

DISCUSSION PAPER SERIES

IZA DP No. 15716

**Revenue Drift, Incentives, and Effort
Allocation in Social Enterprises**

Theodor Vladasel
Simon C. Parker
Randolph Sloof
Mirjam van Praag

NOVEMBER 2022

DISCUSSION PAPER SERIES

IZA DP No. 15716

Revenue Drift, Incentives, and Effort Allocation in Social Enterprises

Theodor Vladasel

*Universitat Pompeu Fabra and Barcelona
School of Economics*

Simon C. Parker

*Ivey Business School, Western University,
University of Aberdeen and IZA*

Randolph Sloof

*Universiteit van Amsterdam and Tinbergen
Institute*

Mirjam van Praag

*Vrije Universiteit Amsterdam, CBS, CEPR,
IZA and Tinbergen Institute*

NOVEMBER 2022

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Revenue Drift, Incentives, and Effort Allocation in Social Enterprises*

Revenue drift, where insufficient attention is given to economic, relative to social, goals, threatens social enterprise performance and survival. We argue that financial incentives can address this problem by redirecting employee attention to commercial tasks and attracting workers less inclined to fixate on social tasks. In an online experiment with varying incentive levels, monetary rewards succeed in directing worker effort to commercial tasks; high-powered incentives attract less prosocial employees, but low-powered incentives do not alter workforce composition. Social enterprises combining monetary rewards with a social mission not only attract more workers, but are also able to guard against revenue drift.

JEL Classification: D22, J33, L21, L31

Keywords: incentives, multitasking, experiment, social enterprise, prosociality

Corresponding author:

Simon C. Parker
#2253 Ivey Business School
Western University
1255 Western Road
N6G 0N1 London, ON
Canada
E-mail: sparker@ivey.uwo.ca

* We thank audiences at Copenhagen Business School, Vrije Universiteit Amsterdam, the SEI Junior Faculty Consortium, DRUID Academy, 1st IESE-LUISS Conference on Hybrid Organizations, Oxford Residence Week for Entrepreneurship Scholars, Organization Science Special Issue Conference on 'Experiments in Organizational Theory', and Stefano Brusoni, April Franco, Orsola Garofalo, Henry Sauermann, Markus Simeth, and Angelique Slade Shantz for valuable comments on previous drafts. We gratefully acknowledge the financial support of the Mærsk-McKinney-Møller Chair in Entrepreneurship at Copenhagen Business School; Vladasel also acknowledges support from the Spanish Agencia Estatal de Investigación (AEI) through the Severo Ochoa Programme for Centres of Excellence in R&D (BSE CEX2019-000915-S). Part of this research was carried out while Vladasel was a Fox International Fellow at Yale.

1 Introduction

Social enterprises that simultaneously pursue profits and purpose face the difficult challenge of allocating scarce employee attention and effort among often-competing commercial and social tasks (Austin et al., 2006; Short et al., 2009; Besley and Ghatak, 2017; McMullen and Bergman, 2017). In an incentive system geared towards mission motivation, employees frequently trade off operational concerns for beneficiary needs and ultimately allocate insufficient effort to promoting vital economic goals (Tracey et al., 2011; Smith et al., 2013; Stevens et al., 2015; Staessens et al., 2019). This type of ‘revenue drift’ (Ebrahim et al., 2014) is common across missions and business models, affecting companies tackling homelessness (Tracey and Jarvis, 2006), work integration (Bruneel et al., 2016), youth development (Beer et al., 2017), and fair trade (Davies and Doherty, 2019). Since revenue drift threatens social enterprises’ long-run financial viability and survival, as well as their ability to deliver social impact, understanding and addressing employee effort allocation is fundamental for social enterprises to achieve their promise.

Despite considerable evidence that monetary rewards can shape worker effort (e.g., Shaw and Gupta, 2015; Lazear, 2018), this potentially valuable tool has been largely overlooked in social enterprise contexts (Battilana and Lee, 2014; Brolis, 2018; Roumpi et al., 2020). We consider the link between monetary rewards and effort allocation, asking: can financial incentives for commercial action help social enterprises facing revenue drift obtain employee effort allocations closer to their desired balance over tasks? If so, how ‘steep’ should incentives be? To answer this question, we consider important social enterprise features and draw on organizational theory and economics (Eisenhardt, 1989; Kaplan and Henderson, 2005) to study ‘incentive’ and ‘selection’ effects: rewards’ attention-directing role induces *shifts in a given workforce’s effort allocation*, while their sorting role engenders *changes in workforce composition* by attracting possibly less compassionate workers (Lazear, 2000; Cadsby et al., 2007; Gerhart and Fang, 2014).

Classical multitasking models predict that commercial task incentives induce a complete shift away from social tasks (Kerr, 1975; Holmström and Milgrom, 1991), but it is theoretically unclear

whether and how strongly agents respond to incentives in prosocial environments (Miller et al., 2012; Besley and Ghatak, 2017; Brolis, 2018). We argue that when tasks are complementary for firms but place competing demands on employee attention, the absence of incentives skews effort allocations as employees favor social impact over revenue generation. We propose that social enterprises facing such revenue drift can deploy low-powered incentives (small bonuses) to elicit a less skewed division of employee effort across tasks via gentle shifts along both incentive and selection channels. In contrast, high-powered incentives (large bonuses) risk distorting effort too far towards commercial imperatives and may attract a less prosocial workforce.

We test our predictions in a pre-registered online experiment. This approach overcomes the limited availability of data on compensation practices and individual effort allocations in social enterprises, as well as the endogeneity of incentive scheme adoption; moreover, it allows us to unpack the channels linking incentives to effort allocation. We recreate essential social enterprise features in an online laboratory using salient missions and a labor market framing, where subjects act as employees of firms whose realistic descriptions match typical for-profits, nonprofits, and social enterprises. Subjects move sliders to allocate effort between commercial and social tasks associated with own and ‘good cause’ payoffs, respectively, and our treatments manipulate the strength of commercial incentives (own payoffs) in social enterprises and employees’ ability to choose their preferred organizational form, keeping good cause payoffs constant.

We find that, in the absence of incentives, social enterprise employees are highly prosocial and a large fraction allocate effort to social tasks only, thus replicating our baseline revenue drift expectation. Monetary rewards attract more workers and elicit a more balanced effort allocation regardless of strength, while we uncover a small, but significant decline in the workforce’s average compassion with high-powered incentives. Overall, sorting effects play a minor role. Instead, our primary incentive channel effect on effort allocation is robust to a large set of sensitivity analyses, as well as an additional experiment where we increase the number of sliders, impose a strict time limit, allow subjects to choose total effort, and explore stronger incentives.

The paper makes three main contributions to social entrepreneurship. First, we cast social enterprises as repositories of a multitasking problem and draw on incentive theory to highlight the revenue drift caused by an overly strong social mission emphasis. Albeit widespread (Tracey and Jarvis, 2006; Bruneel et al., 2016; Davies and Doherty, 2019), revenue drift has not received as much or as systematic attention as ‘mission drift’ (Ebrahim et al., 2014; Grimes et al., 2019).¹ Monetary incentives can help correct revenue drift, adding to a burgeoning literature on rewards in social enterprises (Brolis, 2018; Wolfolds, 2018; Moulick et al., 2019; Roumpi et al., 2020) and on alternative solutions to the trade-off between economic and social goals, such as job separation, monitoring, or hiring and socialization tools (Battilana and Dorado, 2010; Ebrahim et al., 2014; Battilana et al., 2015; Newman et al., 2018). Conceptualizing social enterprises as multitasking settings with a mission emphasis can aid future theory-building in social enterprise compensation practices and organizational design (Battilana and Lee, 2014).

Second, we implement an experimental design that causally isolates the normative, attention-directing role of incentives (Ocasio, 1997; Kaplan and Henderson, 2005; Ethiraj and Levinthal, 2009) from their effect on workforce composition (Lazear, 2000; Cadsby et al., 2007). Contrary to common beliefs among social enterprise actors that incentives may be perceived as controlling, unfair, or incongruent with organizational values (Austin et al., 2006; Tracey et al., 2011; Dees, 2012; Brolis, 2018), our results show that monetary rewards need not deter socially motivated workers from joining social enterprises. Instead, low-powered incentives work mainly by directing attention towards commercial tasks, potentially acting as a coordination device, which may be particularly important in multiple goal organizations. Our distinctive focus on (prosocial) agents’ effort allocation in a mission-driven context further complements work on social preferences and responses to prosocial incentives in non-mission settings (Andreoni and Miller, 2002; Besley and Ghatak, 2005; Tonin and Vlassopoulos, 2015; Cassar and Meier, 2018; Cassar, 2019).²

¹ Mission drift occurs when firms succumb to powerful interest group pressures diverting organizational actions away from their original social goals and toward profit-seeking (Ebrahim et al., 2014; Ramus and Vaccaro, 2017).

² Our question differs from that asked by Andreoni and Miller (2002), for instance, who directly elicit individual preferences over sharing with others in non-mission environments; our question also differs from those asked by Tonin and Vlassopoulos (2015) or Cassar (2019), who study effort responses to prosocial incentives (contributions

Third, we trace out our findings’ implications for social entrepreneurs. By embracing heterogeneity in other-regarding values (Miller et al., 2012; Stevens et al., 2015; Besley and Ghatak, 2017) and studying the channels linking incentives to effort (Gerhart and Fang, 2014), we show that monetary rewards do not necessarily crowd out workers’ social motivation or cause mission drift. This result challenges a typical view and normative pressure in social entrepreneurship that monetary rewards should be avoided (Bacchiega and Borzaga, 2001; Brolis, 2018) and reinforces the need to develop commercial skills in social ventures (Åstebro and Hoos, 2021). Interestingly, incentives attract *more* employees to social enterprises, possibly helping them alleviate talent attraction and retention problems, so we tentatively recommend that social entrepreneurs worried about revenue drift consider including low-powered incentives in their compensation schemes, possibly alongside other hiring and socialization or governance practices (Battilana and Dorado, 2010; Ebrahim et al., 2014; Smith and Besharov, 2019).

2 Theoretical Background

2.1 Social Enterprises

Social enterprises tackle social challenges by harnessing market forces (Dees, 2001; Short et al., 2009; Zahra et al., 2009). Despite diversity in business models, for-profit social enterprises are a common hybrid organization with clearly stated economic and social goals, or a *double bottom line* (Austin et al., 2006; Dacin et al., 2011; Mair et al., 2012). Social ventures occupy a continuum along the trade-off between economic and social goals (Baron, 2007; Ebrahim et al., 2014; Grimes et al., 2019; Shepherd et al., 2019), but most espouse business models where activities pertaining to each goal are not fully aligned (Gamble et al., 2020).³ Commercial and charitable pressures place competing demands on attention and resources (Dacin et al., 2011; Stevens et al., 2015), generating tensions throughout the organization. Due to social enterprises’

to charity). Our paper comes closer to work by Jones et al. (2018) on incentive effectiveness and ability sorting in multitasking environments with and without a mission; our innovation is to consider a setting tailored to social enterprises where i) a mission is present, but the organization explicitly requires effort on both tasks, and ii) we allow individuals to sort across for-profits, non-profits, and social enterprises as a function of prosociality.

³ This trade-off is largely avoided in ideal social enterprises, where commercial and social activities are perfectly aligned, creating blended value (Battilana et al., 2012; Ebrahim et al., 2014).

typically small scale, staffing constraints, and managers' inability to perfectly monitor employees (Smith et al., 2013; Doherty et al., 2014; Santos et al., 2015; Battilana, 2018), and the challenges of splitting jobs between workers (Ethiraj and Levinthal, 2009; Battilana et al., 2015), employees' discretion over how to trade off different tasks generally affects firm outcomes (Battilana and Dorado, 2010; Besharov and Smith, 2014; Henderson and Van den Steen, 2015).

Commercial and social tasks are complements for firms as revenue ultimately services social impact; yet, for workers attending both tasks they are substitutes in terms of attention and effort. For example, in work integration firms, the same employees often ensure financial sustainability through business activity and disadvantaged workers' skill development (Tracey et al., 2011; Pache and Santos, 2013; Battilana et al., 2015; Bruneel et al., 2016).⁴ When making decisions in the field, loan officers in commercial microfinance organizations must balance loan size, interest rate, repayment risk, and potential profits on one hand, with reaching the target underprivileged population on the other (Battilana and Dorado, 2010; Pache and Santos, 2010; Canales, 2014; Wry and Zhao, 2018). 'Base of the pyramid' firm employees must sell products at prices above cost, yet within the reach of the target population (Santos et al., 2015; McMullen and Bergman, 2017; Prado et al., 2022). Finally, environmental social enterprise employees must balance efforts to increase sales with costly ecological protection measures (York et al., 2016).

To guide employee effort in the absence of incentives, social ventures may use alternative, non-financial tools. Hiring and socialization practices that recognize individual motivation heterogeneity are promising candidates (Battilana and Dorado, 2010; Besharov, 2014; Hsieh et al., 2018; Newman et al., 2018), but may suffer from serious limitations. For example, new hires may have been previously socialized into a different, difficult-to-undo mindset, and asking workers to strive for balance may be impractical when founders display biases of their own. Effectively managing individual trade-offs through hiring and socialization thus requires constant dedicated

⁴ Regarding individual decision-making, one housing worker quoted in Beer et al. (2017) notes: "We need to get support workers on board with the importance of rent collection... everyone needs to have a common message [with the young people]". Another housing worker states: "They [support workers] say 'Our job is not to collect rent, it is to get them on a course'. They need to understand that money is what keeps [Youth Futures] going".

attention, often a scarce resource.⁵ Hence, while some social enterprises might rely on these instruments, we believe it is worth investigating the role of financial incentives, whose effectiveness in motivating effort has been highlighted in economics and management ([Jenkins Jr et al., 1998](#); [Lazear, 2000](#); [Gerhart and Rynes, 2003](#); [Shaw and Gupta, 2015](#)).

2.2 Employee Effort and Rewards

We study a setting where employees have discretion over how to divide effort between commercial and social tasks and assume total effort is limited: even for the most motivated employees, there are only so many hours in a day, so spending more time on one task carries the opportunity cost of spending less time on the other.⁶ This creates a trade-off between social and commercial tasks. For simplicity, we assume it is as challenging for employees to exert effort in both activities; we can relax this assumption without changing the logic of our approach. In this framework, social enterprises want employees to devote effort to commercial and social tasks in a specific manner: depending on business model or cost structure, the desired effort split could be, e.g., 30:70, 40:60, or 50:50. Even if this desired effort ‘balance’ is firm-specific, effort on only one task implies drift of some kind when the firm requires both tasks to receive attention, so our hypotheses below do not depend on what exactly ‘balance’ looks like for any firm.

Our focus on monetary rewards is driven by the idea that incentives are integral to firms’ structural distribution of attention ([Ocasio, 1997](#)), perform a normative function by directing employee effort ([Kaplan and Henderson, 2005](#)), and can help complex organizations reach their goals even when not all outcomes can be measured accurately ([Ethiraj and Levinthal, 2009](#)). In contrast to a baseline scenario where a fixed wage is paid regardless of effort allocation, incentives comprise a combination of fixed wage and variable pay for effort on a given outcome, ranging from low-powered (small bonus tied weakly to task effort) to high-powered (large bonus

⁵ Behavior-based governance tools guide individual decisions ([Ebrahim et al., 2014](#); [Smith and Besharov, 2019](#)), but effective monitoring requires resources social ventures may not possess, given scarce managerial attention and the costs of measuring multidimensional performance ([Austin et al., 2006](#); [Santos, 2012](#); [Doherty et al., 2014](#)).

⁶ When we specifically allow subjects to choose total effort in an additional study ([Appendix H](#)), we confirm that total effort is not an important margin of adjustment.

tied closely to task effort). In principle, incentives can be attached to either task (Wolffolds, 2018), but tend to be applied to commercial tasks, as social impact is difficult to measure in a timely and standardized way (Austin et al., 2006; Dacin et al., 2011; Beer et al., 2017). Rewards can affect two action margins. The ‘incentive’ channel relates to how incentives affect *a given pool* of employees’ effort allocation; the ‘selection’ channel relates to effects on the employee pool *composition* itself, recognizing that workers can self-select into different firm types (Lazear, 2000; Gerhart and Rynes, 2003; Cadsby et al., 2007; Gerhart and Fang, 2014). Prosocially motivated individuals who harbor a ‘desire to protect and promote others’ well-being’ (Grant, 2007) tend to select into social enterprises or nonprofits, whereas those extrinsically motivated by outcomes such as money tend to select into for-profits (Barigozzi et al., 2018).

2.3 No Incentives in Social Enterprises

Most social enterprises are founded by compassionate individuals seeking to maximize the social return on their investments (Miller et al., 2012; Grimes et al., 2013; Ruskin et al., 2016; Bacq and Alt, 2018). Due to imprinting, their other-oriented values permeate the entire organization and shape its attention allocation to social goals (Zahra et al., 2009; Stevens et al., 2015). Despite often acknowledging the importance of market mechanisms’ for financial sustainability, social enterprises rely heavily on the mission they champion to acquire resources (Dees, 2001; Renko, 2013; Smith et al., 2013; Allison et al., 2015; Fosfuri et al., 2016; Barberá Tomás et al., 2019). It is thus unsurprising that commercial activity is regarded as instrumental, rather than fundamental, and that the social enterprises context does not punish inferior financial performance to the same extent as commercial marketplaces (Austin et al., 2006).

In turn, the social mission emphasis attracts employees who share a deep concern for others (Besley and Ghatak, 2005, 2017; Tonin and Vlassopoulos, 2015; Burbano, 2016; Cassar, 2019), as well as purely socially motivated volunteers (Doherty et al., 2014), such that social enterprise employees are as prosocial as nonprofit employees (Borzaga and Tortia, 2006; Depedri et al., 2010; Bell and Haugh, 2014; Brolis, 2018; Roumpi et al., 2020). Moreover, social enterprise founders

and employees often have experience working in settings where ‘doing good’ trumps ‘doing well’ (Battilana and Dorado, 2010; Besharov, 2014; Hockerts, 2017). A social sector background may steer their effort allocation decisions, as ‘dangerous idealists’ emphasize the ‘social’ rather than the ‘enterprise’ aspect of the organization (Battilana and Dorado, 2010; Bacq et al., 2016). In short, the social enterprise incentive system is often geared heavily towards prosocial motivation, obscuring the need to generate revenue through commercial activity.

Social enterprises might abstain from using monetary rewards for various reasons. Incentive theory suggests that rewarding commercial tasks is likely to move employee effort away from the less easily-measured and rewarded social tasks (Holmström and Milgrom, 1991). Founders are generally averse to such an incentive channel shift, perceived as incongruent with organizational values or deviant from desired identity, and closely linked to mission drift (Zahra et al., 2009; Tracey et al., 2011; Andersson et al., 2017; Brolis, 2018; Grimes et al., 2019). Founders may also worry that incentives attract financially-driven employees uninterested in social tasks, risking mission drift via a selection channel (Bacchiega and Borzaga, 2001; Austin et al., 2006).

What are the implications of eschewing incentives? Beyond the motivational effects of mission, organizational economics suggests prosocial individuals have a lower psychic cost of exerting effort on the social task (Murdock, 2002; Borzaga and Tortia, 2006; Schnedler, 2008). Absent incentives, their effort allocation is therefore distorted towards the less costly task, leading to an ‘adverse specialization’ (MacDonald and Marx, 2001) that may reduce financial performance and limit the resources needed to survive (Smith et al., 2013). Our baseline prediction is:

Hypothesis 0 *In the absence of incentives, employees predominantly allocate effort to the social task, at the expense of the commercial task.*

Our baseline hypothesis neither asserts what an organization’s optimal balance should be, nor contends an equal split between commercial and social tasks is ideal. Its core thrust is simply that regardless of each social enterprise’s specific desired balance (Grimes et al., 2019; Shepherd et al., 2019), a mission motivation emphasis may lead workers’ effort allocation away from that

target. As long as employees are charged with executing both tasks and enjoy discretion over how they do so, effort allocations solely towards social tasks indicate that a deviation exists and must be addressed.

2.4 Incentives in Social Enterprises

Can incentives mitigate the problems caused by disproportionate mission emphasis and achieve a less skewed effort allocation? Do low- or high-powered incentives work best? Without prosocial motivation, the classic two-task multitasking problem insights would directly apply. The absence of incentives would make workers exert effort on both tasks (Holmström and Milgrom, 1991), while monetary rewards would risk dramatically unbalancing the effort allocation towards the incentivized commercial task (Kerr, 1975). However, the realistic case where a firm emphasizes its mission and some employees are prosocial complicates the analysis.

When workers differ in prosociality, a formal analysis predicts that monetary rewards signal a task's importance (Kaplan and Henderson, 2005) and make it financially attractive for employees to perform it (Holmström and Milgrom, 1991), such that *all* workers' effort allocation responds to monetary rewards (Canton, 2005). The evidence matches this prediction: diverse experimental subjects in mission environments respond (modestly) to financial incentives (Jones et al., 2018), public health agents adjust effort regardless of prosociality when monetary rewards are present (Ashraf et al., 2014), and job training agency employees shift their focus toward more trainable workers when facing placement success incentives (Heckman et al., 1997; Speckbacher, 2011).

Offering monetary rewards to prosocial employees in mission-driven social enterprises is thus likely to induce a shift towards the incentivized commercial task. Yet, the concern remains that workers may be *too* responsive to high-powered rewards. Evidence from psychology, economics, and human resources indicates that strong financial incentives can crowd out prosocial motivation, especially when perceived to have a controlling nature (Deci et al., 1999; Gneezy et al., 2011; Shaw and Gupta, 2015) and in prosocial situations, where they contrast with expectations (Hossain and Li, 2014; Deserranno, 2019; Ganguli et al., 2021).

We argue this outcome can be averted when low-powered financial incentives are used, generating a gentler shift among prosocial employees along the incentive channel by lowering workers' emphasis on the social task. For them, monetary incentives imply a trade-off between the benefit of additional income from commercial activities and the lower satisfaction from reduced effort devoted to social activities. In principle, low-powered incentives locate an intermediate position along this trade-off, without a complete social task abandonment. Low-powered incentives are less likely to be seen as controlling than high-powered ones and may be presented as supportive of social mission by promoting an even-handed approach to achieving organizational goals.

There is another, subtler way that low-powered rewards may shift effort allocation via the incentive channel. In social enterprises, the mission effectively elicits social norms, in contrast to the monetary rewards embedded in commercial settings (Dees, 2012; Bénabou and Tirole, 2016). As Brüggén and Moers (2007) show, exposure to a norm promoting the social mission can mitigate financial incentives' 'distorting' effect, where the distortion refers to the under-provision of social task effort. Social norms can soften the adjustment of (prosocial) employees' effort to incentives, producing gentler shifts (Nellas and Reggiani, 2015; Bénabou and Tirole, 2016). Low-powered incentives are less at odds with social norms than high-powered incentives: combining them with a social norm thus reduces the risk that employees 'over-react' and predominantly expend effort on the commercial task, relative to high-powered incentives.

Together, these arguments suggest that social enterprise employees respond to low-powered monetary incentives by shifting only part of their effort towards the commercial task. In contrast, high-powered ones may lead to disproportionate effort devoted to the commercial task. Thus:

Hypothesis 1 *Relative to no incentives, low-powered incentives generate a shift in effort allocation towards commercial tasks through the incentive channel.*

Hypothesis 2 *Relative to low-powered incentives, high-powered incentives generate an additional shift in effort allocation towards commercial tasks through the incentive channel.*

In addition to directing effort choices through the incentive channel, monetary rewards may operate through a selection channel, attracting more financially-motivated employees to the firm

(Lazear, 2000; Cadsby et al., 2007; Gerhart and Fang, 2014; Bénabou and Tirole, 2016).⁷ Attracting such workers can be unwelcome to social entrepreneurs (Bacchiega and Borzaga, 2001), especially if the perceived incongruence between incentives and mission endangers employees' identification with organizational goals (Akerlof and Kranton, 2005; Besharov, 2014; Andersson et al., 2017) or if incentives are viewed as 'unfair' (Tracey et al., 2011). Founders and prosocial employees often express concerns over practices originating in commercial settings and prevalent in for-profits (Lazear and Shaw, 2007; Besharov, 2014; Gerhart and Fang, 2014).

A social enterprise offering high-powered incentives may attract extrinsically motivated workers, allowing them to reap personal rewards from commercial task effort while exerting less social task effort. Prosocial employees may then consider their other-oriented values under-appreciated or even devalued by the social enterprise, reducing their motivation to perform (Akerlof and Kranton, 2005; Besharov, 2014). The shift towards fewer prosocial agents may contrast with the norm of 'doing good', implying that high-powered incentives deter prosocial workers from joining a social enterprise using such practices (Brolis, 2018) and encourage those currently working in ones that adopt incentives to quit. The outcome is to alter workforce composition, decreasing (increasing) the share of employees focusing on social (commercial) tasks. In the extreme, only financially motivated workers would join the social enterprise deploying high-powered incentives and they would allocate their entire effort to commercial tasks.

In contrast, low-powered incentives place a limit on the in-selection of extrinsically-motivated employees into social enterprises, as well as out-selection of prosocial ones. Low-powered rewards are only weakly attractive for extrinsically-motivated employees if they can still do better working in a for-profit firm, where high-powered incentives abound. This restricts their willingness to join the social enterprise, diluting their influence on the organization's goal focus, and reducing the negative impact on prosocial employees' selection (Jones et al., 2018). Moreover, a loose

⁷ Deserranno (2019) finds that financial incentives affect candidates' perception of a new public sector job's positive effects on society and discourage strongly prosocial agents from applying; in contrast, higher wages and career benefits attract motivated workers to government jobs (Dal Bó et al., 2013; Ashraf et al., 2020).

coupling of pay and performance is less likely to challenge the social norm of ‘doing good’, maintaining a prosocial tone for the organization’s values. The overall effect of a small compositional change following the introduction of low-powered incentives is a slight increase in the proportion of employees focusing more on the commercial task:

Hypothesis 3 *Relative to no incentives, low-powered incentives generate a shift in effort allocation towards commercial tasks through the selection channel.*

Hypothesis 4 *Relative to low-powered incentives, high-powered incentives generate an additional shift in effort allocation towards commercial tasks through the selection channel.*

This set of hypotheses works in the same direction as the first, but differs in the channel by which effort is affected: workforce composition, rather than existing workers’ effort allocation. Our experiment disentangles these channels and identifies which one is more salient in practice.

3 Experimental Design

Our analysis uses an online experiment with a labor market framing, replicating core contextual elements of social enterprises to enhance external validity (Ariely and Norton, 2007; Levitt and List, 2007).⁸ Subjects allocate effort to commercial and social tasks as hypothetical employees of fictional companies, whose descriptions typify those of for-profits, nonprofits, or social enterprises. These firms (‘contracts’) provide similar services but have different goals, corresponding to a realistic labor market choice. For instance, a workforce integration social enterprise is described as follows: ‘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’. In this case, attending solely to one task implies some form of drift. The equivalent for-profit and nonprofit descriptions emphasize revenue generation by providing services and the charitable goal of improving disadvantaged groups’

⁸ Lee et al. (2020) and Boulongne (2022) similarly use online experiments to study how potential investors or customers perceive social enterprises’ investment potential and legitimacy, respectively.

welfare, respectively. Both tasks are available to subjects in each contract and are described in relation to the services the company provides and the attendant payoffs.⁹

The experiment consists of four parts, summarized in Table 1. In Part 1, subjects choose the good cause they can earn money for throughout the experiment (i.e. the good cause payoff) and which provides the realistic context for the fictional firms, thereby ensuring the salience of the social task. In Part 2, all subjects perform the effort allocation task under *each* of the three different contracts (for-profit, nonprofit, social enterprise), displayed randomly to avoid order bias effects. Worker self-selection is ruled out in this setting, allowing us to study the incentive channel by varying the social enterprise bonus between (randomly selected) subjects. In contrast, in Part 3, subjects pick their preferred contract from those encountered in Part 2 and perform the effort allocation task again, allowing us to study the selection channel. Finally, in Part 4 we collect information about demographics and social preferences. We describe our experimental design below and provide the detailed experimental instructions in Online Appendix A.

Task We use a task inspired by Gill and Prowse (2012) to measure effort allocation. Our task consists of 15 horizontal sliders, labeled from 0 to 100 and initially positioned at 50, that can be moved to positions equivalent to exerting commercial or social effort. The former requires placing the slider at exactly 25 and carries a payoff for the subject, determined by the commercial task incentive level (‘bonus’) in a given contract. The latter requires placing the slider at exactly 75 and carries a payoff for the selected good cause. Incorrectly placed sliders carry no payoffs.¹⁰ Each slider has a number to its right, showing its current position. Subjects can move the mouse in any way they like to drag sliders and can readjust each slider’s position as many times as they wish. Subjects drag sliders in a way that reflects their preference, such that they perform an

⁹ To avoid priming, we did not use ‘for-profit’, ‘nonprofit’ or ‘social enterprise’ labels in our study. The latter label likely evokes a strong motivation to exert social task effort (Lee et al., 2020), amplifying revenue drift when incentives are absent. Label avoidance makes it harder to detect revenue drift, rendering our tests conservative.

¹⁰ The workforce reintegration commercial task is described as follows: ‘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 ...’. The equivalent social task is described as follows: ‘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’. As instructed, subjects are only rewarded for sliders placed at precisely 25 or 75.

actual action, as opposed to just stating their preference. While neither action has a monetary cost, both tasks require the subject to physically move a slider, i.e. real effort. The social task entails foregoing the commercial task monetary rewards, capturing a salient trade-off.

Treatments The experiment features two different dimensions, designed to tackle i) the total effect of incentives on effort allocation, and ii) the contributions of the incentive and selection channels. First, beyond the participation fee (equivalent to a fixed wage), we vary the strength of the social enterprise commercial task incentive between subjects. At one extreme, the social enterprise contract offers ‘no incentives’ (£0 bonus), broadly in line with current social enterprise practice and identical to the experimental nonprofit contract; at the other extreme, the social enterprise contract offers the same ‘high-powered incentives’ (£1 bonus) as the experimental for-profit contract; we refer to intermediate levels as ‘low-powered incentives’ (£0.25 or £0.50 bonus). Subjects are randomly assigned to a given incentive level that they face throughout the experiment. While the maximum £1 payment per slider serves as a normalization and is easily understood by subjects, the *relative* strength of incentives allows us to test our hypotheses, not the absolute value of payments per slider.

Second, we allow subjects to choose their preferred contract from the previously encountered for-profit, nonprofit, and social enterprise contracts. Part 3 incentives and company descriptions are the same as in Part 2, but we now allow for individual self-selection across company types (captured by the binary variable *Sorting*). Thus, in Part 3, employees perform the slider task in their preferred contract only. We fix the for-profit bonus at £1, the nonprofit bonus at £0, and the good cause payoff in all contracts at £1 per slider throughout the experiment. The £1 for-profit bonus allows us to benchmark the social enterprise incentive strength (i.e. the bonus is 0%, 25%, 50% or 100% of the for-profit bonus) across the range of possible incentives. The £1 good cause payoff, symmetrical to the for-profit bonus, reflects the high social returns to social task effort and is kept constant across contracts to ensure that the only difference between the

various contract terms is the extent to which commercial effort is rewarded.

Before each slider task is carried out, we explain how payoffs are calculated. In addition, we inform subjects that each slider offers a commercial task bonus between £0 and £1 when we initially describe the task. As the maximum possible bonus per slider is thus known, subjects can infer the total possible benefit for the organization; they are also aware when a social enterprise uses high-powered incentives (£1), allowing us to detect selection channel effects. We measure *Commercial effort* and *Social effort* as counts of the number of sliders (out of the 15 overall) moved to the commercial and, respectively, social task.

Preferred Mission To ensure social task salience, we allow subjects to choose their preferred good cause (Tonin and Vlassopoulos, 2015; Cassar, 2019). The three options available – The Big Issue Foundation, Fairtrade Foundation, and Water Aid – are selected as charities whose social goals match representative social enterprise missions, namely workforce reintegration, fair and equitable trade, and environmental protection (Mair et al., 2012; Shepherd et al., 2019). These missions thus likely resonate with a broad share of potential social enterprise employees.¹¹ They also have a clear multitasking component and can be pursued through a for-profit, nonprofit, or social enterprise model, allowing us to construct realistic and comparable descriptions of each contract. Furthermore, the actual charity organizational form allows us to credibly commit to donating the good cause payoffs generated by subjects in the experiment.

Social Preferences As our theoretical framework assigns an important role for individual motivation, we elicit subjects’ social preferences in various ways. Following the social enterprise literature (Miller et al., 2012; Grimes et al., 2013), we measure *Compassion* using Perry’s (1996) compassion scale, a sum of eight items on a five-point scale.¹² For example, one item asks subjects how strongly they agree with the following statement: ‘I seldom think about the welfare of people

¹¹ More specific, narrower social concerns are likely to attract employees with even stronger matching preferences, potentially exacerbating the revenue drift problem. Our choices are, therefore, rather conservative.

¹² These items, with Cronbach $\alpha = 0.75$, are available in Online Appendix A.

whom I don't know personally'. We measure *Altruism* in a standard incentivized dictator game, where subjects decide how to split a £10 endowment with another randomly paired subject (Galizzi and Navarro-Martinez, 2019). We also use a set of hypothetical and direct questions. We measure *Hypothetical altruism*, where subjects make a hypothetical donation after winning a £1,000 lottery, self-reported *Willingness to share* with others without expecting anything in return on a 0-10 scale (Falk et al., 2018), *Inequality aversion* as the recipient's minimum acceptable amount in a hypothetical ultimatum game (Fehr and Schmidt, 1999), and actual *Prosocial behavior* through past social sector experience, including donations, volunteering, and working for or with nonprofits and social enterprises (Tonin and Vlassopoulos, 2015).

These measures capture different aspects of social motivation, so we aggregate them into a composite *Social motivation* measure using common factor analysis with an orthogonal varimax rotation. This approach produces a single factor with eigenvalue larger than 1, explaining more than 80% of variance, and on which *Compassion*, *Hypothetical altruism*, and *Willingness to share* load strongly (see Online Appendix E for details and robustness checks). For brevity, we discuss *Compassion* and *Social motivation* in our main analysis and report results for other measures as robustness checks. We measure self-reported risk taking (*Risk*) and future discounting (*Time*) preferences on a 0-10 scale (Dohmen et al., 2011; Falk et al., 2018), as they may affect individuals' perception of incentives and effort allocation.

Procedure We recruit subjects on Prolific Academic, a United Kingdom-based (UK) online platform designed for surveys and experiments and geared towards researchers and startups. Prolific compares favorably with Amazon MTurk in response times, data quality, and access to diverse and representative respondents (Peer et al., 2017; Palan and Schitter, 2018). Social enterprises are an established organizational form in the UK (Tracey et al., 2011; Roumpi et al., 2020; Ganguli et al., 2021) and an online experiment allows us to reach a broad, realistic sample of potential employees whose behavior matches that of student or population-representative

samples (Mullinix et al., 2015; Farrell et al., 2017; Snowberg and Yariv, 2018). We restrict the subject pool to UK residents aged 18-64 who are students or active in the labor force. To achieve balance and perform comparisons between men and women, we stratify the randomization using subjects' reported gender on Prolific. Finally, to ensure high-quality answers, we require participants to have a history of taking Prolific studies seriously as evidenced by approved submissions in past studies and consider eligible only those with platform approval rates higher than 90%; we also include attention and manipulation checks.

Subjects receive a flat £3 participation fee (around \$4/€3.50 at the time of the experiment). In addition, the sliders allow subjects to earn up to an additional £60 (15 sliders \times 4 contracts \times £1 per slider) and the dictator game produces own payoffs up to £20 (£10 as giver, £10 as receiver). To use these rather large sums as incentives, we randomly select 40 out of the 796 subjects for bonus payment, with a maximum potential bonus of £80. We therefore compensate some participants with larger sums but a smaller likelihood (≈ 1 in 20), a procedure that Charness et al. (2016) find fully equivalent to paying smaller sums with certainty; our additional study uses a larger likelihood of payment (≈ 1 in 10) with similar results.¹³

We impose several restrictions to arrive at our final sample. First, we require subjects to have placed at most 10 (out of 60) sliders in an incorrect position, considering sliders placed at 23-27 and 73-77 as indicative of strong intentions to exert commercial or social effort and thus correct. Second, subjects must not have failed both attention and manipulation checks. Third, we require consistency between our gender variable and the Prolific variable used for stratification; in other words, a subject must have answered both questions in the same way. Finally, to ensure subjects paid attention, we require them to have completed the experiment in between 10 and 40 minutes. This leaves us with a sample of 708 subjects (out of 796 responses) for whom data quality is likely to be sufficiently high, distributed roughly evenly across treatments.¹⁴

¹³ The expected total payoff is £7 for a duration of around 20 minutes. The maximum own payoff is attained when subjects exert only commercial effort in the £1 treatment. If subjects exert only social effort, the maximum good cause payoff is £60. In practice, the average own and good cause payoffs are £29 and, respectively, £33.

¹⁴ To ensure transparency and commitment in our analysis, we preregistered our design and hypotheses within the Open Science Framework ([link](#)). Online Appendix B provides an analysis of subjects' good cause choice and

4 Results

4.1 Descriptive Statistics

Table 2 presents the descriptive statistics for our experimental sample. Out of 708 subjects, 49.9% are female and 16.5% are students. Subjects are fairly well educated, cover the range of incomes, and tend to be young. The average *Compassion* score is 29.2 (out of 40) and subjects donate on average £4.2 (out of £10) in the dictator game; 24.4%, 10.5%, and 16% of subjects have previous experience working in a nonprofit, in a social enterprise, or with a social organization, respectively. *Compassion* is positively correlated with most social preference measures and loads strongly on *Social motivation*, together with *Willingness to share*, *Hypothetical altruism*, and *Prosocial behavior*. Subjects took on average 18.5 minutes to complete the study. 25%, 18.6%, and 56.4% of subjects chose the workforce reintegration, fair trade, and environmental good causes, respectively, with similar choices across treatments.

4.2 The Incentive Channel: Social Enterprise Effort Allocation

Figure 1 plots average social enterprise *Social effort* across incentive levels and sorting conditions, with 95% confidence intervals. Absent incentives, the effort allocation is skewed towards the social task, as subjects exert more than 70% of their effort on this task (11 out of 15 sliders), as predicted by our baseline hypothesis. However, subjects allocate effort roughly equally between commercial and social tasks at all incentive levels, with confidence intervals that include a fully balanced allocation, plotted as a dashed line. Table 3, Panel A reports average *Social effort* across contracts, together with the number of subjects in each condition. The *t*-tests in the first row of Panel B shows that more balanced social enterprise effort allocations are elicited regardless of incentive strength or whether self-selection is possible ($p < 0.001$). In the remaining rows of

a randomization check. We find that the good cause chosen is not systematically related to individual traits and that our randomization procedure was successful; moreover, calculating standard errors through randomization inference produces similar results. Online Appendix C details the results of power calculations performed prior to running the experiment: the sample sizes we obtain allow us to detect relatively small changes in social effort (around half of a standard deviation) with more than 80% power at the 5% significance level. In addition, an ex post calculation of minimum detectable effects confirms our experiment is well-powered. Online Appendix D shows that results are robust to tightening or relaxing sample restrictions on gender, slider placement, and attention checks, and discusses the similar results we obtained in our pilot study.

Panel B, differences in *Social effort* between the £0.25, £0.50, and £1 treatments are negligible (p -values between 0.286 and 0.956). These findings provide initial evidence for incentive channel effects, or an attention-directing role of incentives in social enterprises.

Table 3 also allows us to compare behavior in different organizational forms. For-profit and nonprofit workers exert 30% and, respectively, 90% of their effort on the social task in each treatment; social enterprise social effort levels are in between and significantly different from for-profit and nonprofit levels ($p < 0.001$ in Panel C).¹⁵ To quantify where social enterprises lie on the for-profit/nonprofit continuum, we use t -tests to compare social enterprise social effort with the average of for-profit and nonprofit social effort in Panel D, where a positive difference indicates the social enterprise is closer to a nonprofit. Indeed, we find a positive difference for the £0 bonus ($p \leq 0.001$) and a negative one (p -values between 0.000 and 0.118) for steeper incentives. Social enterprises occupy the middle ground between for-profits and nonprofits with regards to effort allocation when financial incentives are in place and highlight the adverse specialization problem social enterprises face when they do not use monetary rewards.

We further probe the nature of adverse specialization: does imbalance result from a skewed effort allocation for all workers or only a minority? Figure 2 plots the distribution of social enterprise social effort by treatment. Between 30% and 40% of subjects in the £0 treatment exert only social effort, whereas the distribution is roughly normal and centered around a balanced allocation for other subjects. Adverse specialization thus only affects a fraction of workers, rather than shifting the entire distribution of social effort upwards. Our theoretical framework suggests that subjects with higher other-regarding preferences are more likely to exert social effort and, thus, to exhibit adverse specialization. In Online Appendix F, we show that this is indeed the case for individuals in the top *Compassion* decile, especially if they self-select into the

¹⁵ Interestingly, company descriptions matter beyond the incentives offered, even when commercial and social tasks produce the exact same payoffs per slider. The £0 bonus social enterprise is equivalent to the nonprofit in the contract terms offered, but subjects exert significantly more *Social effort* in the nonprofit contract relative to the social enterprise contract (Table 3, Panel C, columns (1) and (5)). Similarly, whereas the £1 bonus social enterprise is equivalent to the for-profit in contract terms, subjects exert significantly more *Social effort* in the social enterprise contract relative to the for-profit contract (Table 3, Panel C, columns (4) and (8)).

social enterprise. The remaining panels of Figure 2 suggest that all positive incentives reduce the fraction of social enterprise workers exerting only social effort, producing significantly different distributions, centered more tightly around a balanced effort allocation.

As hypothesis 1 predicts, low-powered incentives induce a shift in effort towards commercial tasks, alleviating the adverse specialization occurring in the absence of monetary rewards. Yet, contrary to hypothesis 2 and a common view in social enterprises, the commercial task does not gain prominence with stronger incentives: the effort allocation remains balanced even when the social enterprise pays as much as a for-profit, mitigating mission drift concerns. We now analyze the link between monetary rewards and the other-regarding preferences of workers attracted to the social enterprise and the potential for self-selection to influence effort allocation.

4.3 The Selection Channel: Worker Sorting

One reason social entrepreneurs may be reluctant to use financial incentives relates to their potential attraction of less prosocial individuals. To assess the validity of this concern, Table 4 displays average *Compassion* (standardized to have a mean of 0 and a standard deviation of 1) and *Social motivation* levels across treatments. In Panel A, we find no significant differences in mean social preferences across treatments in Part 2 of the experiment (Kruskal-Wallis $p = 0.635$ and 0.725), indicating successful randomization. Panel B displays social preferences separately for subjects selecting into the for-profit, social enterprise, and nonprofit contracts in Part 3. Social enterprise social preferences are similar for the £0, £0.25, and £0.50 treatments, but are significantly lower in the £1 treatment by about one quarter of a standard deviation, as the t -tests in Panel C show (p -values between 0.002 and 0.177 relative to the £0 treatment).¹⁶ This shift is consistent with our prediction that more extrinsically motivated employees are attracted to social enterprises when this organization uses higher-powered incentives. Nonetheless, more workers join the social enterprise as incentives become stronger: while a £0 social enterprise

¹⁶ Equality of variance F -tests do not indicate a wider social preference dispersion across treatments, but Kolmogorov-Smirnov tests for equality of distributions suggest a different distribution of *Social motivation* in the £1 treatment ($p = 0.060, 0.081, \text{ and } 0.063$). In Online Appendix E, we show that the £1 bonus social enterprise attracts more (fewer) individuals from the bottom (top) 25% of the *Social motivation* distribution.

bonus attracts 28.8% of subjects (49 out of 170), the £0.25, £0.50, and £1 bonuses attract 41%, 50.8%, and, respectively, 71.4% of subjects, mainly at the expense of for-profits.

We investigate the selection channel further by comparing social preferences across Part 3 contract choices in Panel D of Table 4. As already visible in Panel B, outside of the £1 treatment, self-selected social enterprise workers' social motivation is higher than that of for-profit workers ($p < 0.05$) and indistinguishable from that of nonprofit workers ($p > 0.1$), supporting our central contention that the hybrid organizational form attracts highly motivated employees. However, social enterprise workers' *Social motivation* in the £1 treatment is lower than that of nonprofit workers ($p = 0.014$) and more similar to that of for-profit workers ($p = 0.080$). As noted above, this is driven by a large shift in subjects choosing the social enterprise relative to the for-profit when social enterprise and for-profit contract terms are the same. In Panel E we inquire once again whether social enterprises are closer to for-profits or nonprofits in social preferences by comparing the former with the mean of the latter. While social enterprises are indistinguishable from this average in terms of *Compassion*, they are significantly closer to nonprofits with regards to *Social motivation* when the bonus is £0, £0.25, or £0.50 ($p = 0.025, 0.025, \text{ and } 0.068$). Thus, the scope for selection channel effects appears limited, with only high-powered incentives leading to a small, but significant decrease in social enterprise employee social motivation.

As Figure 1 shows, allowing for employee self-selection does not alter the relationship between monetary rewards and *Social effort*, although it leads to an effort allocation closer to full balance. To formally test for differences, columns (1) and (2) of Table 5 regress social enterprise *Social effort* on treatment dummies, the sorting condition, and their interactions, with and without demographic and good cause controls. As some participants perform the social enterprise contract twice, we cluster standard errors at the individual level. If the selection channel were unimportant, we would expect the *Sorting* variable and its interaction with each bonus level to be jointly insignificant. The first three rows replicate the results in columns (1) to (4) in Panel A of Table 3 in a regression framework, showing that incentive channel effects are both significant

and substantial ($p = 0.000$). In the next four rows, where we add the main and interaction sorting effects, we find some evidence of additional influence on effort allocation, mainly in the £0.50 treatment ($p = 0.039$ and 0.049). The joint significance test p -values at the bottom of Table 5 show that sorting differentially affects effort allocation only in the £0.50 treatment, where subjects exert higher *Social effort*. Overall, therefore, most of the action is concentrated in the incentive channel effects of monetary incentives on effort allocation.

Columns (3)-(6) of Table 5 present the results using two other dependent variables. First, we account for potential differences in total effort – which may shift social effort downwards in absolute, but not necessarily relative terms – by computing social effort as a share of total effort. The results in columns (3) and (4) perfectly match those obtained using units of social effort, suggesting that total effort does not represent an important margin of adjustment, in line with our assumption in Section 2. Second, since one could be concerned with deviations from balance in the direction of either commercial or social task effort, we also consider the absolute value of the difference between social and commercial effort, where a lower value implies better balance. Columns (5) and (6) provide additional evidence against self-selection as the main mechanism by which incentives affect effort allocation; however, they suggest that sorting does contribute to better absolute balance in the £0.50 and £1 treatments.

In summary, low-powered incentives have no effect on selection and high-powered incentives engender a loss of prosociality, but do not skew employee effort too far towards the commercial task. These results contradict our predictions, but may allay concerns of mission drift for social entrepreneurs considering using incentives. In our mission-driven context, monetary rewards work mainly through the incentive, rather than the selection channel, directing attention towards the remunerated commercial task in a way that reduces social effort concentration.

Sensitivity and heterogeneity Our results are robust to a wide set of alternative sampling restrictions (Appendix D), factor extraction/rotation techniques and social preference measures

(Appendix E), outcome measures and estimation techniques (Appendix F), and multiple hypothesis testing adjustments for the many comparisons we perform (Appendix G). As women are often found to be more other-oriented than men and more likely to engage with social, rather than commercial activities (Croson and Gneezy, 2009; Dimitriadis et al., 2017), we stratified our randomization by gender. However, we find no substantial differences between women and men in incentives' effect on effort allocation in mission-driven organizations (Appendix G). While we did not stratify our randomization by prior social sector experience or social mission, results are similar along these dimensions (Appendix G). To investigate the role of even larger incentives and analyze total effort as a possible margin of adjustment, we perform an additional study (Appendix H), where we do not allow for sorting, but add a £2 per slider treatment, increase the number of sliders, and impose a time constraint, thus inducing variation in total effort. We replicate our incentive channel results under these alternative experimental parameters: the effort allocation remains balanced even with a larger bonus, while total effort is unaffected. Overall, our results paint a clear picture: low-powered incentives induce a balanced effort allocation without an associated reduction in social motivation.

5 Discussion

To strike their desired balance between imperfectly aligned economic and social value creation goals, social enterprises must allocate scarce employee effort to both commercial and social tasks. Owing to their embeddedness in a environment espousing an emphasis on mission, a reluctance to employ practices stemming from a commercial setting, and high levels of other-regarding preferences among founders and employees, social enterprises often pay insufficient attention to revenue generation. As employees often prioritize purpose over profits, revenue drift may hinder social enterprises' ability to deliver on their social mission and threaten their survival. Why then do so few social enterprises employ monetary rewards and why does the scholarly literature pay relatively little attention to incentives in social enterprises?

We build on incentive theory to argue that social enterprise employees allocate most of their effort to social rather than commercial tasks due to social mission salience and high prosociality when monetary rewards are absent, then theorize about incentives' impact on effort allocation. We test our predictions experimentally, isolating incentive and selection channel effects. Both low- and high-powered incentives produce a more balanced effort allocation, with employee effort split roughly equally between tasks. While low-powered incentives do not affect social enterprise workforce composition, high-powered incentives lead to a small, but significant downwards shift in workers' social motivation distribution; overall, sorting plays only a minor role. Our research suggests that financial incentives could help social entrepreneurs concerned with revenue drift strike a better balance in the allocation of employee effort.

Theoretical implications The tension between social and economic value creation and its implications for firm performance are core social enterprise problems ([Short et al., 2009](#); [Dacin et al., 2011](#); [McMullen and Bergman, 2017](#); [Shepherd et al., 2019](#)). While the risk of mission drift engendered by commercial goal pursuit has received substantial attention (e.g., [Ebrahim et al., 2014](#); [Grimes et al., 2019](#)), few studies address revenue drift beyond noting threats to economic performance when firms put purpose ahead of profits ([Tracey et al., 2011](#); [Smith et al., 2013](#); [Battilana et al., 2015](#); [Stevens et al., 2015](#); [Staessens et al., 2019](#)). Even so, the usual incentive tools used by commercial ventures to guide employee effort are often seen as inappropriate in social enterprises, due to perceived incongruence with social impact and potential to attract less prosocial agents ([Bacchiega and Borzaga, 2001](#); [Austin et al., 2006](#); [Tracey et al., 2011](#); [Brolis, 2018](#); [Ganguli et al., 2021](#)). Our work highlights one explanation for revenue drift emergence, a phenomenon that deserves additional inquiry.

By modifying incentive theory to reflect the social enterprise context, we argue that monetary rewards can stave off revenue drift when commercial and social activities are imperfectly aligned. Existing studies show that firms' social missions elicit higher employee effort ([Besley and Ghatak,](#)

2005; Burbano, 2016; Cassar, 2019), but the *nature* of that effort has been largely overlooked. Social enterprises relying solely on mission to attract and motivate employees are vulnerable to revenue drift, posing obstacles to growth and survival. Combining mission and monetary rewards can instead help workers match their effort allocation to firm preferences. Thus, incentive tools arising from a commercial setting can be selectively coupled with practices prescribed by social goals (Pache and Santos, 2013) as social enterprises chart a path to their desired balance (Grimes et al., 2019). More generally, multiple-goal organizations should consider using multiple tools to motivate employees in the pursuit of their specific goals (Battilana et al., 2020).

Inasmuch as it engenders a deviation from firms' core focus on social mission, a move towards commercial action may still be considered a form of mission drift. However, a lower emphasis on social actions can afford social enterprises the resources needed for growth, allowing them to achieve social impact more reliably and sustainably in the long-run (Santos, 2012; McMullen and Bergman, 2017). This shift is especially important in light of evidence that social enterprises often use financial resources inefficiently and rarely deliver on their earned income commitment (Foster and Bradach, 2005; Mair and Martí, 2006; Staessens et al., 2019). Our results reinforce the need to move away from viewing mission drift as unequivocally bad and towards appreciating situations where it may be helpful or even necessary (Grimes et al., 2019).

Our experimental design causally isolates the channels through which incentives affect effort allocation. Monetary rewards signpost to employees which tasks are valuable for the firm: they perform a normative function (Kaplan and Henderson, 2005; Ethiraj and Levinthal, 2009) and are integral to organizations' structural distribution of attention (Ocasio, 1997); they also serve a sorting function, potentially attracting less prosocial agents (Lazear, 2000; Cadsby et al., 2007; Ganguli et al., 2021). Incentives help lower workers' social focus if motivation crowd-out is small, but may backfire if self-selected workers are extrinsically motivated. Our evidence aligns with an attention-directing role of monetary rewards, as low-powered incentives suffice to gently shift the effort allocation without reducing prosociality. By conveying a given task's importance,

low-powered incentives act as a coordination device and may be especially valuable when firms face conflicting goals or when a dominant goal prescribes and legitimates non-optimal practices (Kogut and Zander, 1996; Lounsbury, 2007), conditions that characterize social enterprises.

Monetary rewards for commercial action are but one tool available to social entrepreneurs. Additional means of guiding employee effort include hiring and socialization practices (Battilana and Dorado, 2010; Hsieh et al., 2018), governance mechanisms (Ebrahim et al., 2014; Battilana et al., 2015; Smith and Besharov, 2019), and (where quantifiable) rewards for achieving social impact or operational goals (Wolffolds, 2018). Assessing these practices' complementarity and comparative efficacy with incentive tools across social enterprises with diverse business models is important for future research. Our multitasking framework can aid such investigations, as it can incorporate different organizational forms, activity costs, and employee preferences to predict when deviations from desired balance are more likely and what solutions work best.

Practical implications Facing growing competitive pressures, social enterprises are increasingly turning their attention to operational sustainability. Monetary rewards – implemented as a share of generated revenue – offer a way for social enterprises to overcome a predominant focus on social impact to the detriment of revenue generation. Low-powered incentives are especially appealing, as they expand the potential employee pool and restore balance at minimal cost; they are unlikely to reduce employee social motivation and increase risk taking or myopic behavior, and can be expected to have minimal adverse effects on firm activities outside the scope of our study. Rather than attracting less motivated workers or over-emphasizing commercial tasks' relevance, low-powered incentives signal certain actions' value to employees. Social entrepreneurs could thus devote attention to internal effort allocations and compensation design, rather than worry excessively about the pool of potential applicants.

Since paying employees larger bonuses can attract more candidates, incentives may aid social enterprises attract talent. Social entrepreneurs who deploy monetary rewards may also reassure

potential investors of their commitment to commercial success, increasing the latter's willingness to invest. Moreover, redirecting part of employees' attention to revenue generation can improve social enterprises' cash flow, reduce dependence on outside finance, and improve their growth and survival prospects. More generally, efforts to develop entrepreneurial skills in social ventures may serve both profits and purpose (Åstebro and Hoos, 2021).

Limitations Reflecting our interest in how monetary rewards affect effort allocation, we studied a wide set of incentive intensities in social enterprises, from as weak as in nonprofits to as strong as in for-profits. While we conceptually covered the full range of relevant intensities, our experimental parameters might still have limited this range, as the good cause payoff was always set to £1 and the own payoff was set to be at most £1 per slider. The relatively high return to social action was chosen to ensure this task was on a level playing field with the commercial task, but one may worry our monetary incentives are overly weak. Stronger incentives, exceeding the £2 bonus from our additional study, could lead to different findings and explain why we did not find support for hypotheses 2 and 4. Yet, this does not affect our conclusion that low-powered incentives may help, rather than hurt social enterprises.

To causally isolate if and how monetary incentives affect effort allocation, we purposely simplified the nature of social enterprises at the expense of their hybrid complexity and considered a one-shot task. One could question our study's external validity, given the multitude and relative difficulty of tasks employees perform, the dimensions in which they differ, possible differences between short- and long-term responses to incentives, and firms' nuanced real life presentation of objectives. Nonetheless, we believe our experiment captures the fundamental tension between commercial and social tasks and employee effort allocation in a simple, transparent way. While we attempted to alleviate external validity concerns by recruiting participants from a representative potential employee pool and offering realistic choices of company types and missions, field replications with alternative designs, missions, or tasks are a natural extension of our work.

Future research Our experiment raises several interesting questions. First, social enterprise success may depend on coordinated action and group work. Are individual incentives better than team ones? Does their effectiveness vary with employee heterogeneity? Second, the introduction of incentives is often accompanied by justifications, helping the practice gain legitimacy. Do goal or communication clarity matter for their success? How do social enterprises implement and justify rewards changes? Third, incentive adoption likely affects stakeholders more broadly. Do targeted communities commend such social ventures for professionalism or downgrade them for courting mission drift? Fourth, pecuniary rewards may have other, possibly negative effects on social ventures. Self-selected employees' inequality aversion is unaffected by incentive strength (Appendix E), but bonuses may reveal large productivity gaps and engender possibly dangerous comparisons. Do incentives create internal conflicts between employees with different social preferences? Do they generate long run distortions as commercial activity becomes entrenched or does their impact wane over time? What are joiners' commitment and retention rates? For social enterprises to achieve their promise, it is crucial to answer these questions. Fifth, the large variation in the ability to measure social impact may affect incentives effectiveness. Research on social impact measurement and drift potential in a multitasking framework may prove valuable for understanding social enterprise tensions and challenges.

Conclusion We posit and provide evidence that social enterprises can address revenue drift – a dominant focus on social impact at the expense of revenue generation – by deploying monetary rewards. Low-powered incentives redirect employee attention to commercial tasks and reinforce social enterprises' commitment to achieving their mission via market-based mechanisms without attracting less prosocial employees. Our findings question a common social entrepreneurship view that incentives are incompatible with social impact and attract the 'wrong' workers. This study adds to a growing literature on the challenges of managing social enterprise tensions and opens up interesting questions on their compensation practices and organizational design.

References

- Akerlof, G. A. and Kranton, R. E. (2005). Identity and the Economics of Organizations. *Journal of Economic Perspectives*, 19(1):9–32.
- Allison, T. H., Davis, B. C., Short, J. C., and Webb, J. W. (2015). Crowdfunding in a Prosocial Microlending Environment: Examining the Role of Intrinsic Versus Extrinsic Cues. *Entrepreneurship Theory and Practice*, 39(1):53–73.
- Andersson, O., Huysentruyt, M., Miettinen, T., and Stephan, U. (2017). Person–Organization Fit and Incentives: A Causal Test. *Management Science*, 63(1):73–96.
- Andreoni, J. and Miller, J. (2002). Giving According To GARP. *Econometrica*, 70(2):737–753.
- Ariely, D. and Norton, M. I. (2007). Psychology and Experimental Economics: A Gap in Abstraction. *Current Directions in Psychological Science*, 16(6):336–339.
- Ashraf, B. N., Bandiera, O., Davenport, E., and Lee, S. S. (2020). Losing Prosociality in the Quest for Talent? Sorting, Selection, and Productivity in the Delivery of Public Services. *American Economic Review*, 110(5):1355–1394.
- Ashraf, N., Bandiera, O., and Jack, B. K. (2014). No Margin, No Mission? A Field Experiment on Incentives for Public Service Delivery. *Journal of Public Economics*, 120(1):1–17.
- Åstebro, T. and Hoos, F. (2021). Impact Measurement Based on Repeated Randomized Control Trials: The Case of a Training Program to Encourage Social Entrepreneurship. *Strategic Entrepreneurship Journal*, 15(2):254–278.
- Austin, J., Stevenson, H., and Wei-Skillern, J. (2006). Social and Commercial Entrepreneurship: Same, Different, or Both? *Entrepreneurship Theory and Practice*, 30(1):1–22.
- Bacchiega, A. and Borzaga, C. (2001). Social Enterprises as Incentive Structures: An Economic Analysis. In Borzaga, C. and Degourny, J., editors, *The Emergence of Social Enterprise*, pages 273–295. Routledge, London, UK.
- Bacq, S. and Alt, E. (2018). Feeling Capable and Valued: A Prosocial Perspective on the Link between Empathy and Social Entrepreneurial Intentions. *Journal of Business Venturing*, 33(3):333–350.
- Bacq, S., Hartog, C., and Hoogendoorn, B. (2016). Beyond the Moral Portrayal of Social Entrepreneurs: An Empirical Approach to Who They Are and What Drives Them. *Journal of Business Ethics*, 133(4):703–718.
- Barberá Tomás, D., Castelló, I., de Bakker, F. G., and Zietsma, C. (2019). Energizing Through Visuals: How Social Entrepreneurs Use Emotion-Symbolic Work for Social Change. *Academy of Management Journal*, 62(6):1789–1817.
- 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.
- Baron, D. P. (2007). Corporate Social Responsibility and Social Entrepreneurship. *Journal of Economics and Management Strategy*, 16(3):683–717.
- Battilana, J. (2018). Cracking the Organizational Challenge of Pursuing Joint Social and Financial Goals: Social Enterprise as a Laboratory to Understand Hybrid Organizing. *M@n@gement*, 21(4):1278–1305.
- Battilana, J. and Dorado, S. (2010). Building Sustainable Hybrid Organizations: The Case of Commercial Microfinance Organizations. *Academy of Management Journal*, 53(6):1419–1440.
- Battilana, J. and Lee, M. (2014). Advancing Research on Hybrid Organizing – Insights from the Study of Social Enterprises. *Academy of Management Annals*, 8(1):397–441.
- Battilana, J., Lee, M., Walker, J., and Dorsey, C. (2012). In Search of the Hybrid Ideal. *Stanford Social Innovation Review*, 10(3):51–55.
- Battilana, J., Obloj, T., Pache, A.-C., and Sengul, M. (2020). Beyond Shareholder Value Maximization: Accounting for Financial/Social Tradeoffs in Dual-Purpose Companies. *Academy of Management Review*, forthcoming.
- Battilana, J., Sengul, M., Pache, A.-C., and Model, J. (2015). Harnessing Productive Tensions in Hybrid Organizations: The Case of Work Integration Social Enterprises. *Academy of Management Journal*, 58(6):1658–1685.

- Beer, H., Micheli, P., and Besharov, M. L. (2017). Pursuing Social Purposes: The Complexity of Performance Measurement in Social Enterprises. *Working Paper*.
- Bell, B. and Haugh, H. (2014). Working for a Social Enterprise: An Exploration of Employee Rewards and Motivations. In Denny, S. and Seddon, F., editors, *Social Enterprise: Accountability and Evaluation Around the World*, pages 67–84. Routledge, Abingdon, UK.
- 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.
- Besharov, M. L. (2014). The Relational Ecology of Identification: How Organizational Identification Emerges When Individuals Hold Divergent Values. *Academy of Management Journal*, 57(5):1485–1512.
- Besharov, M. L. and Smith, W. K. (2014). Multiple Institutional Logics in Organizations: Explaining Their Varied Nature and Implication. *Academy of Management Journal*, 39(3):364–381.
- Besley, T. and Ghatak, M. (2005). Competition and Incentives with Motivated Agents. *American Economic Review*, 95(3):616–636.
- Besley, T. and Ghatak, M. (2017). Profit with Purpose? A Theory of Social Enterprise. *American Economic Journal: Economic Policy*, 9(3):19–58.
- Bénabou, R. and Tirole, J. (2016). Bonus Culture: Competitive Pay, Screening, and Multitasking. *Journal of Political Economy*, 124(2):305–370.
- Borzaga, C. and Tortia, E. (2006). Worker Motivations, Job Satisfaction, and Loyalty in Public and Nonprofit Social Services. *Nonprofit and Voluntary Sector Quarterly*, 35(2):225–248.
- Boulongne, R. (2022). A Cognitive Approach to the Expected Value of Work Integration Social Enterprises. *Strategic Entrepreneurship Journal*, forthcoming.
- Brüggen, A. and Moers, F. (2007). The Role of Financial Incentives and Social Incentives in Multitask Settings. *Journal of Management Accounting Research*, 19(1):25–50.
- Brolis, O. (2018). Do Social Enterprises Attract Workers Who Are More Pro-socially Motivated than Their Counterparts in For-profit Organizations to Perform Low-skilled Jobs? *International Journal of Human Resource Management*, 29(20):2861–2879.
- 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.
- Bruneel, J., Moray, N., Stevens, R., and Fassin, Y. (2016). Balancing Competing Logics in For-Profit Social Enterprises: A Need for Hybrid Governance. *Journal of Social Entrepreneurship*, 7(3):263–288.
- Burbano, V. C. (2016). Social Responsibility Messages and Worker Wage Requirements: Field Experimental Evidence from Online Labor Marketplaces. *Organization Science*, 27(4):1010–1028.
- Cadsby, C. B., Song, F., and Tapon, F. (2007). Sorting and Incentive Effects of Pay for Performance: An Experimental Investigation. *Academy of Management Journal*, 50(2):387–405.
- Canales, R. (2014). Weaving Straw into Gold: Managing Organizational Tensions Between Standardization and Flexibility in Microfinance. *Organization Science*, 25(1):1–28.
- Canton, E. (2005). Power of Incentives in Public Organizations When Employees Are Intrinsically Motivated. *Journal of Institutional and Theoretical Economics*, 161(4):664–680.
- Cassar, L. (2019). Job Mission as a Substitute for Monetary Incentives: Benefits and Limits. *Management Science*, 65(2):896–912.
- Cassar, L. and Meier, S. (2018). Nonmonetary Incentives and the Implications of Work as a Source of Meaning. *Journal of Economic Perspectives*, 32(3):215–238.
- Charness, G., Gneezy, U., and Halladay, B. (2016). Experimental Methods: Pay One or Pay All. *Journal of Economic Behavior and Organization*, 131(A):141–150.
- 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.
- Crosan, R. and Gneezy, U. (2009). Gender Differences in Preferences. *Journal of Economic Literature*, 47(2):448–474.

- Dacin, M. T., Dacin, P. A., and Tracey, P. (2011). Social Entrepreneurship: A Critique and Future Directions. *Organization Science*, 22(5):1203–1213.
- Dal Bó, E., Finan, F., and Rossi, M. (2013). Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service. *Quarterly Journal of Economics*, 128(3):1169–1218.
- Davies, I. A. and Doherty, B. (2019). Balancing a Hybrid Business Model: The Search for Equilibrium at Cafédirect. *Journal of Business Ethics*, 157(4):1043–1066.
- Deci, E. L., Ryan, R. M., and Koestner, R. (1999). A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation. *Psychological Bulletin*, 125(6):627–668.
- Dees, J. G. (2001). The Meaning of Social Entrepreneurship. *Working Paper*.
- Dees, J. G. (2012). A Tale of Two Cultures: Charity, Problem Solving, and the Future of Social Entrepreneurship. *Journal of Business Ethics*, 111(3):321–334.
- Depedri, S., Tortia, E., and Carpita, M. (2010). Incentives, Job Satisfaction, and Performance: Empirical Evidence in Italian Social Enterprises. *Working Paper*.
- Deserranno, E. (2019). Financial Incentives as Signals: Experimental Evidence from the Recruitment of Village Promoters in Uganda. *American Economic Journal: Applied Economics*, 11(1):277–317.
- 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.
- Doherty, B., Haugh, H., and Lyon, F. (2014). Social Enterprises as Hybrid Organizations: A Review and Research Agenda. *International Journal of Management Reviews*, 16(4):417–436.
- Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences. *Journal of the European Economic Association*, 9(3):522–550.
- Ebrahim, A., Battilana, J., and Mair, J. (2014). The Governance of Social Enterprises: Mission Drift and Accountability Challenges in Hybrid Organizations. *Research in Organizational Behavior*, 34(1):81–100.
- Eisenhardt, K. M. (1989). Agency Theory: An Assessment and Review. *Academy of Management Review*, 14(1):57–74.
- Ethiraj, S. K. and Levinthal, D. (2009). Hoping for A to Z While Rewarding Only A: Complex Organizations and Multiple Goals. *Organization Science*, 20(1):4–21.
- Falk, A., Becker, A., Dohmen, T., Huffman, D., and Sunde, U. (2018). The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences. *Working Paper*.
- Farrell, A. M., Grenier, J. H., and Leiby, J. (2017). Scoundrels or Stars? Theory and Evidence on the Quality of Workers in Online Labor Markets. *The Accounting Review*, 92(1):93–114.
- Fehr, E. and Schmidt, K. M. (1999). A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics*, 114(3):817–868.
- Fosfuri, A., Giarratana, M. S., and Roca, E. (2016). Social Business Hybrids: Demand Externalities, Competitive Advantage, and Growth Through Diversification. *Organization Science*, 27(5):1275–1289.
- Foster, W. and Bradach, J. (2005). Should Nonprofits Seek Profits? *Harvard Business Review*, 83(2):92–100.
- 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.
- Gamble, E. N., Parker, S. C., and Moroz, P. W. (2020). Measuring the Integration of Social and Environmental Missions in Hybrid Organizations. *Journal of Business Ethics*, 167(2):271–284.
- Ganguli, I., Huysentruyt, M., and Coq, C. L. (2021). How Do Nascent Social Entrepreneurs Respond to Rewards? A Field Experiment on Motivations in a Grant Competition. *Management Science*, 67(10):6294–6316.
- Gerhart, B. and Fang, M. (2014). Pay for (Individual) Performance: Issues, Claims, Evidence and the Role of Sorting Effects. *Human Resource Management Review*, 24(1):41–52.

- Gerhart, B. and Rynes, S. L. (2003). *Compensation: Theory, Evidence, and Strategic Implications*. Sage, Thousand Oaks, CA.
- Gill, D. and Prowse, V. (2012). A Structural Analysis of Disappointment Aversion in a Real Effort Competition. *American Economic Review*, 102(1):469–503.
- Gneezy, U., Meier, S., and Rey-Biel, P. (2011). When and Why Incentives (Don't) Work to Modify Behavior. *Journal of Economic Perspectives*, 25(4):191–210.
- 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.
- Grant, A. M. (2007). Relational Job Design and the Motivation to Make a Prosocial Difference. *Academy of Management Review*, 32(2):393–417.
- Grimes, M. G., McMullen, J. S., Vogus, T. J., and Miller, T. L. (2013). Studying the Origins of Social Entrepreneurship: Compassion and the Role of Embedded Agency. *Academy of Management Review*, 38(3):460–464.
- Grimes, M. G., Williams, T. A., and Zhao, E. Y. (2019). Anchors Aweigh: The Sources, Variety, and Challenges of Mission Drift. *Academy of Management Review*, 44(4):819–845.
- Heckman, J. J., Heinrich, C., and Smith, J. (1997). Assessing the Performance of Performance Standards in Public Bureaucracies. *American Economic Review*, 87(2):389–395.
- Henderson, R. and Van den Steen, E. (2015). Why Do Firms Have ‘Purpose’? The Firm’s Role as a Carrier of Identity and Reputation. *American Economic Review: Papers & Proceedings*, 105(5):326–330.
- 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.
- Holmström, B. and Milgrom, P. (1991). Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics, and Organization*, 7(3):24–52.
- Hossain, T. and Li, K. K. (2014). Crowding Out in the Labor Market: A Prosocial Setting Is Necessary. *Management Science*, 60(5):1148–1160.
- Hsieh, Y.-C., Weng, J., and Lin, T. (2018). How Social Enterprises Manage Their Organizational Identification: A Theoretical Framework of Identity Management Approach through Attraction, Selection, and Socialization. *International Journal of Human Resource Management*, 29(20):2880–2904.
- Ioannidis, J. P., Stanley, T. D., and Doucouliagos, H. (2017). The Power of Bias in Economics Research. *Economic Journal*, 127(605):F236–F265.
- Jenkins Jr, G. D., Mitra, A., Gupta, N., and Shaw, J. D. (1998). Are Financial Incentives Related to Performance? A Meta-Analytic Review of Empirical Research. *Journal of Applied Psychology*, 83(5):777–787.
- Jones, D., Tonin, M., and Vlassopoulos, M. (2018). Paying for What Kind of Performance? Performance Pay and Multitasking in Mission-Oriented Jobs. *Working Paper*.
- Kaplan, S. and Henderson, R. (2005). Inertia and Incentives: Bridging Organizational Economics and Organizational Theory. *Organization Science*, 16(5):509–521.
- Kerr, S. (1975). On the Folly of Rewarding A, While Hoping for B. *Academy of Management Journal*, 18:769–783.
- Kogut, B. and Zander, U. (1996). What Firms Do? Coordination, Identity, and Learning. *Organization Science*, 7(5):502–518.
- Lazear, E. P. (2000). Performance Pay and Productivity. *American Economic Review*, 90(5):1346–1361.
- Lazear, E. P. (2018). Compensation and Incentives in the Workplace. *Journal of Economic Perspectives*, 32(3):195–2014.
- 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.

- Lazear, E. P. and Shaw, K. L. (2007). Personnel Economics: The Economist’s View of Human Resources. *Journal of Economic Perspectives*, 21(4):91–114.
- Lee, M., Adbi, A., and Singh, J. (2020). Categorical Cognition and Outcome Efficiency in Impact Investing Decisions. *Strategic Management Journal*, 41(1):86–107.
- 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.
- Lounsbury, M. (2007). A Tale of Two Cities: Competing Logics and Practice Variation in the Professionalizing Mutual Funds. *Academy of Management Journal*, 50(2):289–307.
- MacDonald, G. and Marx, L. M. (2001). Adverse Specialization. *Journal of Political Economy*, 109(4):864–899.
- Mair, J., Battilana, J., and Cardenas, J. (2012). Organizing for Society: A Typology of Social Entrepreneurial Models. *Journal of Business Ethics*, 111(3):353–373.
- Mair, J. and Martí, I. (2006). Social Entrepreneurship Research: A Source of Explanation, Prediction, and Delight. *Journal of World Business*, 41(1):36–44.
- McMullen, J. S. and Bergman, B. J. (2017). Social Entrepreneurship and the Development Paradox of Prosocial Motivation: A Cautionary Tale. *Strategic Entrepreneurship Journal*, 11(3):243–270.
- Miller, T. L., Grimes, M. G., McMullen, J. S., and Vogus, T. J. (2012). Venturing for Others With Heart and Head: How Compassion Encourages Social Entrepreneurship. *Academy of Management Review*, 37(4):616–640.
- Moulick, A. G., Alexiou, K., Kennedy, E. D., and Linda, D. (2019). A Total Eclipse of the Heart: Compensation Strategies in Entrepreneurial Nonprofits. *Journal of Business Venturing*, forthcoming.
- Mullinix, K. J., Leeper, T. J., Druckman, J. N., and Freese, J. (2015). The Generalizability of Survey Experiments. *Working Paper*.
- Murdock, K. (2002). Intrinsic Motivation and Optimal Incentive Contracts. *RAND Journal of Economics*, 33(4):650–671.
- Nellas, V. and Reggiani, T. (2015). Multitasking in Motivational Jobs. *Working Paper*.
- Newman, A., Mayson, S., Teicher, J., and Barrett, R. (2018). Recruiting, Managing and Rewarding Workers in Social Enterprises. *International Journal of Human Resource Management*, 29(20):2851–2860.
- Ocasio, W. (1997). Towards an Attention-Based View of the Firm. *Strategic Management Journal*, 18:187–206.
- 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.
- Pache, A.-C. and Santos, F. (2013). Inside the Hybrid Organization: Selective Coupling as a Response to Competing Institutional Logics. *Academy of Management Journal*, 56(4):972–1001.
- Palan, S. and Schitter, C. (2018). Prolific.ac — A Subject Pool for Online Experiments. *Journal of Behavioral and Experimental Finance*, 17(1):22–27.
- Peer, E., Brandimarte, L., Samat, S., and Acquisti, A. (2017). Beyond the Turk: Alternative Platforms for Crowdsourcing Behavioral Research. *Journal of Experimental Social Psychology*, 70(1):153–163.
- 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.
- Prado, A. M., Robinson, J. A., and Shapira, Z. (2022). Serving Rural Low-Income Markets Through a Social Entrepreneurship Approach: Venture Creation and Growth. *Strategic Entrepreneurship Journal*, forthcoming.

- Ramus, T. and Vaccaro, A. (2017). Stakeholders Matter: How Social Enterprises Address Mission Drift. *Journal of Business Ethics*, 143(2):307–322.
- Renko, M. (2013). Early Challenges of Nascent Social Entrepreneurs. *Entrepreneurship: Theory and Practice*, 37(5):1045–1069.
- Romano, J. P. and Wolf, M. (2005). Stepwise Multiple Testing As Formalized Data Snooping. *Econometrica*, 73(4):1237–1282.
- Roumpi, D., Magrizos, S., and Nicolopoulou, K. (2020). Virtuous Circle: Human Capital and Human Resource Management in Social Enterprises. *Human Resource Management*, 59(5):401–421.
- Ruskin, J., Seymour, R. G., and Webster, C. M. (2016). Creating Value for Others? An Exploration of Social Entrepreneurial Motives. *Journal of Small Business Management*, 54(4):1015–1037.
- Santos, F., Pache, A.-C., and Birkholz, C. (2015). Making Hybrids Work: Aligning Business Models and Organizational Design for Social Enterprises. *California Management Review*, 57(3):36–59.
- Santos, F. M. (2012). A Positive Theory of Social Entrepreneurship. *Journal of Business Ethics*, 111(3):335–351.
- Schnedler, W. (2008). When Is It Foolish to Reward for A While Benefiting from B? *Journal of Labor Economics*, 26(4):595–619.
- Shaw, J. D. and Gupta, N. (2015). Let the Evidence Speak Again! Financial Incentives Are More Effective Than We Thought. *Human Resource Management Journal*, 25(3):281–293.
- Shepherd, D. A., Williams, T. A., and Zhao, E. Y. (2019). A Framework for Exploring the Degree of Hybridity in Social Entrepreneurship. *Academy of Management Perspectives*, 33(4):491–512.
- Short, J. C., Moss, T. W., and Lumpkin, G. (2009). Research in Social Entrepreneurship: Past Contributions and Future Opportunities. *Strategic Entrepreneurship Journal*, 3(2):161–194.
- Smith, W. K. and Besharov, M. L. (2019). Bowing before Dual Gods: How Structured Flexibility Sustains Organizational Hybridity. *Administrative Science Quarterly*, 64(1):1–44.
- Smith, W. K., Gonin, M., and Besharov, M. L. (2013). Managing Social-Business Tensions: A Review and Research Agenda for Social Enterprise. *Business Ethics Quarterly*, 23(3):407–442.
- Snowberg, E. and Yariv, L. (2018). Testing the Waters: Behavior Across Participant Pools. *NBER Working Paper*, 24781.
- Speckbacher, G. (2011). The Use of Incentives in Nonprofit Organizations. *Nonprofit and Voluntary Sector Quarterly*, 42(5):1006–1025.
- Staessens, M., Kerstens, P. J., Bruneel, J., and Cherchye, L. (2019). Data Envelopment Analysis and Social Enterprises: Analysing Performance, Strategic Orientation, and Mission Drift. *Journal of Business Ethics*, 159(2):325–341.
- Stevens, R., Moray, N., Bruneel, J., and Clarysse, B. (2015). Attention Allocation to Multiple Goals: The Case of For-Profit Social Enterprises. *Strategic Management Journal*, 36(7):1006–1016.
- Tonin, M. and Vlassopoulos, M. (2015). Corporate Philanthropy and Productivity: Evidence from an Online Real Effort Experiment. *Management Science*, 61(8):1795–1811.
- Tracey, P. and Jarvis, O. (2006). An Enterprising Failure. Why a Promising Social Franchise Collapsed. *Stanford Social Innovation Review*, 4(1):66–70.
- Tracey, P., Phillips, N., and Jarvis, O. (2011). Bridging Institutional Entrepreneurship and the Creation of New Organizational Forms: A Multilevel Model. *Organization Science*, 22(1):60–80.
- Wolffs, S. (2018). Employee Incentives in Microfinance Institutions: Examining the Importance of Diversification and Profit Status. *Academy of Management Best Paper Proceedings*, 2018(1):186–190.
- Wry, T. and Zhao, E. Y. (2018). Taking Trade-offs Seriously: Examining the Contextually Contingent Relationship Between Social Outreach Intensity and Financial Sustainability in Global Microfinance. *Organization Science*, 29(3):507–528.
- York, J. G., Neil, I. O., and Sarasvathy, S. D. (2016). Exploring Environmental Entrepreneurship: Identity Coupling, Venture Goals, and Stakeholder Incentives. *Journal of Management Studies*, 53(5):695–737.

- Young, A. (2019). Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results. *Quarterly Journal of Economics*, 134(2):557–598.
- Zahra, S. A., Gedajlovic, E., Neubaum, D. O., and Shulman, J. M. (2009). A Typology of Social Entrepreneurs: Motives, Search Processes and Ethical Challenges. *Journal of Business Venturing*, 24(5):519–532.

Figures and Tables

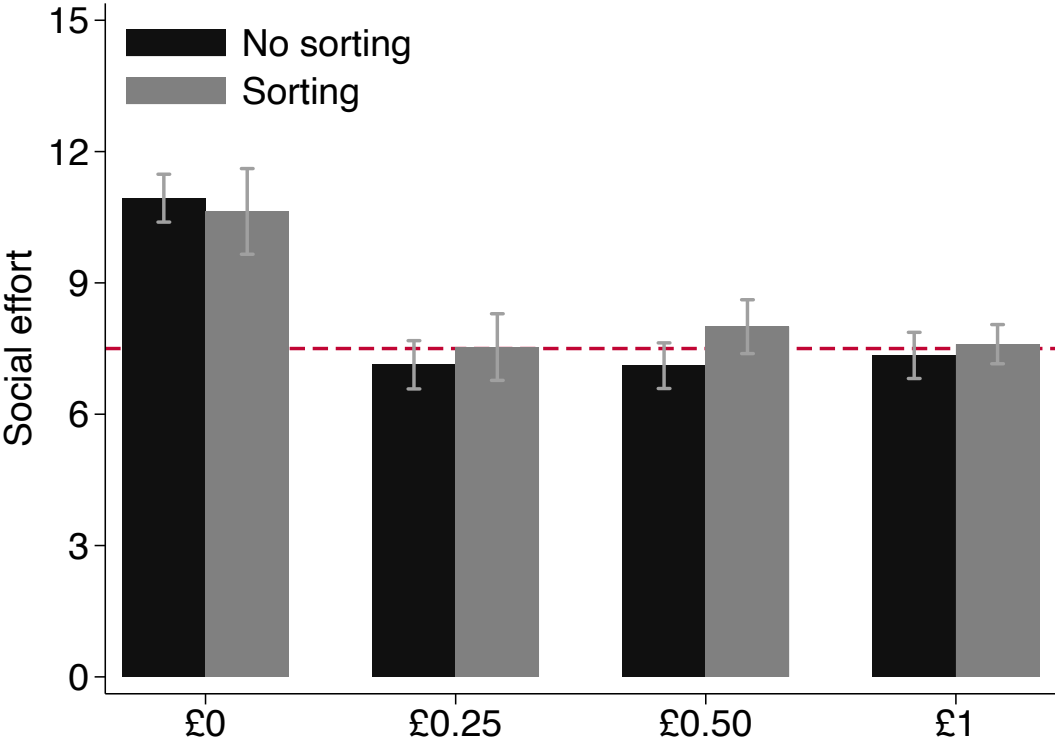


Figure 1: Social effort in social enterprises by treatment and sorting condition, with 95% confidence intervals; the dashed line represents a fully balanced effort allocation.

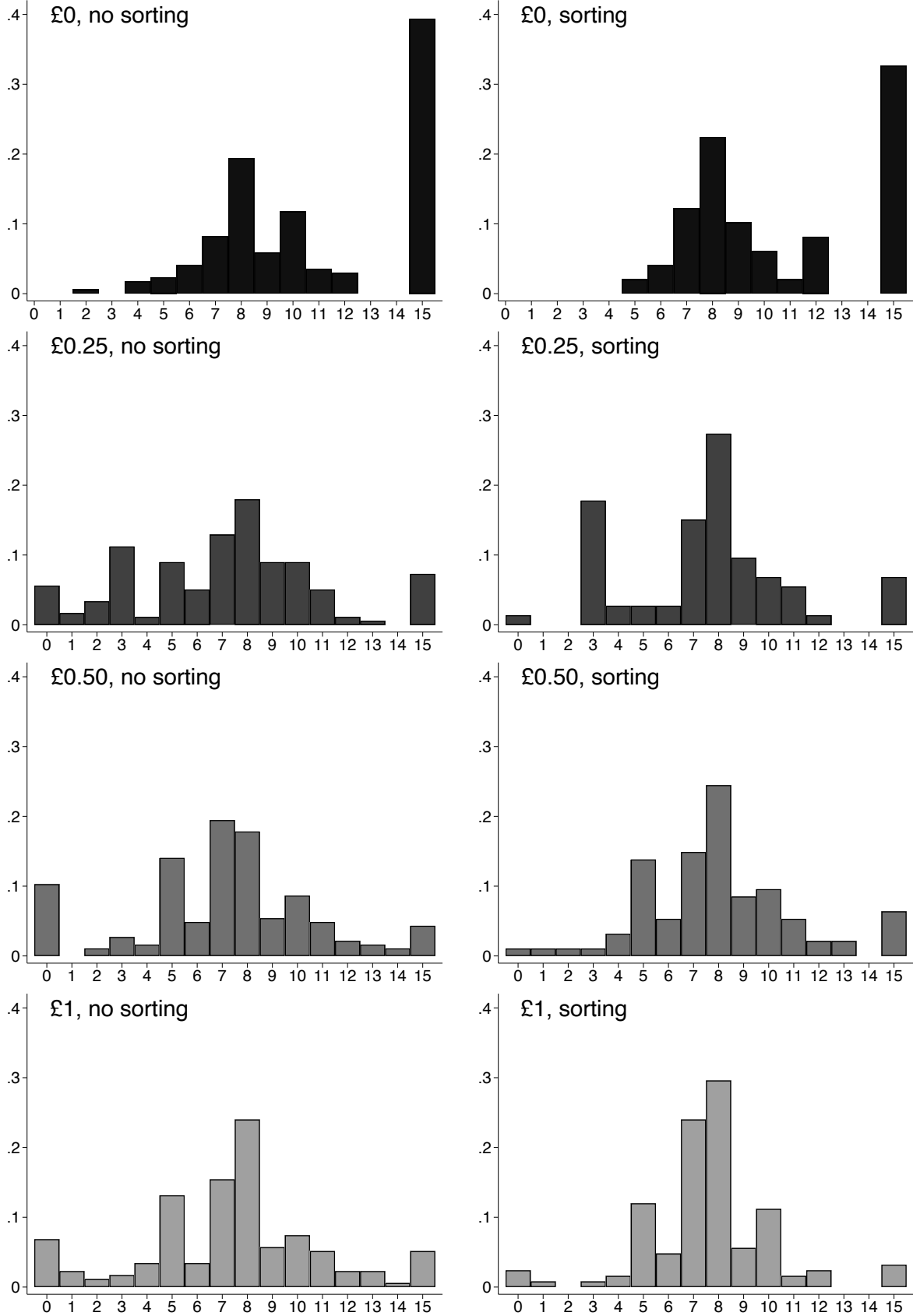


Figure 2: Distribution of social effort in social enterprises. The £0 distribution is different from the £0.25, £0.50, and £1 distributions (Kolmogorov-Smirnov $p < 0.001$, no sorting; $p < 0.01$, sorting). The latter distributions are not different from each other ($p > 0.1$).

Table 1: **Experimental Design**

	Actions and measures	Own payoff (commercial task)	Good cause payoff (social task)
Part 1	Choose good cause Practice slider task Comprehension check		
Part 2	For-profit contract	£1	£1
	Nonprofit contract	£0	£1
	Social enterprise contract	£0/£0.25/£0.50/£1	£1
Part 3	Choose preferred contract		
	For-profit contract (if selected)	£1	£1
	Nonprofit contract (if selected)	£0	£1
	Social enterprise contract (if selected)	£0/£0.25/£0.50/£1	£1
Part 4	Demographics		
	Social preferences		
	<i>Compassion</i> : sub-scale of public service motivation scale		
	<i>Altruism</i> : incentivized £10 dictator (giver)		
	<i>Inequality aversion</i> : hypothetical £10 ultimatum (receiver)		
	<i>Hypothetical altruism</i> : hypothetical £1,000 lottery		
	<i>Willingness to share</i> : without expecting anything in return		
	<i>Prosocial behavior</i> : observed prosocial behavior		
	Risk and time preferences		
	Attention and manipulation checks		
Recruitment: Prolific Academic, UK-based online platform (link)			
Stratification: by gender, for comparison purposes			
Target number of subjects: 800 in total, 200 per bonus level, 100 per gender × bonus level			
Restrictions: UK resident, age 18-64, active labor force (not homemaker, disabled, retired)			
Prior approval rate: > 90%, to ensure high-quality answers			
Participation fee: £3 for 15-20 minutes			
Bonuses: 5% or 1 in 20 subjects, up to £80, from slider task and dictator (giver/receiver)			
Good causes: The Big Issue Foundation, Fairtrade Foundation, Water Aid			
Preregistration: Open Science Framework (anonymous link)			

We randomize the order of Part 2 contracts and Part 3 options. The good cause choice determines the charity receiving the good cause payoffs and the fictional (social) enterprises' mission, i.e. workforce reintegration, fair and equitable trade, and water quality and environment. Each subject is randomly allocated a social enterprise incentive level (£0, £0.25, £0.50, or £1) as they enter the experiment and is only exposed to this incentive throughout. Subjects never see the labels used (for-profit, nonprofit, social enterprise), and only infer company type from the description.

Table 2: **Descriptive Statistics**

	Mean	Std. dev.	<i>N</i>	Min	Max
A. Demographics					
Female	0.499	(0.500)	708	0	1
Student	0.165	(0.372)	708	0	1
Education:					
High school	0.359				
Bachelor degree	0.398				
Master degree	0.127				
Income:					
< £10,000	0.215				
£10,000 – £25,000	0.329				
£25,000 – £50,000	0.307				
£50,000 – £75,000	0.077				
> £75,000	0.025				
Age:					
18 – 24	0.216				
25 – 34	0.356				
35 – 44	0.226				
45 – 54	0.140				
55 – 64	0.062				
B. Social preferences					
Compassion	29.195	(4.969)	708	12	40
Altruism	4.207	(2.181)	708	0	10
Inequality aversion	2.698	(2.275)	708	0	10
Hypothetical altruism	134.859	(159.739)	708	0	1,000
Willingness to share	6.404	(2.354)	708	0	10
Non-profit employment	0.244	(0.430)	708	0	1
Social enterprise employment	0.105	(0.306)	708	0	1
Volunteer	0.226	(0.419)	708	0	1
Donate	0.520	(0.500)	708	0	1
Work with social organization	0.160	(0.366)	708	0	1
Prosocial behavior	1.254	(1.159)	708	0	5
Social motivation (factor)	0.000	(0.757)	708	-2.450	2.307
Compassion (standardized)	0.000	(1.000)	708	-3.460	2.174
Risk taking	5.130	(2.474)	708	0	10
Time discounting	6.532	(2.193)	708	0	10
C. Experimental parameters					
Practice time	40.049	(31.011)	708	0.000	608.147
Comprehension check time	67.419	(37.004)	708	18.617	300.324
Questions time	28.359	(23.547)	708	8.341	280.372
Experiment time	1,111	(351.909)	708	600	2,399
Own payoff (£)	28.894	(12.342)	40	3	60
Good cause payoff (£)	33.025	(12.305)	40	6	60

Times given in seconds. Education and income coded as ‘other’ for 11.58% and 4.66% of subjects.

Table 3: Social Effort, by Contract and Treatment

	No sorting				Sorting			
	(1) £0	(2) £0.25	(3) £0.50	(4) £1	(5) £0	(6) £0.25	(7) £0.50	(8) £1
A. Social effort levels across contracts								
For-profit	4.918 (4.506)	5.152 (4.479)	5.227 (4.435)	5.640 (4.613)	3.956 (3.836)	4.131 (3.735)	4.100 (3.672)	4.846 (4.846)
<i>N</i>	170	178	185	175	90	84	70	26
Social enterprise	10.935 (3.612)	7.129 (3.727)	7.108 (3.595)	7.343 (3.534)	10.633 (3.408)	7.534 (3.262)	8.000 (3.002)	7.600 (2.527)
<i>N</i>	170	178	185	175	49	73	94	125
Nonprofit	13.306 (2.939)	12.719 (3.169)	12.984 (3.303)	12.697 (3.503)	13.355 (2.537)	13.762 (2.364)	13.367 (3.851)	13.458 (3.413)
<i>N</i>	170	178	185	175	31	21	21	24
B. Social enterprise social effort <i>t</i>-tests of equality of means across treatments, <i>p</i>-values								
vs £0		0.000	0.000	0.000		0.000	0.000	0.000
vs £0.25			0.956	0.581			0.340	0.874
vs £0.50				0.533				0.286
C. Social enterprise social effort <i>t</i>-tests of equality with for-profit and nonprofit, <i>p</i>-values								
vs for-profit	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
vs nonprofit	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
D. Social enterprise social effort <i>t</i>-tests of equality with for-profit and nonprofit average, sign and <i>p</i>-values								
	+0.000	-0.000	-0.000	-0.000	+0.001	-0.006	-0.118	-0.013

Standard deviations in parentheses. We use matched pair *t*-tests for the no sorting condition, since all individuals performed the slider task in all contract types. In the bottom row, we compare social enterprise social effort with the average of for-profit and nonprofit social effort to show where the social enterprise lies on the continuum between for-profit and nonprofit: ‘+’ means the social enterprise is closer to the nonprofit than to the for-profit, and ‘-’ means the social enterprise is closer to the for-profit than to the nonprofit. All *p*-values are two-sided.

Table 4: Motivation and Contract Choice

	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
A. Motivation across contracts, without sorting								
All	0.038	-0.025	0.050	-0.065	0.020	-0.036	0.049	-0.035
	(0.953)	(1.015)	(1.086)	(0.937)	(0.752)	(0.768)	(0.827)	(0.670)
<i>N</i>	170	178	185	175	170	178	185	175
B. Motivation across contracts, with sorting								
For-profit	-0.185	-0.255	-0.266	-0.318	-0.237	-0.282	-0.254	-0.293
	(0.949)	(1.088)	(1.159)	(0.989)	(0.768)	(0.745)	(0.871)	(0.763)
<i>N</i>	90	84	70	26	90	84	70	26
Social enterprise	0.272	0.150	0.209	-0.031	0.292	0.196	0.222	-0.046
	(0.863)	(0.902)	(0.970)	(0.920)	(0.637)	(0.705)	(0.742)	(0.624)
<i>N</i>	49	73	94	125	49	73	94	125
Nonprofit	0.311	0.287	0.392	0.036	0.333	0.145	0.284	0.303
	(0.965)	(0.917)	(1.114)	(0.960)	(0.605)	(0.796)	(0.763)	(0.682)
<i>N</i>	31	21	21	24	31	21	21	24
C. Social enterprise motivation <i>t</i>-tests of equality of means across treatments, <i>p</i>-values								
vs £0		0.458	0.699	0.047		0.443	0.575	0.002
vs £0.25			0.692	0.177			0.815	0.012
vs £0.50				0.063				0.004
D. Social enterprise motivation <i>t</i>-tests of equality with for-profit and nonprofit, <i>p</i>-values								
vs for-profit	0.006	0.013	0.005	0.156	0.000	0.000	0.000	0.080
vs nonprofit	0.854	0.547	0.449	0.745	0.782	0.777	0.736	0.014
E. Social enterprise motivation <i>t</i>-tests of equality with for-profit and nonprofit average, sign and <i>p</i>-values								
	+0.147	+0.320	+0.319	-0.506	+0.025	+0.025	+0.068	-0.653

Standard deviations in parentheses. We standardize *Compassion* to have a mean of 0 and a standard deviation of 1. In the bottom row, we compare social enterprise motivation with the average of for-profit and nonprofit motivation to show where the social enterprise lies on the continuum between for-profit and nonprofit: ‘+’ means the social enterprise is closer to the nonprofit than to the for-profit, and ‘-’ means the social enterprise is closer to the for-profit than to the nonprofit. All *p*-values are two-sided.

Table 5: Incentive and Selection Channel Effects

	Social effort units		Social effort share		Absolute balance	
	(1)	(2)	(3)	(4)	(5)	(6)
£0.25	-3.806 (0.394)	-3.934 (0.397)	-0.252 (0.026)	-0.260 (0.026)	-2.197 (0.587)	-2.389 (0.587)
£0.50	-3.827 (0.383)	-3.934 (0.389)	-0.254 (0.026)	-0.261 (0.026)	-2.570 (0.589)	-2.823 (0.591)
£1	-3.592 (0.385)	-3.670 (0.392)	-0.237 (0.026)	-0.241 (0.026)	-2.717 (0.591)	-2.885 (0.600)
Sorting	-0.303 (0.471)	-0.240 (0.464)	-0.020 (0.031)	-0.017 (0.031)	-0.935 (0.838)	-0.856 (0.822)
Sorting × £0.25	0.708 (0.588)	0.639 (0.586)	0.047 (0.039)	0.043 (0.039)	0.005 (0.972)	0.177 (0.948)
Sorting × £0.50	1.195 (0.577)	1.128 (0.572)	0.081 (0.038)	0.077 (0.038)	0.047 (0.930)	0.005 (0.913)
Sorting × £1	0.560 (0.537)	0.489 (0.531)	0.034 (0.036)	0.030 (0.035)	-0.778 (0.904)	-0.837 (0.886)
Constant	10.935 (0.277)	11.545 (0.631)	0.729 (0.018)	0.770 (0.042)	7.894 (0.467)	7.899 (0.929)
Tests of joint significance of Sorting + Sorting × treatment, p-values:						
£0.25	0.250	0.266	0.251	0.274	0.059	0.153
£0.50	0.008	0.008	0.007	0.007	0.027	0.034
£1	0.321	0.344	0.408	0.436	0.000	0.000
Controls	No	Yes	No	Yes	No	Yes
N	1,049	1,049	1,049	1,049	1,049	1,049
R^2	0.152	0.177	0.151	0.175	0.066	0.113

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Subject-clustered standard errors in parentheses. The baseline category is the £0 treatment when sorting is not possible. The dependent variable is social enterprise *Social effort* in columns (1)-(2), social enterprise social effort as share of total effort in columns (3)-(4), and absolute balance (absolute value of the difference between social and commercial effort) in columns (5)-(6). Controls include age, gender, studentship, education, income, risk and time preferences, and good cause choice. These models include 708 observations where sorting is not possible and 341 where it is possible (for individuals who select into the social enterprise contract).

ONLINE APPENDIX

(not for publication)

“Revenue Drift, Incentives, and Effort Allocation in Social Enterprises”

Theodor Vladasel, Universitat Pompeu Fabra and Barcelona School of Economics
Simon C. Parker, Ivey Business School, Western University, University of Aberdeen, and IZA
Randolf Sloof, Universiteit van Amsterdam and Tinbergen Institute
Mirjam van Praag, Vrije Universiteit Amsterdam, CBS, CEPR, IZA, and Tinbergen Institute

This online appendix includes additional material for our paper entitled “Revenue Drift, Incentives, and Effort Allocation in Social Enterprises”. We provide the full experimental instructions including company and task descriptions, payoff explanations, social preference measures, and other demographic variables in Appendix A. Appendix B then provides a set of experimental design checks, including an analysis of good cause choice, a randomization check, and a randomization inference-based computation of standard errors. Appendix C details the results of power calculations performed prior to running the experiment, as well as minimum detectable effect sizes (ex post). Appendix D shows that results are robust to tightening or relaxing the sample restrictions and discusses the results of our pilot experiment, Appendix E details our common factor analysis and examines robustness with regards to alternative social preference measures, and Appendix F considers a set of alternative effort measures and regression models. Appendix G discusses the results of multiple hypothesis testing adjustments, reflecting the number of outcomes and treatments we consider, and performs heterogeneity analyses with regards to gender, previous social sector experience, and mission choice. Finally, Appendix H 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

Appendix A Instructions	45
Appendix B Experimental Checks	53
Appendix C Power Calculations	58
Appendix D Robustness Check: Different Samples	62
Appendix E Robustness Check: Social Preferences	66
Appendix F Robustness Check: Effort Measures	73
Appendix G Multiple Hypothesis Testing and Heterogeneity	79
Appendix H Additional Study	85

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.¹⁸

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.

¹⁸ 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.¹⁹

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.

¹⁹ 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 B 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 D.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 B.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 B.1, by treatment. Table B.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 G).

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 B.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 B.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 B.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 B.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	RI	EXP	RI	EXP	RI	EXP	RI	EXP	RI	EXP	RI
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(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 B.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 B.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)
<i>N</i>	708			
LR χ^2 (<i>p</i> -value)	85.25 (0.018)			
Pseudo- <i>R</i> ²	0.061			
Social preferences: χ^2 (<i>p</i> -value)	17.18 (0.374)			
Treatment levels: χ^2 (<i>p</i> -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 B.5: **Randomization Check**

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

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. 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 C 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:

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

Equation C.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 C.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 $0.2 < \delta/\sigma < 1$.²⁰ 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.²¹ 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 C.1, where we calculate the

²⁰ 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.

²¹ The latter groups are smaller, but their mean difference is expected to be large, so power is retained.

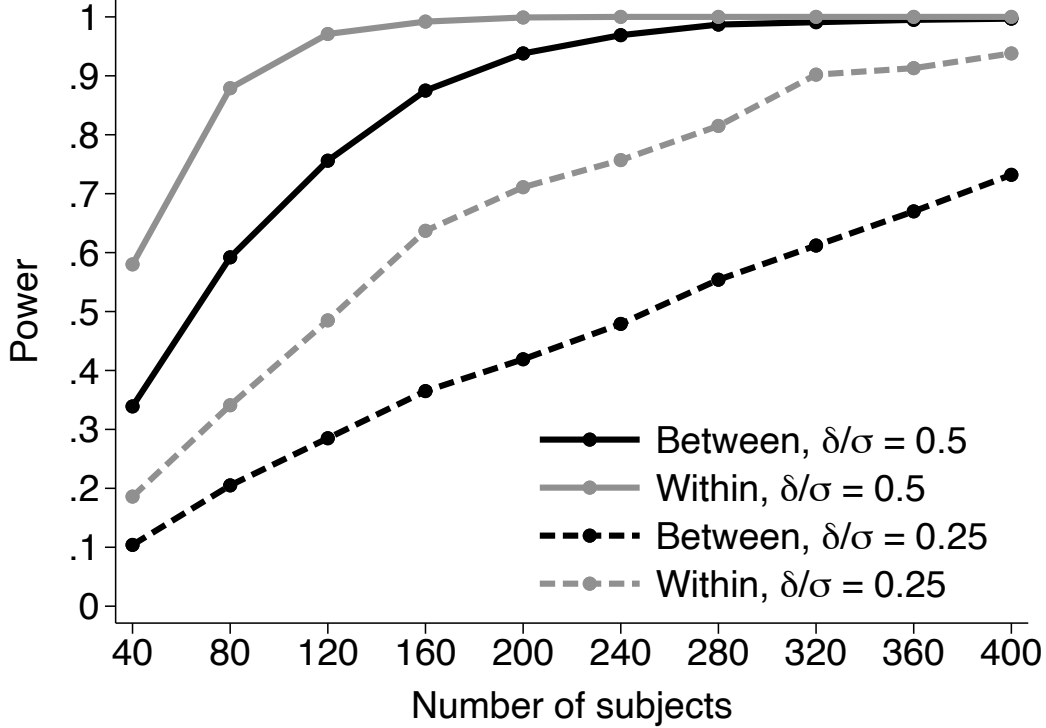


Figure C.1: Power levels for two-sided mean comparisons between- and within-individuals ($\alpha = 0.05, \sigma_a = \sigma_b = \sigma = 2, \delta = 1$ or $0.5, \delta/\sigma = 0.5$ or $0.25, H_0: \mu_a = \mu_b, H_a: \mu_a \neq \mu_b$).

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 C.2 confirms the power calculations in Table C.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 C.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.²²

²² 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.

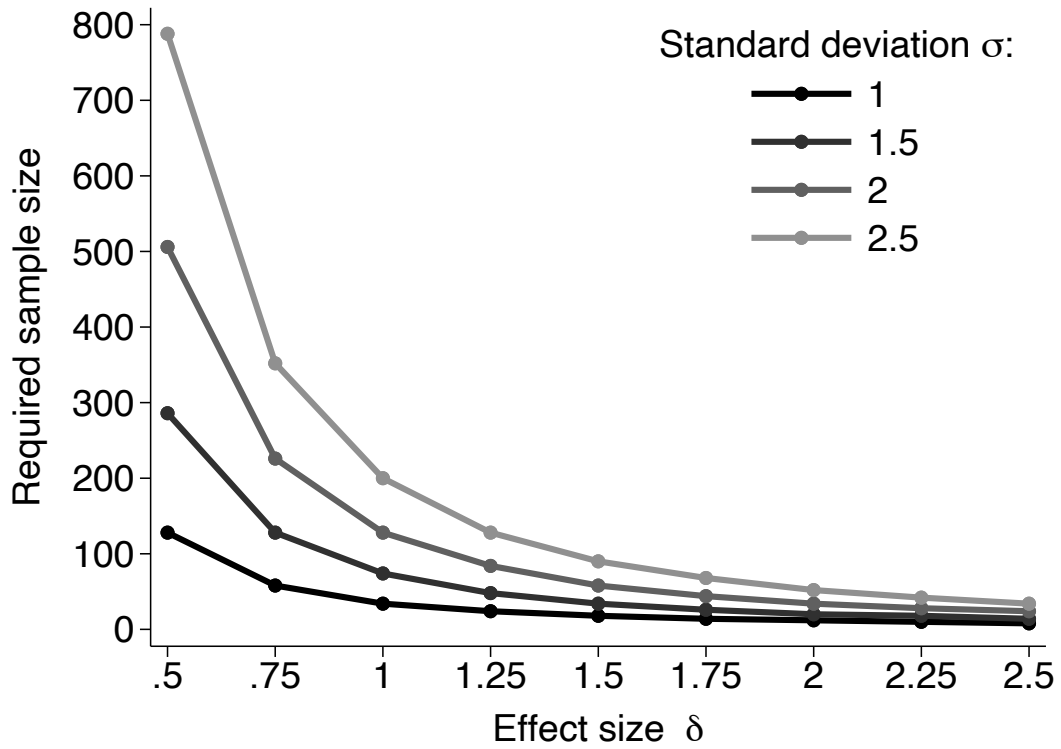


Figure C.2: Required sample size for detecting effect size δ with 80% power in two-sided mean comparison ($1 - \beta = 0.8, \alpha = 0.05, \sigma_a = \sigma_b = \sigma, H_0: \mu_a = \mu_b, H_a: \mu_a \neq \mu_b$).

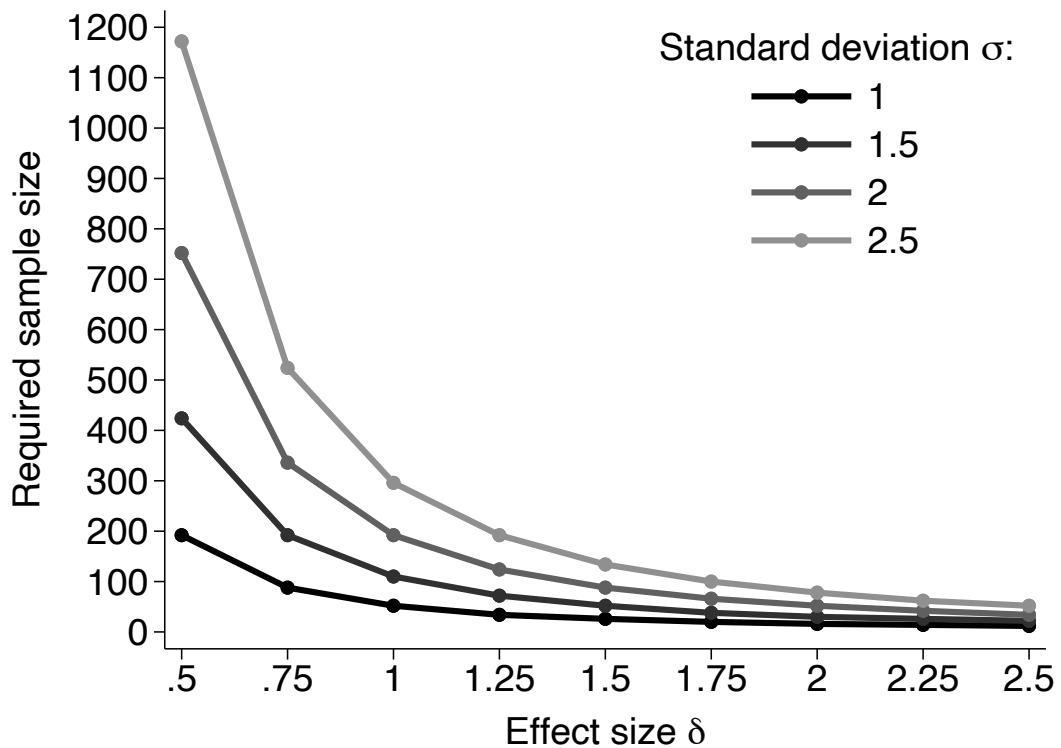


Figure C.3: Required sample size for detecting effect size δ with 80% power in two-sided mean comparison ($1 - \beta = 0.8, \alpha = 0.01, \sigma_a = \sigma_b = \sigma, H_0: \mu_a = \mu_b, H_a: \mu_a \neq \mu_b$).

Table C.1: Power Calculations

	N_a	N_b	μ_a	μ_b	σ	δ/σ	α	2-sided $1 - \beta$	1-sided $1 - \beta$	$\delta_{min}^{80\%}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A.	200	200	7.5	8	2.5	0.20	0.05	0.514	0.638	0.700
Without sorting	200	200	7.5	8.5	2.5	0.40	0.05	0.979	0.991	
	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 sorting (£0.50)	100	40	7.5	8.5	2.5	0.40	0.05	0.565	0.685	
	100	40	7.5	9	2.5	0.60	0.05	0.890	0.939	
	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 sorting (£1)	120	25	7.5	8.5	2.5	0.40	0.05	0.439	0.566	
	120	25	7.5	9	2.5	0.60	0.05	0.774	0.858	
	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 $\delta_{min}^{80\%}/\sigma$ 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 D 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 D.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 D.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.²³ Indeed, 95.04% of the total 47,760

²³ 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 D.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 D.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 D.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 D.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 D.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.

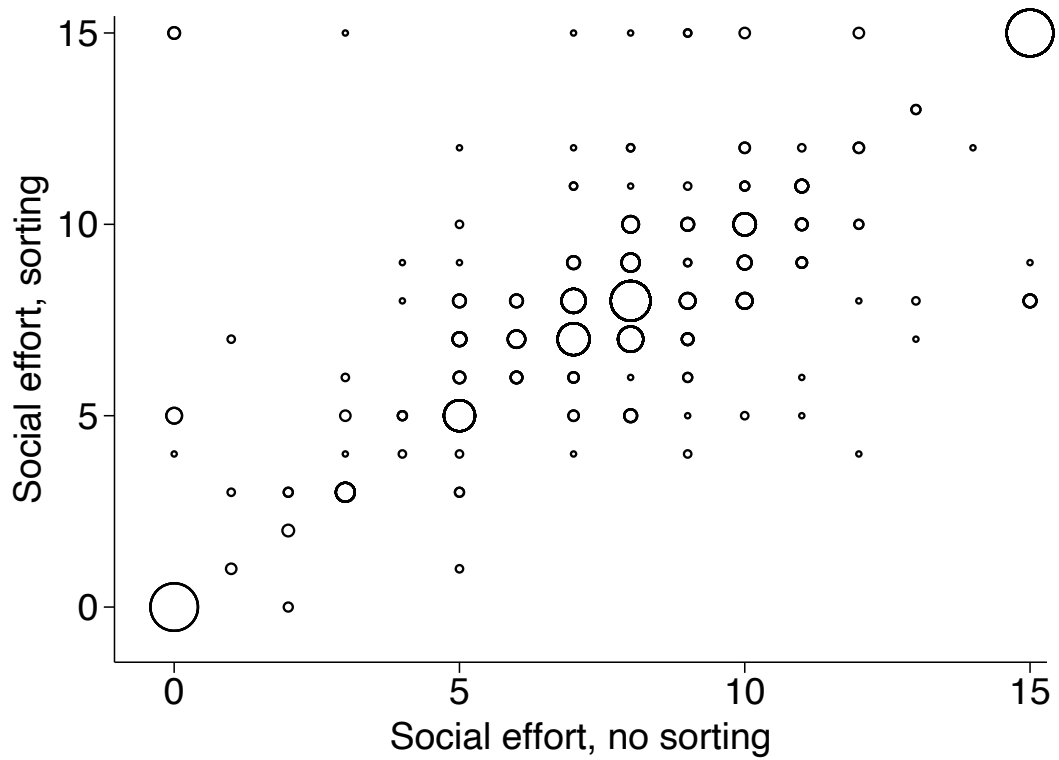


Figure D.1: Answer consistency across social enterprise contracts with and without sorting, for subjects who performed the social enterprise contract twice.

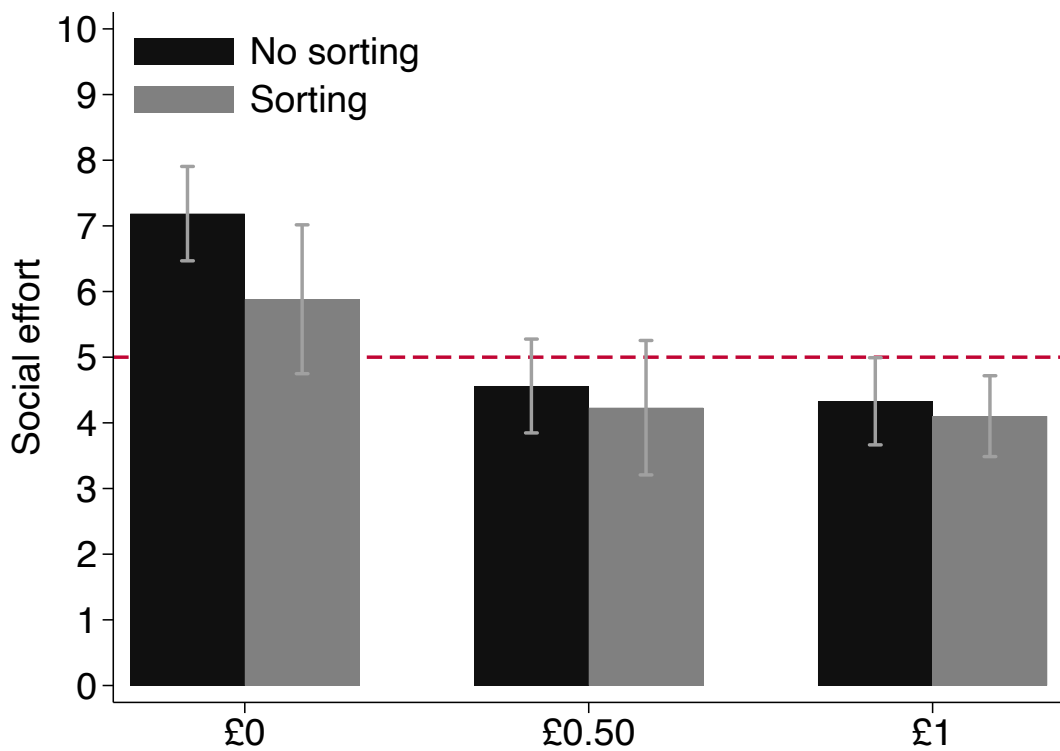


Figure D.2: Pilot data social effort in social enterprises by treatment, with 95% confidence intervals; the dashed line represents a fully balanced effort allocation.

Table D.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 (3.599)	7.099 (3.672)	7.088 (3.762)	7.338 (3.741)	10.350 (3.416)	7.390 (3.140)	7.951 (3.054)	7.504 (2.572)
Share	0.733 (0.239)	0.473 (0.244)	0.472 (0.250)	0.489 (0.249)	0.690 (0.227)	0.492 (0.209)	0.530 (0.203)	0.500 (0.171)
B. $NS = 0$ ($N = 623, N_{SE} = 307$)								
Units	10.913 (3.676)	7.238 (3.703)	7.141 (3.776)	7.463 (3.774)	10.302 (3.447)	7.405 (3.025)	7.940 (3.019)	7.459 (2.620)
Share	0.727 (0.245)	0.482 (0.246)	0.476 (0.251)	0.497 (0.251)	0.686 (0.229)	0.493 (0.201)	0.529 (0.201)	0.497 (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 (3.612)	7.088 (3.760)	7.165 (3.587)	7.323 (3.526)	10.687 (3.421)	7.362 (3.180)	8.000 (3.018)	7.590 (2.557)
Share	0.728 (0.240)	0.474 (0.249)	0.478 (0.239)	0.490 (0.234)	0.712 (0.228)	0.491 (0.210)	0.535 (0.200)	0.506 (0.170)
D. $IS \leq 10$, pass attention check ($N = 717, N_{SE} = 346$)								
Units	10.948 (3.649)	7.173 (3.757)	7.215 (3.592)	7.325 (3.633)	10.775 (3.441)	7.378 (3.073)	7.989 (2.987)	7.500 (2.682)
Share	0.729 (0.243)	0.479 (0.249)	0.481 (0.239)	0.491 (0.242)	0.718 (0.229)	0.492 (0.203)	0.534 (0.198)	0.500 (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 (3.714)	7.088 (3.743)	7.112 (3.659)	7.222 (3.599)	10.274 (3.800)	7.426 (3.333)	7.978 (2.992)	7.500 (2.632)
Share	0.727 (0.239)	0.480 (0.250)	0.474 (0.243)	0.484 (0.239)	0.706 (0.225)	0.496 (0.222)	0.539 (0.204)	0.503 (0.170)
F. $IS \leq 30$ ($N = 767, N_{SE} = 372$)								
Units	10.700 (3.827)	7.149 (3.775)	7.158 (3.635)	7.272 (3.664)	10.321 (3.785)	7.566 (3.298)	7.959 (2.949)	7.407 (2.735)
Share	0.724 (0.242)	0.488 (0.254)	0.478 (0.242)	0.487 (0.244)	0.709 (0.226)	0.505 (0.219)	0.538 (0.201)	0.497 (0.177)
G. Full sample ($N = 796, N_{SE} = 388$)								
Units	10.239 (4.296)	6.984 (3.868)	6.901 (3.790)	7.040 (3.802)	9.419 (4.550)	7.388 (3.457)	7.774 (3.101)	7.208 (2.935)
Share	0.717 (0.252)	0.493 (0.258)	0.479 (0.246)	0.489 (0.245)	0.717 (0.247)	0.505 (0.247)	0.547 (0.209)	0.500 (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 E 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 E.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 E.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 E.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 E.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 E.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 E.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 E.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.²⁴

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 E.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 E.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.

²⁴ 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.

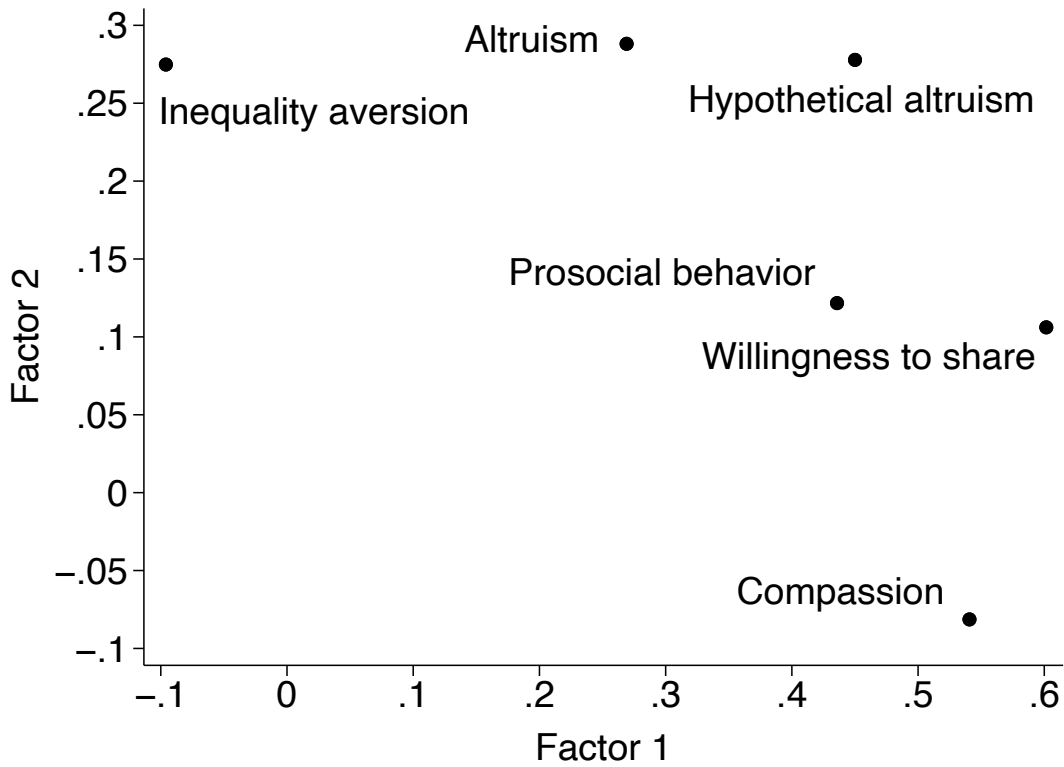


Figure E.1: Social preference loadings on the first two factors obtained from a common factor analysis with orthogonal varimax rotation.

Table E.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	<i>Social motivation</i>		

Factor loadings for a common factor analysis with an orthogonal varimax rotation; results are similar with oblique (non-orthogonal) rotations (see Table E.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.

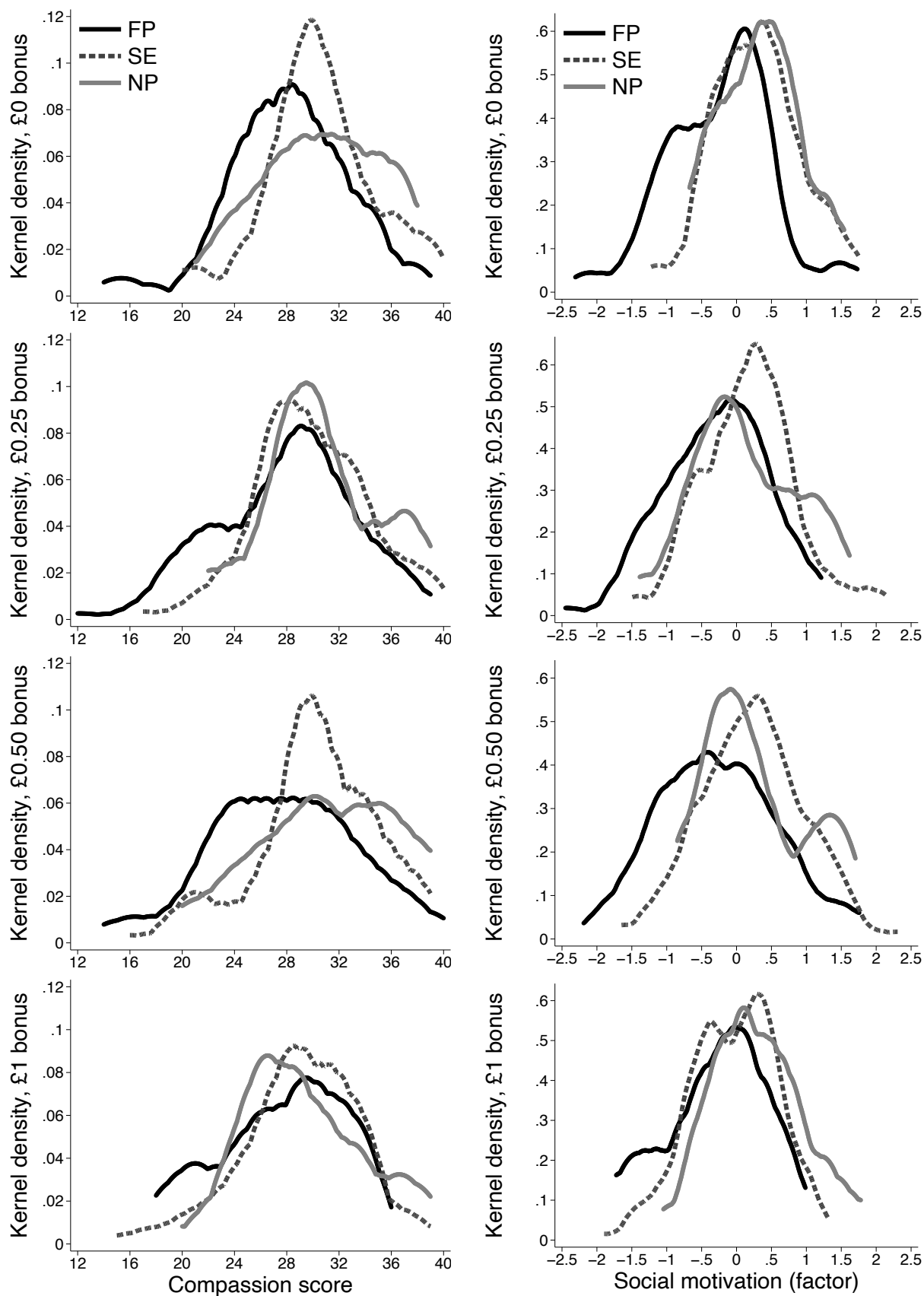


Figure E.2: Distribution of *Compassion* and *Social motivation* under sorting, by treatment and contract choice (FP = for-profit, NP = nonprofit, SE = social enterprise).

Table E.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 E.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 (0.162)	-0.074 (0.170)	0.022 (0.067)	-0.005 (0.068)	-0.005 (0.082)	0.007 (0.085)
£0.50	-0.064 (0.158)	0.030 (0.153)	0.027 (0.064)	0.001 (0.062)	0.064 (0.080)	0.106 (0.080)
£1	-0.304** (0.148)	-0.184 (0.155)	0.113* (0.064)	0.087 (0.065)	-0.001 (0.075)	0.044 (0.078)
R^2	0.016	0.143	0.013	0.112	0.004	0.110
B. Social motivation						
£0.25	-0.097 (0.123)	-0.073 (0.123)	0.144** (0.059)	0.138** (0.062)	-0.059 (0.089)	-0.039 (0.089)
£0.50	-0.070 (0.119)	-0.025 (0.117)	0.098* (0.051)	0.095* (0.054)	-0.047 (0.086)	-0.012 (0.085)
£1	-0.339*** (0.107)	-0.274*** (0.112)	0.187*** (0.052)	0.177*** (0.058)	-0.204*** (0.078)	-0.170** (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 E.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 (0.347)	-0.006 (0.261)	-0.256 (0.226)	0.080 (0.252)	0.115 (0.420)
£0.50	0.201 (0.302)	0.179 (0.232)	0.039 (0.212)	0.181 (0.222)	-0.101 (0.323)
£1	-0.115 (0.301)	0.030 (0.226)	-0.260 (0.229)	-0.121 (0.190)	-0.374 (0.318)
B. Social motivation					
£0.25	-0.106 (0.205)	-0.151 (0.181)	-0.060 (0.156)	-0.051 (0.172)	-0.045 (0.242)
£0.50	-0.239 (0.207)	-0.067 (0.182)	0.036 (0.161)	0.043 (0.182)	-0.003 (0.205)
£1	-0.317* (0.186)	-0.297* (0.177)	-0.260 (0.163)	-0.259 (0.176)	-0.359* (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 10th, 25th, 50th, 75th, and 90th 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 E.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 (0.308)	-0.316 (0.384)	14.214 (29.871)	-0.150 (0.375)	-0.291 (0.208)	-0.136 (0.437)	-0.076 (0.378)
£0.50	-0.497* (0.294)	-0.149 (0.366)	2.896 (28.500)	-0.039 (0.358)	-0.189 (0.198)	0.167 (0.417)	-0.109 (0.361)
£1	-0.762*** (0.281)	-0.199 (0.350)	-43.925 (27.262)	-0.810** (0.343)	-0.295 (0.189)	-0.043 (0.399)	-0.555 (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 E.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*** (0.163)	-0.418*** (0.159)	-0.523*** (0.164)	-0.289 (0.209)	-0.543*** (0.124)	-0.494*** (0.119)	-0.509*** (0.128)	-0.248 (0.159)
Nonprofit	0.002 (0.211)	0.109 (0.223)	0.186 (0.262)	0.076 (0.214)	0.020 (0.143)	-0.078 (0.188)	0.063 (0.179)	0.353** (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 E.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*** (0.496)	3.716*** (0.248)	2.695*** (0.236)	96.148*** (13.888)	6.015*** (0.264)	1.307** (0.137)	-0.136* (0.081)	5.086*** (0.264)	6.209*** (0.216)
Social enterprise	30.977*** (0.522)	4.822*** (0.237)	2.496*** (0.247)	143.172*** (18.353)	7.169*** (0.262)	1.534*** (0.149)	0.307*** (0.083)	5.176*** (0.280)	6.643*** (0.230)
Nonprofit	31.514*** (0.647)	4.913*** (0.251)	2.685*** (0.313)	192.515*** (22.131)	7.371*** (0.314)	1.527*** (0.153)	0.427*** (0.094)	5.044*** (0.341)	6.607*** (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 F 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.²⁵ 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 F.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 D.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 D.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 F.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 F.3. With or without sorting, there does not appear to a linear association between *Social motivation* and maximum social effort.²⁶ 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*,

²⁵ Note that we do not find differences in social enterprise total effort across treatments (as subjects could choose to move less than 15 sliders), regardless of the sample we use. Our additional study in Appendix H provides further evidence that total effort is not an important adjustment margin.

²⁶ 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 E.6. Figure F.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 F.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 F.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 F.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.

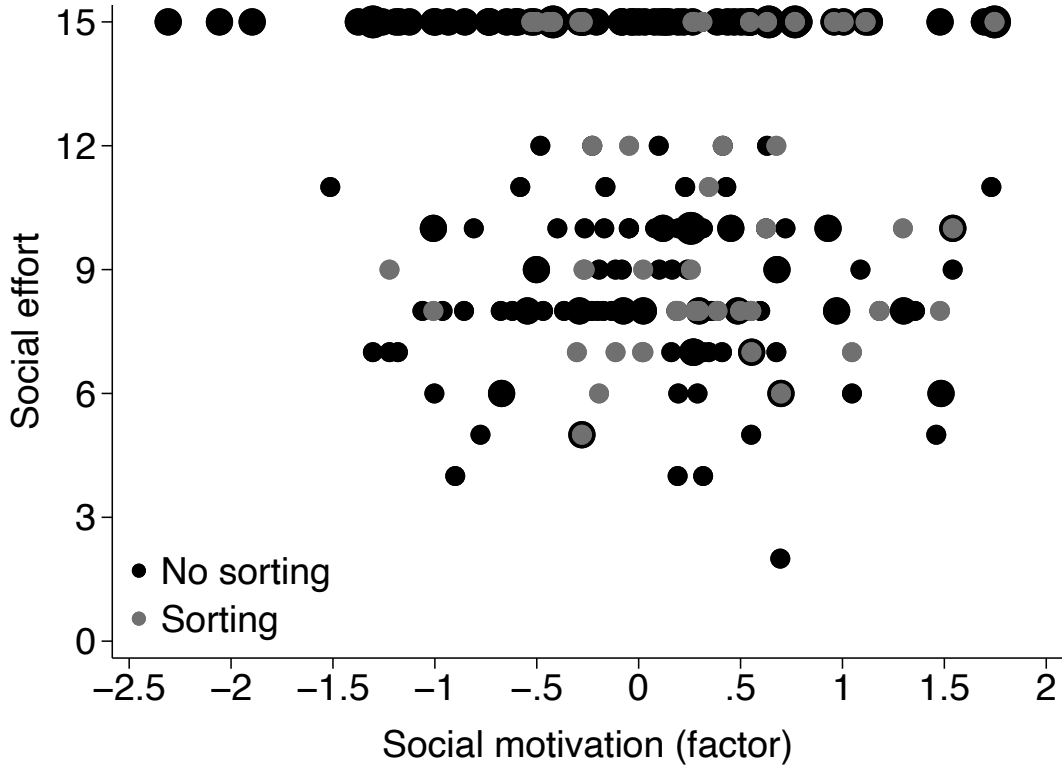


Figure F.1: Social effort in the social enterprise contract, £0 treatment.

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

	No sorting				Sorting			
	£0 (1)	£0.25 (2)	£0.50 (3)	£1 (4)	£0 (5)	£0.25 (6)	£0.50 (7)	£1 (8)
For-profit	0.328 (0.023)	0.344 (0.022)	0.349 (0.022)	0.376 (0.023)	0.266 (0.028)	0.276 (0.027)	0.273 (0.029)	0.325 (0.062)
Social enterprise	0.729 (0.018)	0.477 (0.019)	0.475 (0.018)	0.492 (0.018)	0.709 (0.032)	0.504 (0.025)	0.535 (0.021)	0.507 (0.015)
Nonprofit	0.887 (0.015)	0.854 (0.016)	0.869 (0.017)	0.854 (0.018)	0.890 (0.030)	0.917 (0.034)	0.911 (0.056)	0.897 (0.046)
Social enterprise effort <i>t</i> -tests of equality <i>p</i> -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 <i>t</i> -tests of equality with nonprofit and for-profit average, <i>p</i> -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 F.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*** (0.426)	1.978*** (0.331)	1.881*** (0.319)	1.703*** (0.336)
Nonprofit	8.388*** (0.455)	7.567*** (0.433)	7.757*** (0.425)	7.057*** (0.443)
B. Share of social effort				
Social enterprise	0.401*** (0.028)	0.133*** (0.022)	0.126*** (0.021)	0.116*** (0.022)
Nonprofit	0.559*** (0.030)	0.510*** (0.028)	0.521*** (0.028)	0.478*** (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 F.3: **Adverse Specialization and Motivation**

	No sorting			Sorting		
	(1)	(2)	(3)	(4)	(5)	(6)
A. Compassion						
Compassion	0.027 (0.047)	0.037 (0.040)		0.235** (0.085)	0.114* (0.080)	
Compassion ²		0.095*** (0.025)			0.150*** (0.050)	
Compassion $\geq 90^{\text{th}}$ pct.			0.363*** (0.134)			0.946*** (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 (0.057)	-0.076 (0.056)		0.241 (0.176)	0.145 (0.183)	
Social motivation ²		0.085* (0.047)			0.146 (0.163)	
Social motivation $\geq 90^{\text{th}}$ pct.			-0.021 (0.139)			0.573 (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 F.4: Measures of imbalance

	Absolute imbalance $ S - C $		Absolute imbalance $S - C$	
	No sorting (1)	Sorting (2)	No sorting (3)	Sorting (4)
A. Differences in balance across treatments				
£0 (benchmark)	7.894 (0.349)	6.959 (0.623)	6.871 (0.553)	6.265 (0.844)
£0.25	-2.197*** (0.551)	-2.192*** (0.805)	-7.522*** (0.774)	-6.128*** (1.091)
£0.50	-2.570*** (0.546)	-2.523*** (0.768)	-7.632*** (0.767)	-5.212*** (1.040)
£1	-2.717*** (0.553)	-3.495*** (0.735)	-7.111*** (0.777)	-6.065*** (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*** (1.449)	9.210*** (1.986)	5.805*** (1.710)	10.885*** (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 (1.706)	5.293 (3.885)	-0.461 (2.007)	6.849 (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 $|S - C|$ balance measure is eliminated when adjusting for multiple hypothesis testing.

Table F.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*** (0.509)	-4.822*** (0.509)	-0.310*** (0.034)	-0.319*** (0.034)	-3.074*** (0.749)	-3.284*** (0.742)
£0.50	-4.825*** (0.508)	-4.936*** (0.511)	-0.321*** (0.034)	-0.328*** (0.034)	-3.395*** (0.750)	-3.672*** (0.745)
£1	-4.519*** (0.504)	-4.594*** (0.508)	-0.298*** (0.034)	-0.303*** (0.034)	-3.627*** (0.748)	-3.803*** (0.752)
Sorting	-0.491 (0.645)	-0.412 (0.631)	-0.033 (0.043)	-0.028 (0.042)	-1.210 (1.069)	-1.096 (1.039)
Sorting × £0.25	0.962 (0.757)	0.876 (0.748)	0.064 (0.050)	0.059 (0.050)	0.169 (1.207)	0.346 (1.171)
Sorting × £0.50	1.572** (0.750)	1.485** (0.736)	0.106** (0.050)	0.100** (0.049)	0.102 (1.165)	0.023 (1.134)
Sorting × £1	0.791 (0.709)	0.704 (0.695)	0.050 (0.047)	0.045 (0.046)	-0.652 (1.136)	-0.734 (1.103)
Constant	11.832*** (0.402)	12.434*** (0.760)	0.789*** (0.027)	0.829*** (0.051)	9.074*** (0.640)	9.156*** (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 F.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*** (0.318)	-3.850*** (0.322)	-0.248*** (0.021)	-0.255*** (0.021)	-2.494*** (0.510)	-2.694*** (0.513)
Sorting	-0.303 (0.470)	-0.238 (0.463)	-0.020 (0.031)	-0.016 (0.031)	-0.935 (0.837)	-0.875 (0.822)
Sorting × Bonus	0.824 (0.501)	0.756 (0.495)	0.054 (0.033)	0.050 (0.033)	-0.362 (0.868)	-0.333 (0.851)
Constant	10.935*** (0.277)	11.544*** (0.630)	0.729*** (0.018)	0.769*** (0.042)	7.894*** (0.466)	7.860*** (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 G 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 G.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.²⁷ When we regress social effort on treatment dummies, gender, and their interactions in Table G.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 G.1. In Table G.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.

²⁷ 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.

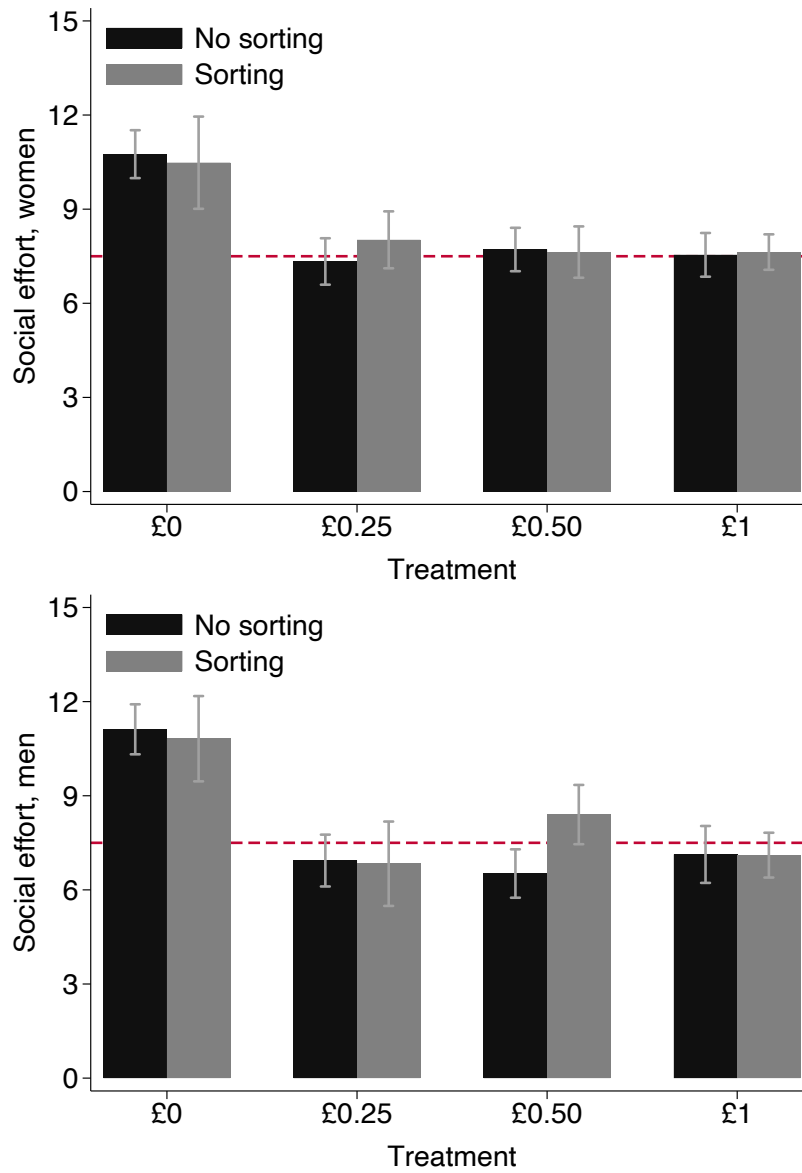


Figure G.1: Social effort in social enterprises by treatment for women (top) and men (bottom), with 95% confidence intervals; the dashed line represents a fully balanced effort allocation.

Mission heterogeneity Because social preferences may be weakly correlated with good cause choice (Table B.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 G.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.

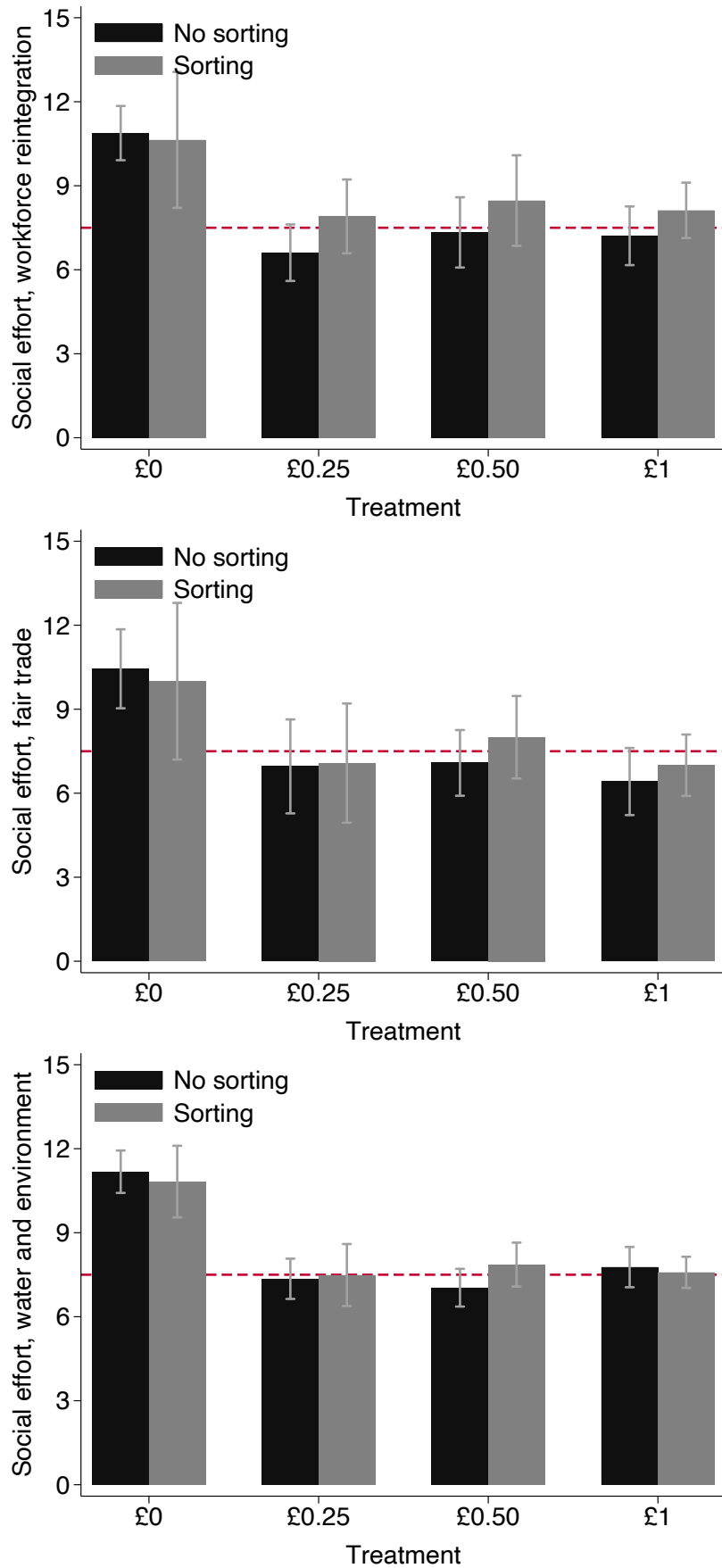


Figure G.2: Social effort in social enterprises by treatment and mission, with 95% confidence intervals; the dashed line represents a fully balanced effort allocation.

Table G.1: Multiple Hypothesis Testing Adjustments

Comparison			<i>p</i> -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 G.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*** (0.577)	-4.296*** (0.593)	-0.276*** (0.038)	-0.284*** (0.039)	-3.985*** (0.919)	-4.539*** (0.913)	-0.263*** (0.061)	-0.300*** (0.061)
£0.50	-4.596*** (0.558)	-4.651*** (0.562)	-0.306*** (0.037)	-0.309*** (0.037)	-2.418*** (0.798)	-2.924*** (0.814)	-0.161*** (0.053)	-0.195*** (0.054)
£1	-3.988*** (0.571)	-3.984*** (0.582)	-0.264*** (0.038)	-0.263*** (0.039)	-3.711*** (0.738)	-4.114*** (0.758)	-0.247*** (0.049)	-0.274*** (0.051)
Female	-0.365 (0.555)	-0.370 (0.587)	-0.024 (0.037)	-0.025 (0.039)	-0.337 (0.960)	-0.517 (0.956)	-0.022 (0.064)	-0.034 (0.064)
Female × £0.25	0.764 (0.787)	0.814 (0.806)	0.049 (0.052)	0.053 (0.054)	1.527 (1.246)	1.816 (1.211)	0.101 (0.083)	0.119 (0.081)
Female × £0.50	1.558** (0.762)	1.572** (0.772)	0.105** (0.051)	0.106** (0.051)	-0.431 (1.143)	-0.374 (1.113)	-0.025 (0.076)	-0.022 (0.074)
Female × £1	0.780 (0.772)	0.741 (0.787)	0.054 (0.051)	0.051 (0.052)	1.230 (1.063)	1.176 (1.054)	0.082 (0.071)	0.078 (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 G.3: **Gender Differences in Social Preferences**

	Compassion (standardized)		Social Motivation	
	(1)	(2)	(3)	(4)
£0.25	-0.334 (0.255)	-0.256 (0.247)	-0.189 (0.198)	-0.108 (0.200)
£0.50	-0.276 (0.253)	-0.208 (0.232)	-0.134 (0.194)	-0.107 (0.188)
£1	-0.549** (0.236)	-0.423* (0.226)	-0.464*** (0.170)	-0.383** (0.171)
Female	-0.064 (0.251)	-0.080 (0.245)	-0.035 (0.185)	-0.015 (0.184)
Female × £0.25	0.365 (0.329)	0.327 (0.328)	0.159 (0.252)	0.066 (0.252)
Female × £0.50	0.403 (0.321)	0.436 (0.304)	0.121 (0.242)	0.151 (0.236)
Female × £1	0.444 (0.301)	0.433 (0.293)	0.226 (0.217)	0.199 (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 H 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 H.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 H.2 are similar to those in our main study in Table B.1 and subject characteristics effectively pass a randomization check. For each covariate in Table H.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 H.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

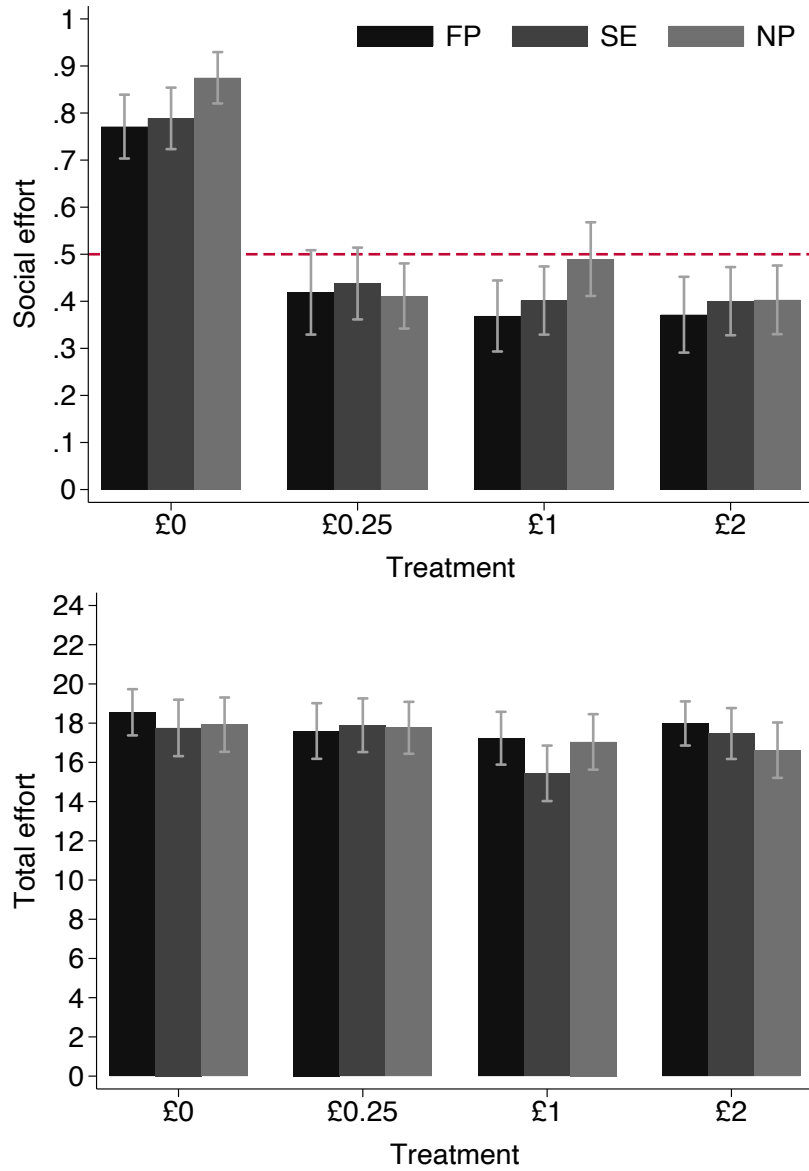


Figure H.1: Social effort as share of total effort (top) and total effort (bottom), with 95% confidence intervals; the dashed line represents a fully balanced effort allocation.

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 H.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 H.1's top panel and in column (8) of Table H.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 H.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 H.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 H.5. Moreover, when we study the effect of incentives on social effort through OLS regressions in column (7) of Table H.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 H.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 H.1.

We investigate differences across organizational forms more formally in Panel A of Table H.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 H.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.

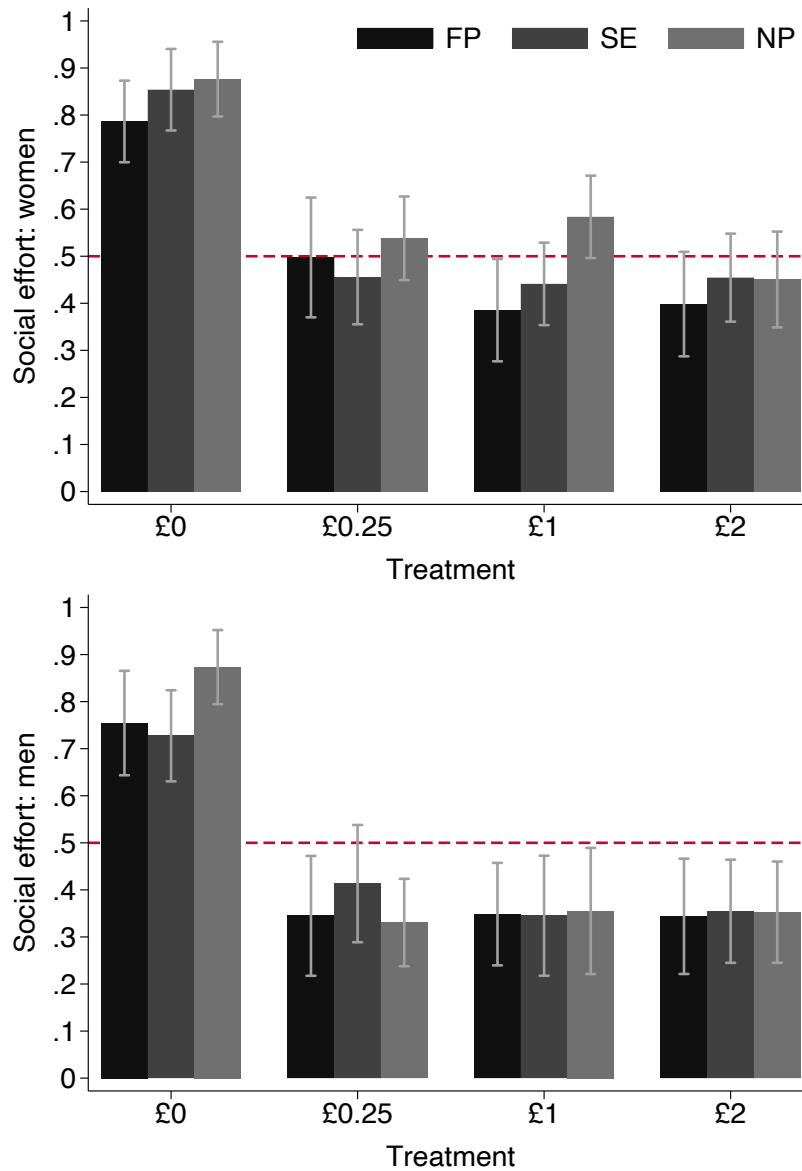


Figure H.2: Social effort for women (top) and men (bottom), with 95% confidence intervals; the dashed line represents a fully balanced effort allocation.

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.

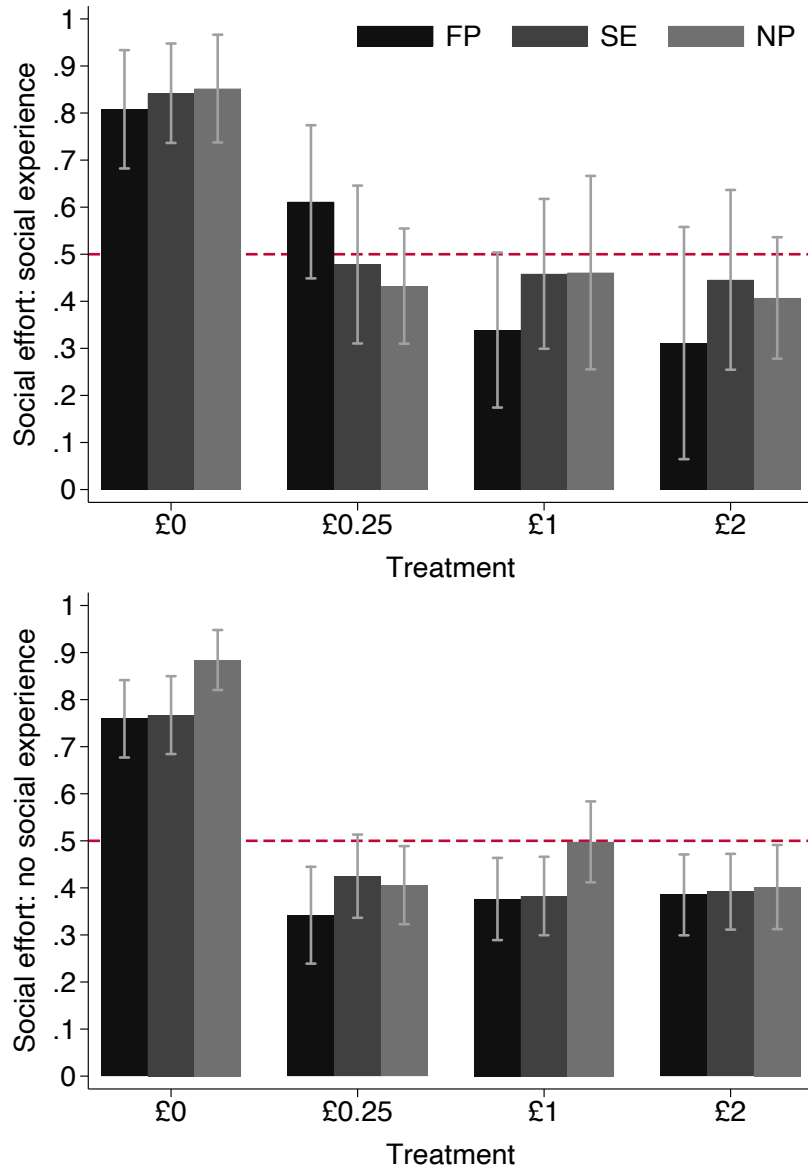


Figure H.3: Social effort for subjects with (top) and without (bottom) social sector experience, with 95% confidence intervals; the dashed line represents a fully balanced effort allocation.

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 H.2 and Table H.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 H.3 shows that subjects with such experience exert nominally higher social effort at all incentive levels, especially high-powered, but Table H.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

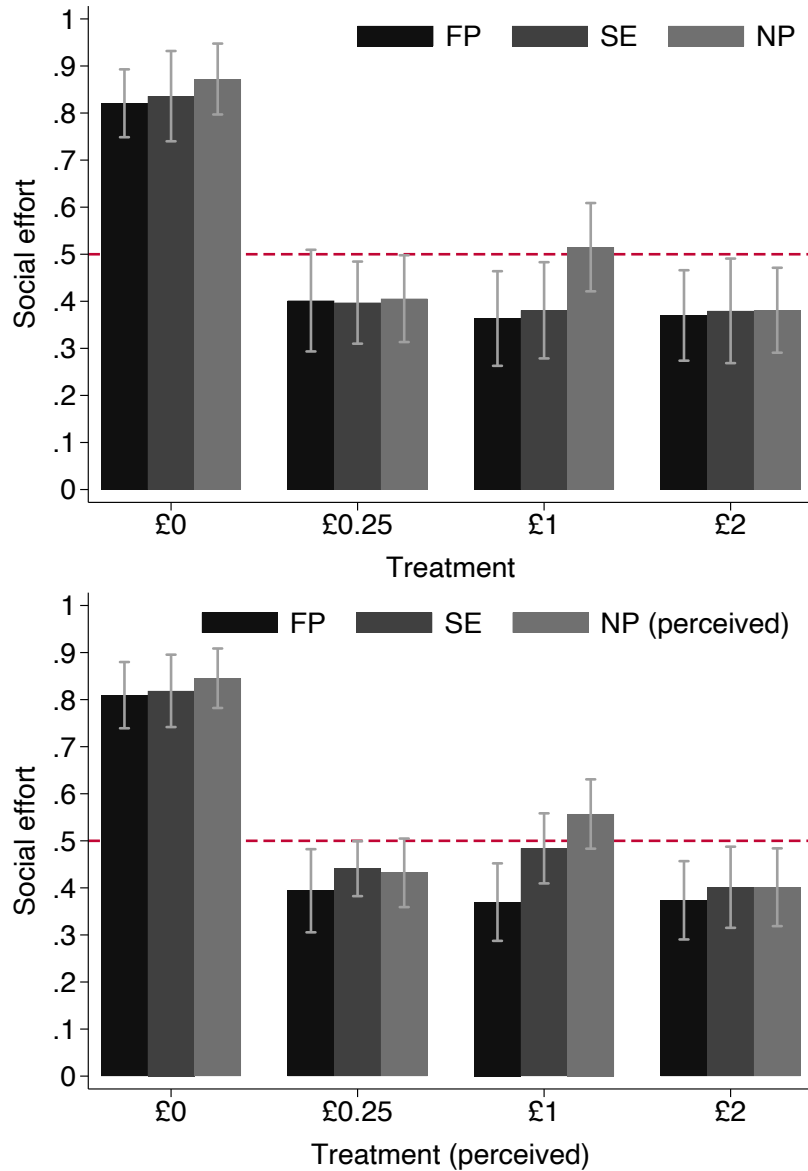


Figure H.4: Social effort for those passing all checks (top) and perceived treatments (bottom), with 95% confidence intervals; the dashed line represents a fully balanced effort allocation.

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 H.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 H.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 H.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 H.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
<i>N</i>	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 H.3: **Additional Study: Randomization Check**

Characteristic	(1) Mean	(2) St. dev.	(3) Joint <i>p</i> -value	(4) <i>R</i> 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 H.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 <i>t</i>-tests of equality of means across treatments, <i>p</i>-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 <i>t</i>-tests of equality of means across organizational forms, <i>p</i>-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 <i>t</i>-tests of equality of means across treatments, <i>p</i>-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 <i>t</i>-tests of equality of means across organizational forms, <i>p</i>-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 H.2 for sample sizes by treatment.

Table H.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 (0.552)	0.136 (0.994)	0.206 (0.944)	-0.390 (0.029)	-0.351 (0.050)	-0.344 (0.050)	-0.343 (0.051)
£1	-1.501 (0.559)	-2.316 (1.011)	-1.904 (0.986)	-0.392 (0.029)	-0.387 (0.049)	-0.393 (0.046)	-0.403 (0.046)
£2	-0.718 (0.536)	-0.286 (0.971)	-0.160 (0.930)	-0.420 (0.028)	-0.388 (0.049)	-0.370 (0.048)	-0.371 (0.049)
For-profit		0.796 (0.931)	0.600 (0.899)		-0.017 (0.047)	-0.027 (0.046)	-0.024 (0.047)
£0.25 × For-profit		-1.090 (1.357)	-1.079 (1.280)		-0.001 (0.076)	0.003 (0.073)	-0.002 (0.074)
£1 × For-profit		0.994 (1.351)	0.745 (1.324)		-0.015 (0.070)	-0.010 (0.066)	-0.007 (0.066)
£2 × For-profit		-0.283 (1.268)	-0.295 (1.208)		-0.011 (0.072)	-0.037 (0.070)	-0.038 (0.071)
Nonprofit		0.167 (1.000)	0.349 (0.926)		0.086 (0.043)	0.049 (0.044)	0.051 (0.044)
£0.25 × Nonprofit		-0.297 (1.382)	-0.530 (1.304)		-0.113 (0.067)	-0.067 (0.066)	-0.069 (0.067)
£1 × Nonprofit		1.436 (1.416)	1.165 (1.375)		0.002 (0.068)	0.021 (0.065)	0.027 (0.065)
£2 × Nonprofit		-1.017 (1.387)	-1.383 (1.324)		-0.083 (0.067)	-0.070 (0.066)	-0.077 (0.066)
R^2	0.010	0.019	0.126	0.244	0.254	0.356	0.361
B. Incentive presence							
Incentive	-0.847 (0.447)	-0.826 (0.822)	-0.628 (0.785)	-0.400 (0.022)	-0.376 (0.039)	-0.370 (0.039)	-0.373 (0.039)
For-profit		0.796 (0.928)	0.598 (0.896)		-0.017 (0.047)	-0.028 (0.046)	-0.025 (0.047)
Incentive × For-profit		-0.130 (1.078)	-0.212 (1.038)		-0.008 (0.057)	-0.013 (0.055)	-0.014 (0.056)
Nonprofit		0.167 (0.996)	0.340 (0.923)		0.086 (0.042)	0.048 (0.044)	0.050 (0.044)
Incentive × Nonprofit		0.062 (1.145)	-0.226 (1.079)		-0.065 (0.052)	-0.037 (0.052)	-0.038 (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 H.6: **Additional Study: Heterogeneity in Social Effort**

	For-profit		Social enterprise		Nonprofit	
	(1)	(2)	(3)	(4)	(5)	(6)
A. Gender						
£0.25	-0.410 (0.083)	-0.361 (0.078)	-0.314 (0.077)	-0.353 (0.078)	-0.543 (0.060)	-0.487 (0.064)
£1	-0.406 (0.076)	-0.397 (0.070)	-0.382 (0.078)	-0.390 (0.072)	-0.518 (0.076)	-0.480 (0.077)
£2	-0.411 (0.081)	-0.377 (0.077)	-0.373 (0.072)	-0.375 (0.072)	-0.521 (0.065)	-0.496 (0.066)
Female	0.032 (0.069)	0.002 (0.061)	0.126 (0.064)	0.043 (0.071)	0.003 (0.055)	0.022 (0.059)
£0.25 × Female	0.121 (0.112)	0.080 (0.108)	-0.084 (0.101)	-0.033 (0.105)	0.205 (0.084)	0.128 (0.087)
£1 × Female	0.005 (0.102)	0.012 (0.094)	-0.030 (0.099)	-0.039 (0.096)	0.226 (0.096)	0.172 (0.098)
£2 × Female	0.023 (0.106)	-0.048 (0.105)	-0.026 (0.095)	-0.024 (0.096)	0.095 (0.091)	0.088 (0.094)
R^2	0.223	0.376	0.255	0.386	0.366	0.443
B. Social sector experience						
£0.25	-0.417 (0.066)	-0.380 (0.062)	-0.342 (0.060)	-0.340 (0.060)	-0.478 (0.052)	-0.428 (0.051)
£1	-0.383 (0.060)	-0.378 (0.054)	-0.384 (0.059)	-0.389 (0.054)	-0.386 (0.054)	-0.385 (0.052)
£2	-0.374 (0.060)	-0.384 (0.058)	-0.375 (0.058)	-0.371 (0.056)	-0.482 (0.055)	-0.453 (0.056)
Experience	0.049 (0.071)	0.040 (0.060)	0.075 (0.065)	0.096 (0.072)	-0.032 (0.063)	-0.075 (0.072)
£0.25 × Experience	0.221 (0.117)	0.205 (0.107)	-0.022 (0.110)	-0.108 (0.112)	0.059 (0.094)	0.028 (0.102)
£1 × Experience	-0.086 (0.112)	-0.037 (0.108)	0.001 (0.107)	-0.079 (0.105)	-0.005 (0.121)	0.005 (0.119)
£2 × Experience	-0.123 (0.137)	-0.085 (0.139)	-0.021 (0.113)	-0.053 (0.121)	0.038 (0.097)	0.021 (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 H.5.