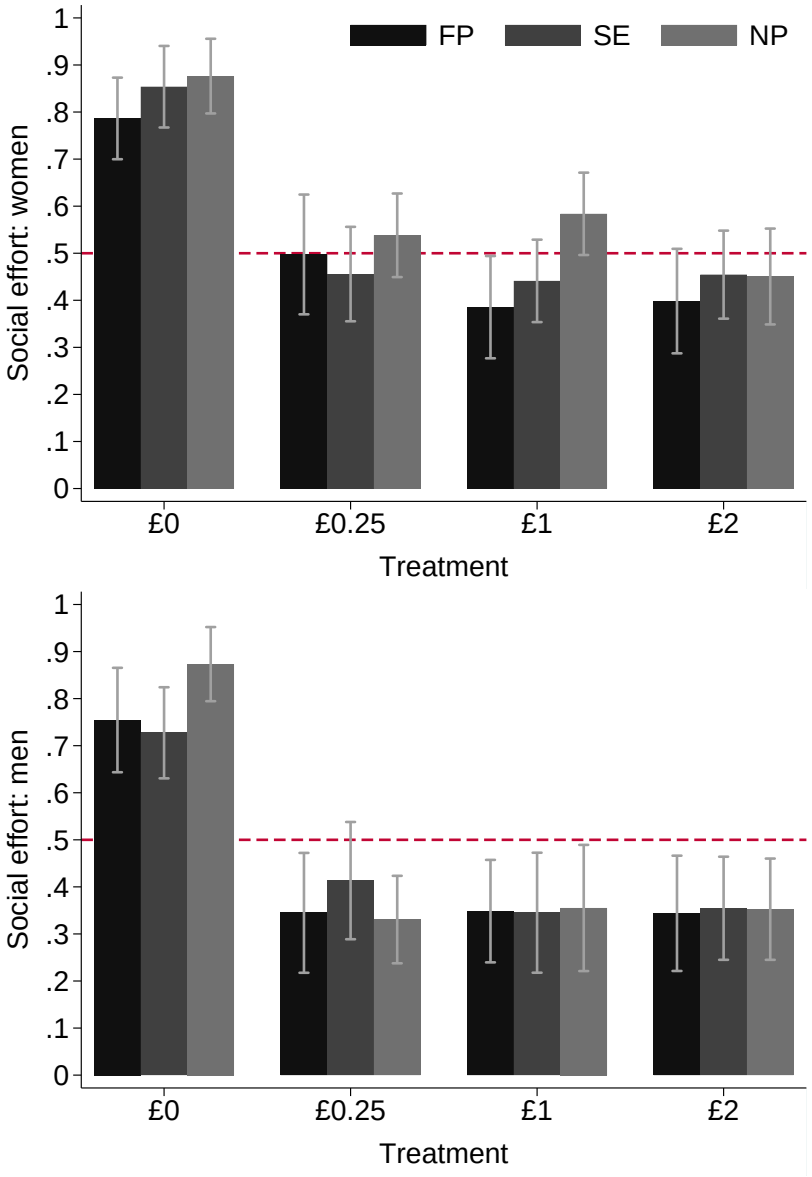

Figure J.2: Social effort for women (top) and men (bottom), with 95% confidence intervals; the dashed line represents a fully balanced effort allocation.

Overall, nonprofits elicit higher social effort when incentives are absent, for-profits and social enterprises always elicit similar responses, while any positive reward has virtually the same effect on effort allocation, regardless of firm type. These findings appear to run counter to the results of our main experiment, where firm labels do matter and we observe differences between social enterprises and for-profits in the £1 treatment. However, we caution against interpreting these organizational form results too strongly. Our primary focus was on understanding the roles of total effort and stronger incentives, with the former requiring that we impose a tight time limit of 1 minute. For this reason, it is likely that subjects were limited in their ability to engage with their allocated organizational form and adjust their effort allocation when slider placement was substantially time-constrained. In this context, subjects were likely predominantly preoccupied with the commercial and social tasks themselves, which directly determine their payoffs, allocating less attention to the organizational form; as they only saw *one* contract, they were also unable to compare *across* companies relative to our main study. Thus, it is likely we are simply capturing subject behavior in situations where both commercial and social tasks are available, a framework that describes social enterprises better than for-profits or nonprofits.

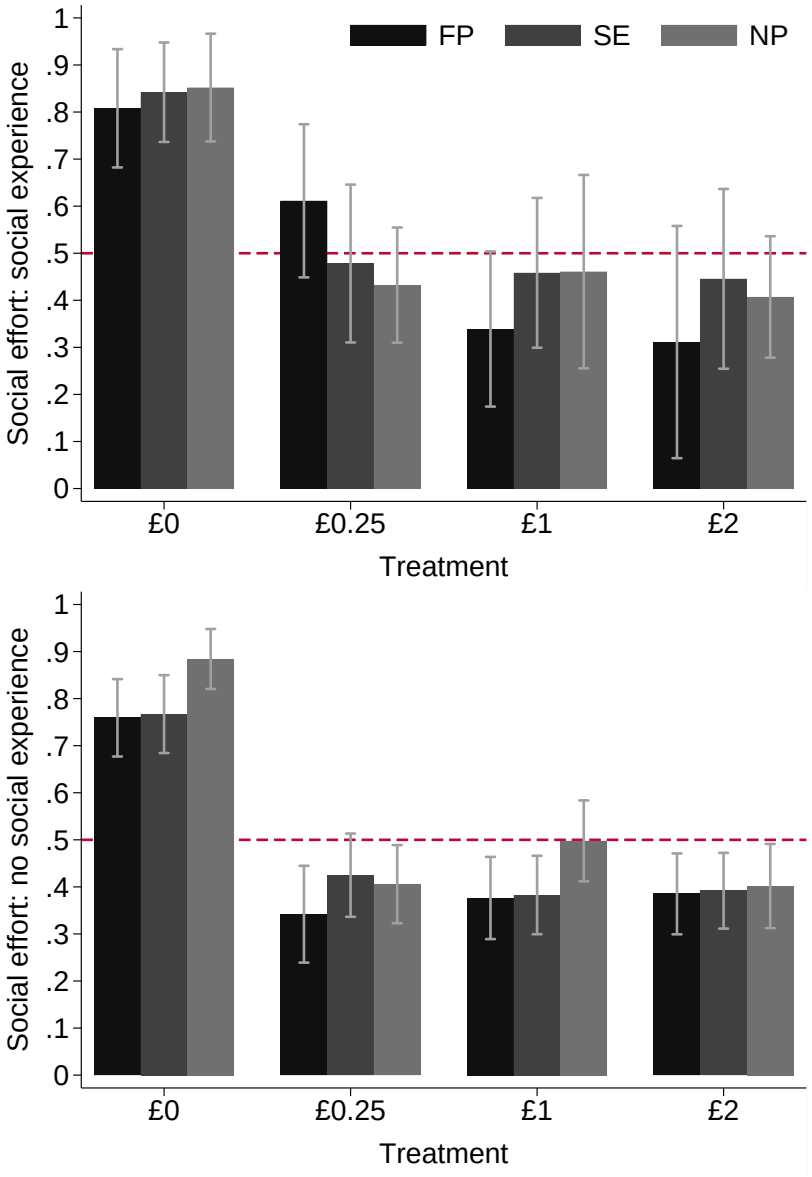

Figure J.3: Social effort for subjects with (top) and without (bottom) social sector experience, with 95% confidence intervals; the dashed line represents a fully balanced effort allocation.

Heterogeneity As we stratified our randomization procedure by gender, we can compare the effects of monetary rewards on effort allocation separately for men and women. Figure J.2 and Table J.6, Panel A show that women exert nominally higher social effort across all organizational types, both with and without incentives, but that this difference is not significant once we add controls, except for a higher share of social effort in the £1 bonus nonprofit. Overall, as in our main study, women and men behave similarly. We also collected data on prior work experience in social enterprises or nonprofits. Although we did not stratify our treatments on this variable – so results should be taken with caution –, we also check whether incentives operate differently for subjects with and without social sector experience. Figure J.3 shows that subjects with such experience exert nominally higher social effort at all incentive levels, especially high-powered, but Table J.6, Panel B shows that the differences are not significant. In other words, workers with social sector experience also respond to incentives.

Attention and manipulation checks The nature of online experiments trades off a potentially more representative sample of respondents against a controlled environment, where we can

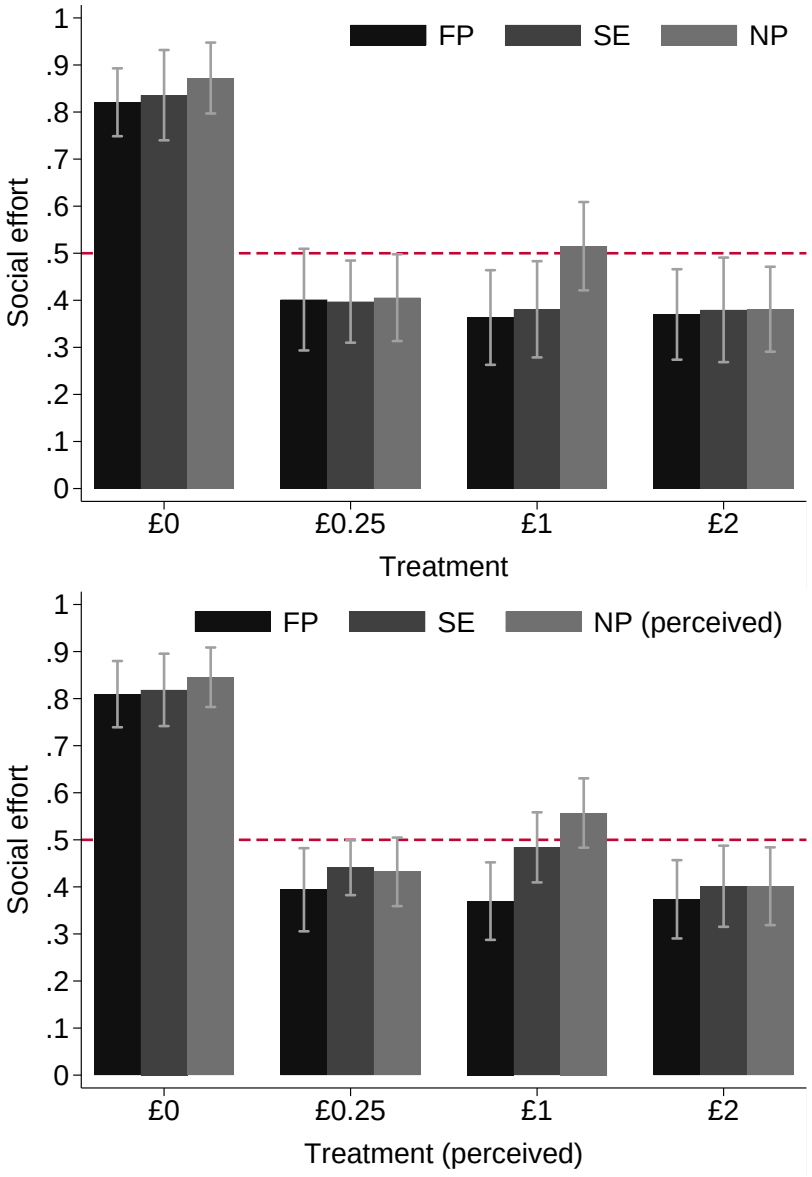

Figure J.4: Social effort for those passing all checks (top) and perceived treatments (bottom), with 95% confidence intervals; the dashed line represents a fully balanced effort allocation.

ensure subjects allocate enough attention to tasks. To alleviate this problem, we included several checks. In our main sample, 3.1% of subjects failed the attention check, 16% the incentive check, and 30.8% the organizational form check. Failing the attention check is independent of failing either other check, but subjects failing one manipulation check are more likely to also fail the other; failing checks is independent of treatment, except for a lower failure rate in £0.25 or £1 bonus for-profits. For these reasons, our analysis so far retains all observations. As a sensitivity analysis, we exclude all subjects who have failed at least one check (41.8%) and reproduce our main analysis in the top panel of Figure J.4: while noisier, these estimates are remarkably similar to our main ones. We also re-assign treatments based on the answers given to manipulation checks in the bottom panel of Figure J.4: when considering the treatments subjects *perceived*, we observe i) high social effort with a £0 incentive, ii) but similar across organizational forms, and iii) a relatively similar shift of effort towards commercial tasks when incentives are present, although slightly more muted for social enterprises and non-profits.

Table J.1: **Additional Study: Experimental Design**

Recruitment: Prolific Academic, UK-based online platform (link)
Between-subject 3×4 design:
- 3 organizational forms (for-profit/FP, social enterprise/SE, nonprofit/NP)
- 4 incentive levels (£0, £0.25, £1, £2)

Target number of subjects: 864 in total, 72 per cell, stratified by gender
Restrictions: UK resident, age 18-64, active labor force (not homemaker, disabled, retired)
Prior approval rate: $> 90\%$, to ensure high-quality answers
Participation fee: £2 for 8 minutes
Bonuses: 10% or 1 in 10 subjects, up to £48 from slider task
Slider task: maximum 24 sliders in 1 minute
Good causes: The Big Issue Foundation, Fairtrade Foundation, Water Aid
Social preferences: compassion, prior social enterprise or nonprofit work experience

Table J.2: **Additional Study: Choice of Good Cause**

	Workforce reintegration			Fair and equitable trade			Water quality and environment			Overall		
	FP	SE	NP	FP	SE	NP	FP	SE	NP	FP	SE	NP
£0	16	15	18	9	11	10	40	40	38	65	66	66
£0.25	18	17	19	16	14	15	36	35	38	70	66	72
£1	16	15	17	13	14	14	40	39	37	69	68	68
£2	17	17	16	13	13	13	35	40	37	65	70	66
N	67	64	70	51	52	52	151	154	150	269	270	272
Total	201			155			455			811		

Number of participants in each treatment that selected the given good cause.

Table J.3: **Additional Study: Randomization Check**

Characteristic	(1) Mean	(2) St. dev.	(3) Joint p-value	(4) R^2
Gender	0.515	(0.500)	0.483	0.012
Age 25-34	0.371	(0.483)	0.036	0.024
Age 35+	0.514	(0.500)	0.165	0.018
Student	0.122	(0.327)	0.173	0.017
Bachelor degree	0.425	(0.494)	0.298	0.015
Master degree	0.208	(0.406)	0.786	0.007
Low income	0.419	(0.493)	0.906	0.006
Medium income	0.418	(0.493)	0.410	0.014
High income	0.115	(0.320)	0.550	0.011
Compassion	29.260	(5.104)	0.288	0.015
Social sector experience	0.234	(0.423)	0.739	0.008
Comprehension time	43.366	(30.851)	0.014	0.019
Question time	20.064	(9.528)	0.788	0.010
Compassion time	47.502	(22.869)	0.935	0.005

Each row presents results from a separate model regressing the characteristic on treatment indicators and testing their joint significance. $N = 811$. The omitted age, education, and income categories are 18-24, high school/other, and those who prefer not to answer.

Table J.4: **Additional Study: Effort Allocation and Total Effort**

	For-profit				Social enterprise				Nonprofit			
	(1) £0	(2) £0.25	(3) £1	(4) £2	(5) £0	(6) £0.25	(7) £1	(8) £2	(9) £0	(10) £0.25	(11) £1	(12) £2
A. Social effort (share)												
	0.771	0.418	0.368	0.371	0.788	0.437	0.401	0.400	0.874	0.411	0.489	0.403
	(0.273)	(0.375)	(0.314)	(0.325)	(0.265)	(0.310)	(0.299)	(0.303)	(0.221)	(0.294)	(0.323)	(0.296)
B. Social effort *t*-tests of equality of means across treatments, *p*-values												
vs £0		0.000	0.000	0.000		0.000	0.000	0.000		0.000	0.000	0.000
vs £0.25			0.395	0.435			0.494	0.477			0.136	0.871
vs £1				0.961				0.978				0.109
C. Social effort *t*-tests of equality of means across organizational forms, *p*-values												
vs social enterprise	0.711	0.751	0.532	0.595					0.045	0.608	0.101	0.956
D. Total effort												
	18.553	17.600	17.231	17.984	17.757	17.893	15.441	17.471	17.924	17.763	17.044	16.621
	(4.753)	(5.945)	(5.605)	(4.543)	(5.857)	(5.566)	(5.847)	(5.441)	(5.633)	(5.635)	(5.842)	(5.745)
E. Total effort *t*-tests of equality of means across treatments, *p*-values												
vs £0		0.307	0.144	0.486		0.891	0.023	0.768		0.867	0.376	0.190
vs £0.25			0.707	0.675			0.014	0.662			0.459	0.240
vs £1				0.396				0.036				0.673
F. Total effort *t*-tests of equality of means across organizational forms, *p*-values												
vs social enterprise	0.394	0.766	0.069	0.554					0.867	0.891	0.112	0.377

Standard deviations in parentheses. See Table J.2 for sample sizes by treatment.

Table J.5: **Additional Study: Incentives and Organizational Form**

	Total effort			Social effort (share)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. Incentive indicators							
£0.25	-0.326	0.136	0.206	-0.390	-0.351	-0.344	-0.343
	(0.552)	(0.994)	(0.944)	(0.029)	(0.050)	(0.050)	(0.051)
£1	-1.501	-2.316	-1.904	-0.392	-0.387	-0.393	-0.403
	(0.559)	(1.011)	(0.986)	(0.029)	(0.049)	(0.046)	(0.046)
£2	-0.718	-0.286	-0.160	-0.420	-0.388	-0.370	-0.371
	(0.536)	(0.971)	(0.930)	(0.028)	(0.049)	(0.048)	(0.049)
For-profit		0.796	0.600		-0.017	-0.027	-0.024
		(0.931)	(0.899)		(0.047)	(0.046)	(0.047)
£0.25 × For-profit		-1.090	-1.079		-0.001	0.003	-0.002
		(1.357)	(1.280)		(0.076)	(0.073)	(0.074)
£1 × For-profit		0.994	0.745		-0.015	-0.010	-0.007
		(1.351)	(1.324)		(0.070)	(0.066)	(0.066)
£2 × For-profit		-0.283	-0.295		-0.011	-0.037	-0.038
		(1.268)	(1.208)		(0.072)	(0.070)	(0.071)
Nonprofit		0.167	0.349		0.086	0.049	0.051
		(1.000)	(0.926)		(0.043)	(0.044)	(0.044)
£0.25 × Nonprofit		-0.297	-0.530		-0.113	-0.067	-0.069
		(1.382)	(1.304)		(0.067)	(0.066)	(0.067)
£1 × Nonprofit		1.436	1.165		0.002	0.021	0.027
		(1.416)	(1.375)		(0.068)	(0.065)	(0.065)
£2 × Nonprofit		-1.017	-1.383		-0.083	-0.070	-0.077
		(1.387)	(1.324)		(0.067)	(0.066)	(0.066)
R^2	0.010	0.019	0.126	0.244	0.254	0.356	0.361
B. Incentive presence							
Incentive	-0.847	-0.826	-0.628	-0.400	-0.376	-0.370	-0.373
	(0.447)	(0.822)	(0.785)	(0.022)	(0.039)	(0.039)	(0.039)
For-profit		0.796	0.598		-0.017	-0.028	-0.025
		(0.928)	(0.896)		(0.047)	(0.046)	(0.047)
Incentive × For-profit		-0.130	-0.212		-0.008	-0.013	-0.014
		(1.078)	(1.038)		(0.057)	(0.055)	(0.056)
Nonprofit		0.167	0.340		0.086	0.048	0.050
		(0.996)	(0.923)		(0.042)	(0.044)	(0.044)
Incentive × Nonprofit		0.062	-0.226		-0.065	-0.037	-0.038
		(1.145)	(1.079)		(0.052)	(0.052)	(0.052)
R^2	0.004	0.007	0.117	0.243	0.249	0.352	0.357
Controls	No	No	Yes	No	No	Yes	Yes
Total effort	No	No	No	No	No	No	Yes

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. $N = 811$. Standard errors in parentheses. Results from linear regressions of total effort and social effort on incentive level or presence and their interactions with organizational form; social enterprises are the baseline category. Controls include age, gender, student status, education, income, compassion, social sector experience, comprehension time, question time, compassion time, and mission.

Table J.6: **Additional Study: Heterogeneity in Social Effort**

	For-profit		Social enterprise		Nonprofit	
	(1)	(2)	(3)	(4)	(5)	(6)
A. Gender						
£0.25	-0.410	-0.361	-0.314	-0.353	-0.543	-0.487
	(0.083)	(0.078)	(0.077)	(0.078)	(0.060)	(0.064)
£1	-0.406	-0.397	-0.382	-0.390	-0.518	-0.480
	(0.076)	(0.070)	(0.078)	(0.072)	(0.076)	(0.077)
£2	-0.411	-0.377	-0.373	-0.375	-0.521	-0.496
	(0.081)	(0.077)	(0.072)	(0.072)	(0.065)	(0.066)
Female	0.032	0.002	0.126	0.043	0.003	0.022
	(0.069)	(0.061)	(0.064)	(0.071)	(0.055)	(0.059)
£0.25 × Female	0.121	0.080	-0.084	-0.033	0.205	0.128
	(0.112)	(0.108)	(0.101)	(0.105)	(0.084)	(0.087)
£1 × Female	0.005	0.012	-0.030	-0.039	0.226	0.172
	(0.102)	(0.094)	(0.099)	(0.096)	(0.096)	(0.098)
£2 × Female	0.023	-0.048	-0.026	-0.024	0.095	0.088
	(0.106)	(0.105)	(0.095)	(0.096)	(0.091)	(0.094)
R^2	0.223	0.376	0.255	0.386	0.366	0.443
B. Social sector experience						
£0.25	-0.417	-0.380	-0.342	-0.340	-0.478	-0.428
	(0.066)	(0.062)	(0.060)	(0.060)	(0.052)	(0.051)
£1	-0.383	-0.378	-0.384	-0.389	-0.386	-0.385
	(0.060)	(0.054)	(0.059)	(0.054)	(0.054)	(0.052)
£2	-0.374	-0.384	-0.375	-0.371	-0.482	-0.453
	(0.060)	(0.058)	(0.058)	(0.056)	(0.055)	(0.056)
Experience	0.049	0.040	0.075	0.096	-0.032	-0.075
	(0.071)	(0.060)	(0.065)	(0.072)	(0.063)	(0.072)
£0.25 × Experience	0.221	0.205	-0.022	-0.108	0.059	0.028
	(0.117)	(0.107)	(0.110)	(0.112)	(0.094)	(0.102)
£1 × Experience	-0.086	-0.037	0.001	-0.079	-0.005	0.005
	(0.112)	(0.108)	(0.107)	(0.105)	(0.121)	(0.119)
£2 × Experience	-0.123	-0.085	-0.021	-0.053	0.038	0.021
	(0.137)	(0.139)	(0.113)	(0.121)	(0.097)	(0.109)
R^2	0.241	0.388	0.241	0.387	0.313	0.435
Controls	No	Yes	No	Yes	No	Yes

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors in parentheses. $N = 269$ (for-profits), 270 (social enterprises), and 272 (nonprofits). Results from linear regressions of social effort on incentive level and its interactions with gender or social sector experience; social enterprises are the baseline category. For controls, see Table J.5.

Additional References

Barigozzi, F., Burani, N., and Raggi, D. (2018). Productivity Crowding-Out in Labor Markets with Motivated Workers. *Journal of Economic Behavior and Organization*, 151:199–218.

Bellemare, C., Bissonnette, L., and Kröger, S. (2016). Simulating Power of Economic Experiments: The powerBBK Package. *Journal of the Economic Science Association*, 2(2):157–168.

Besley, T. and Ghatak, M. (2005). Competition and Incentives with Motivated Agents. *American Economic Review*, 95(3):616–636.

Besley, T. and Ghatak, M. (2018). Prosocial Motivation and Incentives. *Annual Review of Economics*, 10(1):411–438.

Bruhn, M. and McKenzie, D. (2009). In Pursuit of Balance: Randomization in Practice in Development Field Experiments. *American Economic Journal: Applied Economics*, 1(4):200–232.

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

Conway, J. M. and Huffcutt, A. I. (2003). A Review and Evaluation of Exploratory Factor Analysis Practices in Organizational Research. *Organizational Research Methods*, 6(2):147–168.

Croson, R. and Gneezy, U. (2009). Gender Differences in Preferences. *Journal of Economic Literature*, 47(2):448–474.

Delfgaauw, J. and Dur, R. (2010). Managerial Talent, Motivation, and Self-Selection Into Public Management. *Journal of Public Economics*, 94(9-10):654–660.

Dimitriadis, S., Lee, M., Ramarajan, L., and Battilana, J. (2017). Blurring the Boundaries: The Interplay of Gender and Local Communities in the Commercialization of Social Ventures. *Organization Science*, 28(5):819–839.

Frey, B. S. (1994). How Intrinsic Motivation Is Crowded Out and In. *Rationality and Society*, 6(3):334–352.

Frey, B. S. (1997). On the Relationship Between Intrinsic and Extrinsic Work Motivation. *International Journal of Industrial Organization*, 15(4):427–439.

Galizzi, M. M. and Navarro-Martinez, D. (2019). On the External Validity of Social Preference Games: A Systematic Lab-Field Study. *Management Science*, 65(3):955–1453.

Gill, D. and Prowse, V. (2012). A Structural Analysis of Disappointment Aversion in a Real Effort Competition. *American Economic Review*, 102(1):469–503.

Goretzko, D., Pham, T. T. H., and Bühner, M. (2021). Exploratory Factor Analysis: Current Use, Methodological Developments and Recommendations for Good Practice. *Current Psychology*, 40(7):3510–3521.

Heß, S. (2017). Randomization Inference with Stata: A Guide and Software. *Stata Journal*, 17(3):630–651.

Hockerts, K. (2017). Determinants of Social Entrepreneurial Intentions. *Entrepreneurship: Theory and Practice*, 41(1):105–130.

Ioannidis, J. P., Stanley, T. D., and Doucouliagos, H. (2017). The Power of Bias in Economics Research. *Economic Journal*, 127(605):F236–F265.

Lazear, E. P., Malmendier, U., and Weber, R. A. (2012). Sorting in Experiments with Application to Social Preferences. *American Economic Journal: Applied Economics*, 4(1):136–163.

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

List, J. A., Sadoff, S., and Wagner, M. (2011). So You Want to Run an Experiment, Now What? Some Simple Rules of Thumb for Optimal Experimental Design. *Experimental Economics*, 14(4):439–457.

List, J. A., Shaikh, A. M., and Xu, Y. (2019). Multiple Hypothesis Testing in Experimental Economics. *Experimental Economics*, 22(4):773–793.

Mair, J., Battilana, J., and Cardenas, J. (2012). Organizing for Society: A Typology of Social Entrepreneuring Models. *Journal of Business Ethics*, 111(3):353–373.

Pache, A.-C. and Santos, F. (2010). When Worlds Collide: The Internal Dynamics of Organizational Responses to Conflicting Logics. *Academy of Management Review*, 35(3):455–476.

Perry, J. L. (1996). Measuring Public Service Motivation: An Assessment of Construct Reliability and Validity. *Journal of Public Administration Research and Theory*, 6(1):5–22.

Romano, J. P. and Wolf, M. (2005). Stepwise Multiple Testing As Formalized Data Snooping. *Econometrica*, 73(4):1237–1282.

Young, A. (2019). Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results. *Quarterly Journal of Economics*, 134(2):557–598.