Striking a Balance: Revenue Drift, Incentives, and Effort Allocation in Social Enterprises*

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

October 22, 2019

Abstract

Social enterprises often exhibit revenue drift, i.e. an excessive focus on purpose at the expense of profits. Despite the threats this poses for organizational performance and survival, social entrepreneurs are reluctant to use performance-based pay due to the perceived incongruence of incentives with social impact and the risk of mission drift. We argue that monetary incentives can elicit a balanced effort allocation by redirecting employee effort to commercial tasks and by attracting workers who are less prone to exerting excessive social effort. We vary incentive strength in an online, real-effort experiment and find that both modest and strong incentives produce a more balanced effort allocation by redirecting worker attention to commercial tasks. While strong incentives lead to a small decrease in workers' social motivation, modest incentives do not affect social enterprise workforce composition. Social enterprises that combine mission and monetary rewards not only attract more workers, but also succeed in directing worker attention to both commercial and social tasks.

JEL Classification: D22, J33, L21, L31.
Keywords: social enterprise, hybrid organization, social motivation, compassion, incentives, multitasking, online experiment, effort allocation.

* We thank seminar and conference participants at Copenhagen Business School, DRUID Academy, the 1st IESE-LUISS Conference on Hybrid Organizations, and the Oxford Residence Week for Entrepreneurship Scholars, as well as Orsola Garofalo, Markus Simeth, and Angelique Slade Shantz for valuable comments. We gratefully acknowledge the financial support of the Mærsk-McKinney-Møller Chair in Entrepreneurship at Copenhagen Business School. Part of this research was carried out while Vladasel was a Fox International Fellow at Yale.
† Corresponding author. Universitat Pompeu Fabra and Barcelona GSE, theodor.vladasel@upf.edu
‡ Ivey Business School, Western University, University of Aberdeen, and IZA, sparker@ivey.ca
§ Universiteit van Amsterdam and Tinbergen Institute, r.sloof@uva.nl
¶ Vrije Universiteit Amsterdam, CBS, CEPR, IZA, and Tinbergen Institute, c.m.van.praag@vu.nl

1 Introduction

As hybrid organizations combining commercial and social logics (Austin et al., 2006; Dacin et al., 2011; Pache and Santos, 2013; Grimes et al., 2018), social enterprises must allocate scarce employee effort between commercial and social mission tasks in order to deliver on their dual objectives (Battilana and Dorado, 2010; Battilana and Lee, 2014; Besharov, 2014).[1] In practice, many social enterprise employees prioritize purpose over profits, creating the risk that insufficient effort is allocated to promoting economic performance (Battilana et al., 2015; Stevens et al., 2015; Staessens et al., 2018). An excessive focus on social impact to the detriment of generating revenue – or 'revenue drift' (Ebrahim et al., 2014) – may ultimately threaten social enterprises' financial sustainability and survival (Tracey et al., 2011; Smith et al., 2013).

In such circumstances, one might expect social enterprises to offer pecuniary rewards to induce employees to balance their effort between commercial and social tasks. Yet, social enterprises rarely make use of monetary incentives (Battilana and Lee, 2014). For their part, potential employees may perceive a tight coupling of pay and commercial performance as incompatible with social enterprise values and resources (Austin et al., 2006; Tracey et al., 2011; Besharov, 2014). Identity tension is especially likely if employees associate monetary rewards with the competitive 'bonus culture' and profit motive traditionally characterizing a commercial logic (Dees, 2012; Benabou and Tirole, 2016; Dimitriadis et al., 2017). At the same time, social entrepreneurs may be anxious about the danger of mission drift, which may result when mission-driven organizations emphasize commercial over social imperatives (Ebrahim et al., 2014). This may induce them to eschew monetary incentives (Bacchiega and Borzaga, 2001).

In this paper, we question whether these concerns are well-founded. As social sector organizations face increasing competition for talent, pressure to professionalize, and market competition (Hwang and Powell, 2009; Smith et al., 2013; Battilana and Lee, 2014), many hybrids are turning

[1] Social enterprises occupy a continuum along the trade-off between economic and social goals (Besharov and Smith, 2014; Shepherd et al., 2019). Our focus is on firms where this trade-off is most pronounced, i.e. where multiple logics are core to organizational functioning and provide contradictory prescriptions for action.

their attention to the financial sustainability of their operations and the role of staffing practices in meeting their dual objectives (Battilana and Lee, 2014; Smith and Besharov, 2019). Thus, the role of incentives cannot be dismissed outright, especially for social enterprises whose finances are precarious. Specifically, we ask: can monetary incentives for commercial performance help social enterprises achieve a more balanced allocation of employee effort between commercial and social tasks? If so, how sensitive is effort allocation to monetary rewards – in other words, how 'steep' do incentives need to be to attain a balanced effort allocation? To answer these questions, we conceptually theorize about and conduct an experiment to distinguish between two mechanisms that affect the balance of commercial and social tasks. At the intensive margin, by rewarding a particular dimension of effort, performance-based pay performs an attention-directing role (Ocasio, 1997), which shifts the effort allocation *of a given workforce*. At the extensive margin, performance based-pay performs a sorting function, related to a *changing workforce composition* (Lazear, 2000), acknowledging that less socially motivated individuals may select into social enterprises that offer high-powered incentives. The intensive and extensive margin effects therefore capture both employees' effort allocation and who becomes an employee in the first place.

We draw on organizational theory and organizational economics (Eisenhardt, 1989; Kaplan and Henderson, 2005) to analyze these questions conceptually. Given the predominance of socially motivated employees and a mission emphasis in social enterprises (Miller et al., 2012; Smith et al., 2013; Besley and Ghatak, 2017), we predict that an absence of pay for performance practices generates an unbalanced effort allocation, with employees favoring social impact over revenue generation. In a multitasking framework where commercial and social tasks are complementary for the firm (Holmström and Milgrom, 1991), workers' high levels of intrinsic pro-social motivation are expected to induce excessive social task effort, a form of adverse specialization (MacDonald and Marx, 2001), endangering social enterprises' financial performance. We propose that modest incentives (i.e. small performance bonuses) can restore the balance of employee effort between commercial and social tasks, via action on *both* the intensive and extensive mar-

gins. However, we expect that strong incentives (i.e. large performance bonuses) risk causing mission drift by distorting effort too far in the direction of commercial imperatives.

We conduct an incentivized, real-effort, online experiment to test our theoretical predictions. This approach overcomes the limited availability of data on compensation practices and individual effort allocations in social enterprises, as well as the endogeneity of pay for performance adoption decisions; moreover, it allows us to unpack the mechanisms connecting effort allocation with incentives. Our experiment uses a labor market framing, where subjects are asked to behave as employees of companies whose descriptions match typical for-profits, non-profits, and social enterprises. Subjects move a set of sliders (Gill and Prowse, 2012) to allocate effort between a commercial and a social task, associated respectively with own and 'good cause' payoffs. To disentangle the causal mechanisms underlying our hypothesized effects, we manipulate the strength of commercial performance incentives (i.e. own payoffs) within social enterprises and employees' ability to choose their preferred contract, keeping good cause payoffs constant. Our hypotheses are broadly supported. Monetary incentives elicit a more balanced effort allocation, regardless of their steepness, while we find a small, but significant downward shift in employees' social motivation when strong incentives are offered.

The paper makes three main contributions to the organization science literature. First, we extend the logic of incentive theory to social enterprises and highlight the adverse specialization problem stemming from an excessive mission emphasis. Pecuniary rewards allow firms to overcome revenue drift by focusing employee attention on generating the revenue required for delivering their social mission (Tracey et al., 2011; Smith et al., 2013; Stevens et al., 2015). We believe that our conceptualization of the social enterprise as a multitasking setting with motivated workers can aid future theory-building efforts in the area of hybrid organization compensation practices and organizational design (Battilana and Lee, 2014).

Second, we implement a novel experimental design which causally isolates the normative role of incentives (Ocasio, 1997; Kaplan and Henderson, 2005; Ethiraj and Levinthal, 2009; Wolfolds,

2018) from their effect on workforce composition (Lazear, 2000; Cadsby et al., 2007; Deserranno, 2019). Contrary to a common belief among social enterprise actors that monetary rewards may be perceived as unfair, controlling, or incongruent with organizational values (Austin et al., 2006; Tracey et al., 2011; Dees, 2012), our experimental results show that pay for performance need not deter socially motivated workers from joining social enterprises. Instead, (modest) incentives work mainly by directing attention towards commercial tasks.

Third, we trace out the implications of our findings for social entrepreneurs. By embracing heterogeneity in other-regarding values (Miller et al., 2012; Stevens et al., 2015; Besley and Ghatak, 2017) and considering the mechanisms linking incentives to effort (Cadsby et al., 2007), we show that pay for performance does not necessarily crowd-out social motivation by workers and cause mission drift. This challenges a common view and normative pressure in the social enterprise community that monetary rewards should be avoided (Bacchiega and Borzaga, 2001). We therefore tentatively recommend that social entrepreneurs worried about revenue drift and adverse specialization should consider incorporating modest incentives into their worker compensation schemes, possibly alongside socialization practices (Battilana and Dorado, 2010) and other governance mechanisms (Ebrahim et al., 2014; Smith and Besharov, 2019).

2 Theoretical Background

Social enterprises are double-bottom line organizations. They aim to deliver a social mission, as well as the financial performance which enables them to survive and prosper, and so advance their social mission. In what follows, we study the allocation of social enterprise employees' effort between these two objectives. We first characterize the tensions these 'dual mission' organizations often encounter, before clarifying definitions and assumptions about effort, employee social motivation, incentives, and actions within such hybrid organizations. We then discuss effort allocation in the absence and presence of financial incentives offered by social enterprises.

2.1 Social and commercial tensions in social enterprises

The social enterprise hybrid organizational form tackles social challenges through business means (Dees, 2001; Margolis and Walsh, 2003; Zahra et al., 2009). Despite the diversity of their business models (Mair et al., 2012), for-profit social enterprises represent a common organizational form, with clearly stated economic and social goals, or a *double bottom line* (Austin et al., 2006; Dacin et al., 2011; Stevens et al., 2015). In 'integrated' hybrids, profit and purpose are pursued jointly, often through direct transactions with beneficiaries (Smith et al., 2013; Besley and Ghatak, 2017; Eldar, 2017).[2] However, profit and purpose represent goals associated with commercial/for-profit and charitable/non-profit institutional logics, respectively (Pache and Santos, 2010; Battilana and Lee, 2014; Besharov and Smith, 2014). Thus, they often place conflicting demands on organizations' attention and resources (Dacin et al., 2011; Stevens et al., 2015).

The trade-off between commercial and social goals is reflected at all levels of the organization, as employees decide how to allocate their effort. The following examples illustrate this trade-off. Loan officers in commercial microfinance organizations must balance loan size, interest rates, and potential profits on one hand, with reaching the underprivileged target population on the other (Battilana and Dorado, 2010; Pache and Santos, 2010; Besharov and Smith, 2014; Canales, 2014; Wry and Zhao, 2018). In work integration social enterprises (WISEs), employees must ensure financial sustainability through business activity, as well as disadvantaged workers' skill development (Tracey et al., 2011; Pache and Santos, 2013; Battilana et al., 2015). For 'base of the pyramid' firms, employees must ensure products are sold at prices above cost, yet within the reach of the target population (Hockerts, 2015; Santos et al., 2015). Finally, environmental social enterprises must balance generating revenue with costly measures to protect the environment (Pacheco et al., 2014).

[2] By contrast, 'differentiated hybrid' social enterprises rely on cross-subsidization, where commercial revenue is devoted to a social goal (Baron, 2007; Ebrahim et al., 2014). For instance, TOMS Shoes uses its profits to provide shoes and fund investments in hygiene in developing countries (Marquis and Park, 2014).

2.2 Terminology and assumptions

Before developing our hypotheses, it is helpful to define terms and clarify the setting we analyze. First, we assume that social enterprise employees must decide how to allocate their effort between two tasks. One task delivers the commercial objective C, e.g. generating sales; the other delivers the social mission S, e.g. alleviating poverty. If a fixed and finite amount of effort is available, one can equivalently analyze the proportion of time a worker devotes to one task rather than the other. We assume that spending more time on one task necessarily means spending less time on the other, and refer to *balanced effort* as allocations in which roughly similar amounts of effort are allocated to each task.[3] In line with standard principal-agent theory, we assume that managers cannot perfectly measure and direct how employees allocate their effort, which is partly at workers' discretion (Battilana and Dorado, 2010; Henderson and Van den Steen, 2015).

Second, we assume there are two types of employees. The I type is intrinsically motivated by the social enterprise mission and has a strong desire to do good for others. The E type is primarily extrinsically motivated (i.e. by money). I types are more predisposed than E types to allocate effort towards the S task and more willing to join non-profits or social enterprises rather than for-profit firms. Another standard principal-agent theory assumption is that founders are imperfectly informed about which type of employee is which; that is, there are 'hidden types'. This rules out job separation arrangements within the organization whereby different worker types could be assigned to work solely on the C or S task.[4] All employees are assumed to pursue both tasks, due to social enterprises' integrated hybrid nature and typically small scale and staffing constraints (Smith et al., 2013; Santos et al., 2015; Battilana, 2018). For the social enterprise, commercial and social tasks are complements (MacDonald and Marx, 2001).

Third, we consider two types of incentives. Incentives are an integral part of firms' 'structural

[3] We assume a capped level of total effort due to our core focus on effort allocation. While our experiment allows subjects to also adjust this margin, we find no differences in total effort across treatments.

[4] Another possibility is non-separability of missions within job assignments, i.e. both missions are necessarily present simultaneously; yet workers have discretion over which to emphasize. By contrast, one WISE studied by Battilana et al. (2015) does practice job separation and enforces 'spaces of negotiation', but this approach risks disconnecting different areas of the organization (Ethiraj and Levinthal, 2009).

distribution of attention' (Ocasio, 1997) and perform a normative function, directing employee effort (Kaplan and Henderson, 2005). Pecuniary rewards can ensure complex organizations reach their goals even when not all outcomes can be measured accurately (Ethiraj and Levinthal, 2009). We distinguish between 'low-powered' and 'high-powered' incentives. Low-powered incentives consist of a fixed wage unrelated to the allocation of effort to, or performance in, either task (Holmström and Milgrom, 1991). By contrast, high-powered incentives comprise a combination of fixed wage and pay-for-performance on a given outcome, occupying a continuum ranging from 'modest' incentives (i.e. a small bonus tied weakly to task performance) to 'strong' incentives (i.e. a large bonus tied closely to task performance). In principle, high-powered incentives can be attached to either task (Wolfolds, 2018), but in practice tend to be applied to the C task only, given the difficulty of measuring social impact in a timely and standardized manner (Austin et al., 2006; Dacin et al., 2011).

Fourth, incentives affect one or two margins of action, which economists label 'intensive' and 'extensive'. The intensive margin relates to how incentives affect the chosen effort allocation *by a given pool* of employees. The extensive margin relates to how incentives affect the *composition* of the social enterprise's pool of employees itself. Along this margin, employees can self-select into either social enterprises or an alternative organizational form. For example, I types are more likely to select into social enterprises or non-profit organizations than E types, all else equal; the latter may select instead into for-profits (Barigozzi et al., 2018).

2.3 Low-powered incentives in social enterprises

We propose that social enterprise employees respond to incentives in ways that affect both the intensive and extensive margins. We first consider the case where low-powered incentives prevail. This is the predominant case in practice: relatively few social enterprises offer financial bonuses for commercial performance (Bacchiega and Borzaga, 2001; Becchetti et al., 2013).

Most social enterprises are founded by compassionate individuals seeking to maximize the social return on their investments (Miller et al., 2012; Santos, 2012; Grimes et al., 2013; Bacq and

Alt, 2018). Their emphasis on social mission attracts similarly-minded employees (Besley and Ghatak, 2017), who prefer to exert high levels of S task effort (Besley and Ghatak, 2005; Bell and Haugh, 2014; Henderson and Van den Steen, 2015; Tonin and Vlassopoulos, 2015; Burbano, 2016; Cassar, 2019). Moreover, social enterprise founders and employees often have experience working within an institutional logic 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). Despite often acknowledging the importance of market mechanisms for financial sustainability, social enterprises rely heavily on the mission they champion to attract potential customers, investors, and employees (Dees, 2001; Renko, 2013; Smith et al., 2013; Fosfuri et al., 2016).

Social enterprises may offer low-powered incentives for at least two reasons. First, incentive theory suggests that high-powered incentives for the C task risk shifting employees' effort away from the less easily-measured and rewarded S task and towards the more-easily measured and rewarded C task (Holmström and Milgrom, 1991). This represents an intensive margin effect. Founders may be averse to this shift on the grounds of it being perceived as incongruent with organizational values (Tracey et al., 2011; Hossain and Li, 2014; Andersson et al., 2017). Second, founders may worry that high-powered incentives attract financially-motivated employees uninterested in the S task (Bacchiega and Borzaga, 2001; Austin et al., 2006). This extensive margin outcome risks causing 'mission drift', whereby the enterprise abandons social concerns in favor of profit-seeking activities (Ebrahim et al., 2014). Founders may object to this outcome if it threatens to change the composition and hence the long-term purpose of the enterprise.

What are the implications of eschewing high-powered incentives for social enterprises? Beyond the pro-social motivational effects of missions, organizational economics suggests that I types have a lower psychic cost of effort for the S task (Murdock, 2002; Schnedler, 2008). In the absence of incentives to do otherwise, their effort allocation is distorted towards the less costly

task, leading to 'adverse specialization' (MacDonald and Marx, 2001). For social enterprises, the risk is that I employees exert too much effort on the S task and too little on the C task, reducing financial performance and hence limiting the resources the enterprise needs to survive (Smith et al., 2013). Ebrahim et al. (2014) refer to this outcome as 'revenue drift', which we summarize as our baseline hypothesis:

Hypothesis 0 *Low-powered incentives are associated with excessive effort devoted to the S task, at the expense of the C task.*

There are numerous examples of revenue drift. Tracey and Jarvis (2006) and Tracey et al. (2011) discuss how a social enterprise tackling homelessness threatened its own survival by focusing excessively on beneficiary needs at the expense of required operational investments. Beer et al. (2017) document how employees of a youth development social enterprise focused on ensuring disadvantaged individuals had adequate housing, but overlooked the importance of rent collection to keep the business going.[5] A fair trade social enterprise insufficiently heeding customer demands and prioritizing producers instead suffered from operational difficulties for several years, limiting their ability to serve beneficiaries (Davies and Doherty, 2018). Staessens et al. (2018) find that Belgian WISEs predominantly use inputs to achieve worker reintegration outcomes rather than to generate revenue, results echoed by Battilana et al. (2015) and Stevens et al. (2015). Overall, social enterprises' cash flow problems, often listed as an impediment to growth along with lack of access to finance, appear partly attributable to an excessive focus on social impact (Smith et al., 2013).

This raises the question of whether high-powered incentives – coupling pay with financial performance – can be used to avoid the adverse specialization problem, helping social enterprises strike a more balanced allocation of effort across the C and S tasks. And if so, would modest or strong high-powered incentives work best? We explore these questions next.

[5] One housing worker 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*".

2.4 High-powered incentives in social enterprises

Can high-powered incentives mitigate the problems caused by excessive attention to the social task and thus achieve a balanced effort allocation? To answer this question, we turn to incentive theory. If employees had no preference regarding performing C or S tasks, i.e. there was no heterogeneity along the I-E dimension, then insights from the classical two-task multitasking problem would apply. In that case, low-powered incentives would lead workers to exert effort on both tasks (Holmström and Milgrom, 1991), while any kind of high-powered incentive would run the risk of dramatically unbalancing the effort allocation by directing all employee effort towards the C task. However, the classical analysis needs to be modified when dealing with the more realistic case where some employees are I types and some are E types.

A formal analysis of the case where there is a mixture of E and I employees suggests that both types' effort is responsive to high-powered incentives (Canton, 2005). Such incentives signal the importance of the incentivized task (Kaplan and Henderson, 2005) and make it more financially attractive for employees to perform it (Holmström and Milgrom, 1991; Canton, 2005). In social enterprises, this is equivalent to a distortion of effort towards the C task, which may lead to mission drift. Not only are E types likely to shift their effort to the C task in response to high-powered incentives, but I types may also be induced the same way. The reason is that strong financial incentives may 'crowd out' intrinsic motivation by I types. Evidence from psychology and economics (Deci et al., 1999; Osterloh and Frey, 2000; Gneezy et al., 2011) indicates that strong incentives may be perceived to have a controlling nature, which generally undermines intrinsic motivation for the S task. The outcome on the intensive margin can be for demoralized I types to respond by allocating too much effort to the C task and too little to the S task.

The core of our argument is that this outcome does not necessarily arise when more modest high-powered financial incentives are used. These can generate a gentler shift among I types on the intensive margin, from strong effort concentration on the S task towards a more balanced allocation between the two tasks. For the I types predominant in social enterprises, monetary

incentives imply a trade-off between the benefit of additional income from commercial activities and the lower satisfaction from reduced effort dedicated to the social mission. In principle, modest incentives can locate an intermediate position along the trade-off. Modest incentives are less likely to be perceived as controlling, and may even be presented as supportive of the social mission by promoting a balanced C-S perspective towards achieving the organization's goals.

There is another, subtler way that modest incentives may generate a more balanced effort allocation along the intensive margin. Brüggen and Moers (2007) show that exposure to a social norm promoting the social mission can mitigate the 'distorting' effect of financial incentives, where the distortion is associated with under-provision of S task effort. Both E and I types can be affected by a social norm, reducing their tendency to respond to incentives by making sharp shifts in their effort allocation away from S towards C; but other-oriented workers' social effort is crowded out to a smaller extent, such that they provide a more balanced effort allocation (Nellas and Reggiani, 2015; Benabou and Tirole, 2016). In social enterprises, the S task effectively elicits social norms, while monetary incentives are associated with a dangerous, agentic 'bonus culture', deeply embedded in a commercial logic (Dees, 2012; Benabou and Tirole, 2016). Hence, modest incentives reduce the risk that employees 'over-react' and devote excessive effort to the C task, compared with strong incentives.

Together, these arguments suggest that social enterprise employees respond to modest high-powered monetary incentives by shifting part of their effort towards the C task, while continuing to expend significant effort on the S task; as a result, their effort allocation is more balanced between tasks. In contrast, strong high-powered incentives may lead to excessive effort devoted to the C task at the expense of the S task. Thus:

Hypothesis 1 *Relative to low-powered incentives, modest high-powered incentives generate a more balanced effort allocation between C and S tasks along the intensive margin.*

Hypothesis 2 *Relative to modest high-powered incentives, strong high-powered incentives generate excessive effort devoted to the C task along the intensive margin.*

In addition to directing worker effort choices along the intensive margin, incentives may also

perform a sorting function, whereby more financially-motivated E types are attracted to the organization (Lazear, 2000; Cadsby et al., 2007; Benabou and Tirole, 2016). Incentives thus operate along the extensive margin as well. Attracting E types to the social enterprise can be unwelcome to social entrepreneurs (Bacchiega and Borzaga, 2001), especially if the perceived incongruence between incentives and social mission endangers employees' identification with organizational goals (Akerlof and Kranton, 2005; Besharov, 2014; Andersson et al., 2017) — or if incentives are perceived as 'unfair' (Tracey et al., 2011). Both founders and I-type employees may express concerns over practices whose origin lies in a commercial logic and which are prevalent in for-profit companies (Lazear and Shaw, 2007; Besharov, 2014; Gerhart and Fang, 2014); and I-type employees may refrain from joining a social enterprise espousing such profit-oriented practices.

A social enterprise offering strong high-powered incentives may attract E types, by allowing them to reap personal rewards from performing the C task while exerting a smaller share of S task effort. But by recruiting more E types through this mechanism, I types may feel that their other-oriented values are under-appreciated, or even devalued, by the social enterprise, which reduces their intrinsic motivation (Akerlof and Kranton, 2005; Besharov, 2014). The shift towards less socially motivated workers may contrast with the social norm of 'doing good', implying that strong high-powered incentives may deter I types from joining a social enterprise that uses such practices and encourage those currently working in ones that adopts such incentives to quit. The overall outcome is to change the composition of the social enterprise workforce, reducing the proportion of employees focusing more on the S task and increasing the proportion of employees focusing more on the C task.

By contrast, modest high-powered incentives place a limit on in-selection of E types into the social enterprise, as well as out-selection of I types. Loosely coupled pay-for-performance may be only weakly attractive for E types, if they can still do better working in a for-profit, where strong high-powered incentives are more common. This restricts the number of E types willing

to join the social enterprise, diluting their impact on the organization's goal focus, and reducing the negative impact on I type selection (Jones et al., 2018). Moreover, a loose coupling of pay and performance is less likely to challenge the social norm of 'doing good', maintaining a pro-social tone for the organization's values. As a result, the overall effect of a small compositional change following the introduction of modest incentives is likely to be a slight decrease in the proportion of employees focusing more on the S task, and a slight increase in the proportion of employees focusing more on the C task. We summarize these predictions as follows:

Hypothesis 3 *Relative to low-powered incentives, modest high-powered incentives generate a more balanced effort allocation between C and S tasks along the extensive margin.*

Hypothesis 4 *Relative to modest high-powered incentives, strong high-powered incentives generate excessive effort devoted to the C task along the extensive margin.*

This second set of hypotheses works in the same direction as the first. The difference lies in the mechanism by which effort is affected. Whereas hypotheses 1 and 2 relate to the intensive margin, hypotheses 3 and 4 relate to the extensive margin. Our experiment is designed to discriminate between these mechanisms and identify which one is more salient in practice.

3 Experimental Design

Our analysis uses an online, incentivized, real effort experiment with a labor market framing. Subjects allocate effort between a commercial and a social action as hypothetical employees of different fictional companies, whose descriptions match those of typical for-profits (FP), non-profits (NP), or social enterprises (SE). These firms, or 'contracts', provide similar services but have different objectives, corresponding to a realistic labor market choice (despite the absence of explicit labels for these organizations). For instance, a workforce integration SE contract 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". The equivalent FP and NP company descriptions emphasize revenue generation by providing services (C task) and, respectively, the charitable goal of improving disadvantaged groups' welfare (S task). Both tasks are available to subjects in each contract and are described in relation to the services the company provides, together with the payoffs they generate.[6]

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 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 (FP, NP, SE), displayed randomly to avoid order bias effects. Worker self-selection is avoided in this setting, allowing us to study intensive margin effects by varying the SE 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 effect of incentives along the extensive margin. 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 real effort task adapted from the slider task introduced by Gill and Prowse (2012) to measure effort allocation. This task consists of 15 horizontal sliders that can be moved to positions equivalent to exerting effort on the commercial or social task. The sliders are labeled from 0 to 100 and initially positioned at 50. The commercial task, C, requires placing the slider at 25 and carries a payoff for the subject, determined by the commercial task incentive level (or 'bonus') in a given contract. The social task, S, requires placing the slider at 75 and carries

[6] To avoid priming, we did not use the 'for-profit', 'non-profit' or 'social enterprise' labels and the associated abbreviations in our experiment (except for the final questions on demographics in Part 4). The social enterprise label, beyond the description of social enterprises' desired effort allocation, is likely to evoke a strong motivation to exert S task effort, which may amplify revenue drift with low-powered incentives. As a result, our avoidance of labels makes it more difficult to detect revenue drift and makes our tests more conservative.

a payoff for the selected good cause.[7] Incorrectly placed sliders (e.g. position 44) carry no payoffs. 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 the position of each slider as many times as they wish. Most importantly, subjects do not simply state how they would like to allocate effort, but must drag sliders in a way that reflects their preferences, such that they expend real effort. While neither action has a monetary cost, both tasks require the subject to physically move a slider. In addition, the social task entails foregoing the monetary rewards of the commercial task. It is in this sense that the slider task allows us to capture real effort.

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 are aware when a social enterprise uses strong incentives (£1), allowing us to detect extensive margin effects. We measure *Commercial effort* and *Social effort* as counts of the number of sliders moved to the C and, respectively, S task. The more similar *Commercial effort* and *Social effort* are, the more balanced the effort allocation.

Treatments The experiment features two different dimensions, designed to tackle i) the effect of incentives on effort allocation, and ii) the contributions of the extensive and intensive margins. First, beyond the participation fee (equivalent to a fixed wage), we vary the strength of the SE C task incentive between subjects. At one extreme, the SE contract offers 'low-powered incentives' (£0 bonus), in line with current social enterprise practice and identical to the NP contract; at the other extreme, the SE contract offers the same 'strong high-powered incentives' (£1 bonus) as the FP contract; we refer to intermediate levels as 'modest high-powered incentives' (£0.25 or £0.50 bonus). Subjects are randomly assigned to a fixed incentive level, which they face throughout

[7] For example, in the workforce reintegration mission, the 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 ... to the good cause'.

the experiment. While the maximum £1 payment per slider serves as a normalization and is easily understood by subjects, it is the *relative* strength of incentives that allows us to test our hypotheses, rather than the absolute value of payments per slider.

Second, we allow subjects to choose their preferred contract from the previously encountered FP, NP, and SE contracts. That is, Part 3 incentives (and, hence, payoffs) and company descriptions are exactly the same as in Part 2, but we now allow for individual sorting 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 FP bonus at £1, the NP bonus at £0, and the good cause payoff in all contracts at £1 per slider throughout the experiment. The £1 FP bonus allows us to benchmark SE incentive strength (i.e. the SE bonus is 0%, 25%, 50% or 100% of the FP bonus) across the range of possible incentives. The £1 good cause payoff, symmetrical to the FP bonus, reflects the high social returns to S 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 financially rewarded (which is the main focus of our experiment).

Preferred Mission To ensure S task saliency, we allow subjects to choose their preferred good cause (Tonin and Vlassopoulos, 2015; Cassar, 2019). The options available – The Big Issue Foundation, Fairtrade Foundation, and Water Aid – are selected as charities whose goals match representative social enterprise missions, namely workforce reintegration, fair and equitable trade, and environmental protection (Mair et al., 2012; Eldar, 2017). These missions are therefore likely to resonate with a broad share of potential social enterprise employees.[8] Moreover, these missions have a clear multitasking component and can be pursued through a for-profit, non-profit, or social enterprise model, allowing us to construct realistic descriptions of the FP, NP, and SE contracts. Furthermore, the actual charity organizational form allows us to credibly commit to donating the good cause payoffs generated by subjects in the experiment.

[8] More specific, narrower social concerns are likely to attract employees with strong matching preferences, potentially exacerbating the revenue drift problem. Our choices are, therefore, rather conservative.

Social Preferences As our theoretical framework suggests individual social motivation affects both effort allocation and self-selection, we elicit subjects' social preferences in various ways. Following the social entrepreneurship 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.[9] For example, one item asks subjects how strongly they agree with the following statement: "I seldom think about the welfare of people 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, 2018). 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., 2016), *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 non-profits and social enterprises (Tonin and Vlassopoulos, 2015).[10]

These measures capture different aspects of motivation, so we aggregate them into a composite *Social motivation* measure using principal component analysis. 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 parsimony, we discuss *Compassion* and *Social motivation* in our main analysis, and report results for other measures as robustness checks. We also measure self-reported risk taking (*Risk*) and future discounting (*Time*) preferences on a 0-10 scale (Dohmen et al., 2011; Falk et al., 2016), as they may affect individuals' perception of incentives and effort allocation.[11]

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

[10] The latter variable captures professional work relationships with such social organizations (for example, as a joint venture with a for-profit where the individual is employed).

[11] The self-reported answers to the general risk, time, and willingness to share questions are reliable predictors of behavior and consistent with incentivized elicitations (Dohmen et al., 2011; Falk et al., 2016).

Procedure We recruit subjects on Prolific Academic, a UK-based online platform designed for surveys and experiments and geared towards researchers and startups. Prolific compares favorably with Amazon Mturk in terms of 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) and an online study allows us to reach a broad population of potential employees (Bitektine et al., 2018). We restrict the subject pool to UK residents aged 18-64 who are students or active in the labor force (i.e. not 'homemakers', disabled, or retired). To achieve gender balance and perform comparisons between men and women, we stratify the randomization by gender, using the gender variable Prolific previously required subjects to report. 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 participants with prior approval rates higher than 90% on the platform; we also include attention and manipulation checks.

All subjects receive a flat £3 participation fee (around $4 or €3.50 at the time of the experiment). In addition, the sliders allow subjects to earn up to an additional £60 (15 sliders × 4 contracts × £1 per slider) and the dictator game produces own payoffs up to £20 (£10 as giver, £10 as receiver). To be able 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. Thus, we compensate some participants with larger sums but a smaller likelihood (about 1 in 20), a procedure equivalent to paying smaller sums with certainty (Charness et al., 2016).[12]

To arrive at our final sample, we impose several restrictions. First, we require subjects to have placed at most 10 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 therefore correct. Second, subjects must not have failed both attention and manipulation checks. Third, we require

[12] 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.

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.[13]

4 Results

4.1 Descriptive statistics

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 a maximum of 40 and subjects donate on average £4.2 in the dictator game; about 24.4%, 10.5%, and 16% of subjects have previous experience working in a non-profit, in a social enterprise, or with a social organization, respectively. *Compassion* is positively correlated with most other 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, respectively, environmental good causes, with similar choices across treatments.

4.2 Intensive margin: Social enterprise effort allocation

Figure 1 plots average social enterprise (SE) *Social effort* across incentive levels and sorting conditions, together with 95% confidence intervals. In the absence of pay for performance, the effort allocation is skewed towards the social task, as subjects exert more than two thirds of their effort on this task (i.e. more than 10 out of 15 sliders), as predicted by our baseline Hypothesis 0. However, subjects allocate effort roughly equally between the commercial and social tasks

[13] To ensure transparency and commitment in our analysis, we preregistered our design and hypotheses within the Open Science Framework (link). Online Appendix B provides descriptive statistics, social preference correlations, an analysis of subjects' good cause choice, and a randomization check. We find that the good cause chosen is not systematically related to individual traits and that our randomization procedure was successful. Online Appendix C then 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. Online Appendix D shows that results are robust to tightening or relaxing the sample restrictions.

at all levels of pay for performance, with confidence intervals that include the level of social effort expected for a perfectly balanced effort allocation, plotted as a dashed line. Similarly, Panel A of Table 2 reports average *Social effort* across contracts, together with the number of subjects in each condition. Using t-tests, the first row of Panel B shows that more balanced SE effort allocations are elicited regardless of incentive strength or whether self-selection is possible ($p < 0.001$). The remaining rows of Panel B show that differences in SE *Social effort* between the £0.25, £0.50, and £1 treatments are not statistically significant. These results provide initial evidence for intensive margin effects, equivalent to an attention-directing role of incentives in social enterprises.

Table 2 also allows us to compare behavior in different organizational forms. For profit (FP) and non-profit (NP) workers exert 30% and, respectively, 90% of their effort on the social task in each treatment; SE social effort levels are in between and significantly different from FP and NP levels ($p < 0.001$ in Panel C). To quantify where SEs lie on the FP/NP continuum, we use t-tests to compare SE social effort with the average of FP and NP social effort in Panel D, where a positive difference indicates SE is closer to NP. Indeed, we find a positive difference in the £0 treatment and a negative one (though not always significant) for steeper incentives. These results suggest that SEs occupy the middle ground between FPs and NPs with regards to effort allocation when monetary incentives are in place and highlight the adverse specialization problem that social enterprises face when they do not use monetary rewards.

Interestingly, company descriptions matter beyond the incentives offered, even when both C and S tasks produce the same payoffs per slider. For instance, while the £0 bonus SE is equivalent to the NP in the contract terms offered, subjects exert significantly more *Social effort* in the NP contract relative to the SE contract (Table 2, Panel C, columns (1) and (5)). Similarly, whereas the £1 bonus SE is equivalent to the FP in contract terms, subjects exert significantly more *Social effort* in the SE contract relative to the FP contract (Table 2, Panel C, columns (4) and (8)). The fact that behavior is significantly different in these organizational forms indicates

that the stated organizational goal *per se* triggers shifts in the effort allocation.[14]

We can further probe the nature of adverse specialization: does imbalance result from a skewed effort allocation for all workers or only a minority of workers? Figure 2 plots the distribution of SE social effort in each treatment. Between 30% and 40% of subjects in the £0 treatment exert only social effort, whereas the distribution of social effort is roughly normal and centered around a fully balanced allocation for other subjects. Thus, adverse specialization only affects a fraction of workers, rather than shifting the entire distribution of social effort upwards. What could drive such behavior? Our theoretical framework suggests that subjects with higher other-regarding preferences are more likely to exert more social effort and, therefore, 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 select into the SE contract. The remaining panels of Figure 2 suggest that incentives, both modest and strong, successfully reduce the fraction of SE workers exerting only social effort. This results in significantly different distributions, centered more tightly around a fully balanced effort allocation.

As Hypothesis 1 predicts, modest incentives induce a more balanced effort allocation between the commercial and social tasks, alleviating the adverse specialization that occurs in the absence of monetary rewards. Surprisingly, contrary to Hypothesis 2 and a common view in social entrepreneurship, 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 turn to analyzing the relationship 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.

[14] Moreover, with a £0, £0.25, or £0.50 SE bonus, subjects would be better off financially by choosing the FP contract, as commercial effort is better remunerated in the FP and social effort produces the same good cause payoff (the social action in an FP could correspond, for instance, to corporate social responsibility (CSR) activities, see, e.g., Kitzmueller and Shimshack, 2012). Subjects would also be better off choosing the FP over the NP: they could choose to allocate, for example, 13 units of their effort to the social task, while still reaping personal rewards from 2 units of commercial effort in the FP contract (whereas the NP contract does not reward commercial action). Nonetheless, individuals choose organizations with a stated social mission over organizations without one, suggesting that labels matter.

4.3 Extensive margin: Worker self-selection

One of the reasons social entrepreneurs are reluctant to use pay for performance relates to their potential attraction of less socially motivated individuals. To assess the validity of this concern, Table 3 displays the average levels of *Compassion* and *Social motivation* across treatments.[15] In Panel A, we find no significant differences in mean social preferences across treatments in Part 2 of the experiment (Kruskal-Wallis $p > 0.1$), suggesting our randomization was successful. Panel B displays social preferences separately for subjects selecting into the FP, SE, and NP contracts in Part 3 of the experiment. SE social preferences are similar for the £0, £0.25, and £0.50 treatments, but are significantly lower in the £1 treatment, as the t-tests in Panel C show ($p < 0.05$ relative to the £0 treatment).[16] This shift is consistent with our prediction that more extrinsically motivated employees are attracted to SEs when incentives in this organization are stronger. Nonetheless, we find that more workers join the SE as incentives become stronger: while a £0 SE bonus attracts 28.8% of subjects (i.e. 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 FPs.

We investigate extensive margin effects further by comparing social preferences across Part 3 contract choices in Panel D of Table 3. As already visible in Panel B, outside of the £1 treatment, self-selected SE workers' social motivation is higher than that of FP workers ($p < 0.05$) and indistinguishable from that of NP workers ($p > 0.1$), supporting our central contention that the SE organizational form attracts highly motivated employees. However, SE workers' *Social motivation* in the £1 treatment is lower than that of NP workers ($p = 0.014$) and similar to that of FP workers ($p > 0.05$). As noted above, this is driven by a large shift in subjects choosing SE relative to FP when SE and FP contract terms are the same. In Panel E we inquire once again whether SEs are closer to FPs or NPs in social preferences by comparing the former with the

[15] For ease of interpretation, we standardize *Compassion* to have a mean of 0 and a standard deviation of 1.

[16] Equality of variances F-tests do not indicate a wider dispersion of social preferences across treatments, but Kolmogorov-Smirnov tests for equality of distributions suggest a different distribution of *Social motivation* in the £1 treatment ($p < 0.1$). In Online Appendix E, we show that with a £1 bonus, the SE attracts more (fewer) individuals from the bottom (top) 25% of the *Social motivation* distribution, shifting the distribution downwards.

mean of the latter. While SEs are indistinguishable from this average in terms of *Compassion*, they are significantly closer to NPs with regards to *Social motivation* when the bonus is £0, £0.25, or £0.50. Thus, the scope for extensive margin effects appears limited, with only strong incentives leading to a small, but significant decrease in SE employee social motivation.

As Figure 1 shows, allowing for employee self-selection does not alter the relationship between monetary incentives and *Social effort*, although self-selection leads to an effort allocation that appears closer to full balance. To formally test for differences across conditions, columns (1) and (2) of Table 4 regress SE *Social effort* on dummies for incentive levels, the sorting condition, and their interactions, with and without demographic and good cause controls. As some participants perform the SE contract twice, we cluster standard errors at the individual level. If the extensive margin is unimportant, we would expect the *Sorting* variable and its interaction with each treatment level to be jointly insignificant. The first three rows replicate the results in columns (1) to (4) in Panel A of Table 2, showing that the intensive margin effects are both significant and substantial. When looking at the next four rows, where we add the main and interaction sorting effects, we find some evidence of additional influence on effort allocation, but mainly in the £0.50 treatment. The p-values for tests of joint significance, reported at the bottom of Table 4, 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 intensive margin effects of monetary incentives on effort allocation.

Columns (3)-(6) of Table 4 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 C or S 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, our results suggest that incentives operate similarly with and without worker self-selection. Modest incentives have no extensive margin effects and strong incentives may engender a loss of employee social motivation, but do not skew employee effort towards the commercial task, contrary to our hypotheses. In the social enterprise context, monetary rewards work mainly at the intensive, rather than the extensive margin, directing employee attention towards the remunerated commercial task in a way that generates a balanced effort allocation in absolute terms, as well as relative to other organizational forms.

Sensitivity and heterogeneity Our results are robust to a wide set of sensitivity analyses. These include alternative sampling restrictions (Online Appendix D), social preference measures (Online Appendix E), effort measures (Online Appendix F), and multiple hypothesis testing adjustments, reflecting the number of outcomes and treatments we consider (Online 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 men and women in the effect of incentives on effort allocation (Online Appendix G). Our results paint a clear picture: modest incentives induce a balanced effort allocation without an associated reduction in social motivation.

5 Discussion

To achieve their economic and social value creation goals, social enterprises must allocate scarce employee effort between commercial and social tasks. Owing to their embeddedness in a social logic espousing an emphasis on social impact, a reluctance to employ practices stemming from

a commercial logic, and high levels of other-regarding preferences among both founders and employees, social enterprises often pay insufficient attention to revenue generation. As purpose takes priority over profits, revenue drift hinders social enterprises' ability to deliver on their social mission and threatens their survival. Why then do so few social enterprises adopt pay for performance practices, and why does the scholarly literature pay relatively little attention to incentives in social enterprises?

We argue that in the absence of pecuniary rewards, social enterprise employees allocate most of their effort to social tasks due to their high social motivation and perceived social mission saliency. We hypothesize that monetary incentives elicit a balanced effort allocation by directing employee effort to the commercial task and by potentially attracting workers with lower levels of social motivation who are less prone to adverse specialization. While we expect modest incentives to operate mainly at the intensive margin by increasing the benefits of exerting commercial effort, strong incentives may work on the extensive margin by attracting less motivated employees who could cause mission drift. Our experiment examines the effect of incentives on effort allocation in social enterprises, identifying the mechanisms through which this effect propagates. Both modest and strong incentives are found to produce a more balanced effort allocation, with employee effort split roughly equally between a commercial and a social task. While modest incentives do not affect social enterprise workforce composition, strong incentives do lead to a small, but significant downwards shift in the distribution of workers' social motivation.

Theoretical implications The tension between social and economic value creation in social enterprises and its implications for firm performance are core questions in the hybrid organizations literature (Pache and Santos, 2010; Dacin et al., 2011; Smith et al., 2013; Battilana et al., 2015; Smith and Besharov, 2019). While the risk of mission drift engendered by the pursuit of commercial goals in social enterprises has received substantial attention (Ebrahim et al., 2014; Ramus and Vaccaro, 2017; Grimes et al., 2018), few studies address revenue drift beyond noting

threats to economic performance when firms put purpose ahead of profits (Tracey et al., 2011; Battilana et al., 2015; Stevens et al., 2015; Staessens et al., 2018). Moreover, the usual incentive tools used by commercial ventures to guide employee effort are often viewed as inappropriate for social enterprises, due to their perceived incongruence with social impact and their potential to attract less motivated workers (Bacchiega and Borzaga, 2001; Austin et al., 2006; Tracey et al., 2011; Smith et al., 2013).

By applying incentive theory to social enterprises, we argue instead that monetary rewards can stave off revenue drift in this type of hybrid organization. Whereas existing studies show that firms often use social missions to elicit higher employee effort (Besley and Ghatak, 2005; Henderson and Van den Steen, 2015; Burbano, 2016; Cassar, 2019), the nature of that effort has been largely overlooked (Jones et al., 2018). Social enterprises relying solely on social mission to attract and motivate employees are vulnerable to revenue drift, which may create obstacles to growth and survival. By contrast, a combination of mission and monetary incentives succeeds in making workers balance their effort between commercial and social tasks. In other words, pay for performance mechanisms stemming from the commercial logic can be selectively coupled with practices prescribed by the social logic (Pache and Santos, 2013).

Inasmuch as it engenders a deviation from social enterprises' core focus on social mission, an outcome of balance may still be considered a source of mission drift. However, we argue that balance will afford social enterprises the resources needed for growth, allowing them to achieve their social impact more reliably. Our results therefore reinforce the need to move away from viewing mission drift as unequivocally bad and towards appreciating the situations where it may be necessary (Grimes et al., 2018).

The experimental design developed in this article causally isolates the channels through which incentives affect effort allocation. On the one hand, monetary rewards signpost to employees what tasks are valuable for the organization; hence, they perform a normative function (Kaplan and Henderson, 2005; Ethiraj and Levinthal, 2009; Wolfolds, 2018) and are integral to firms'

structural distribution of attention (Ocasio, 1997). On the other hand, incentives also perform a sorting function, potentially attracting employees with different pro-social motivations (Lazear, 2000; Cadsby et al., 2007). This can help achieve task effort balance if social motivation crowd-out is small, but may backfire if self-selected workers are extrinsically motivated and exert most of their effort on the incentivized commercial task (Deserranno, 2019). The evidence from our experiment is consistent with an intensive margin, attention-directing role of monetary rewards, with modest incentives being sufficient to balance the effort allocation without reducing employee social motivation. By conveying the importance of a given task to all employees, modest incentives may act as a coordination device and could be especially valuable either when firms face largely conflicting goals (Kogut and Zander, 1996) or when a dominant logic prescribes and legitimates non-optimal practices (Lounsbury, 2007) – both conditions that characterize social enterprises.

Practical implications Facing growing competitive pressures, social enterprises are increasingly turning their attention to operational sustainability. Monetary rewards offer one way for social enterprises to overcome an excessive focus on social impact to the detriment of revenue generation. Modest incentives are particularly appealing, as they expand the potential employee pool and restore balance with minimal cost. Modest incentives are also unlikely to reduce employee social motivation and increase risk taking or myopic behavior, and may thus be expected to have minimal adverse effects on organizational activities outside the scope of our study. Rather than attracting less motivated workers or over-emphasizing the relevance of commercial tasks, modest incentives signal the value of certain actions to employees. Consequently, social entrepreneurs need not worry too much about the pool of potential applicants but devote attention instead to internal effort allocations and compensation design.

As our results show, a social enterprise that pays employees larger bonuses can attract more candidates, such that incentives may aid social enterprises in their search for talent. Moreover,

social entrepreneurs who use performance-based pay may gain an advantage by reassuring potential investors of their commitment to commercial performance, thereby increasing the latter's willingness to invest. Furthermore, redirecting a fraction of employees' attention towards revenue generation can improve social enterprises' cash flow and thus reduce their dependence on outside finance; and by using the revenue generated to make operational investments, social enterprises may improve their growth and survival prospects. Of course, monetary rewards for commercial performance are but one tool available to social entrepreneurs. Additional means of guiding employee effort include governance mechanisms (Ebrahim et al., 2014), hiring and socialization practices (Battilana and Dorado, 2010), and (where quantifiable) rewards for achieving social impact and operational targets (Wolfolds, 2018). Assessing the comparative efficacy and complementarity of such practices with pay for performance mechanisms presents an interesting avenue for future research, with potentially important implications for social enterprise performance.

Limitations Reflecting our interest in how high-powered incentives affect effort allocation, we studied a wide range of incentive strengths in social enterprises, from as weak monetary incentives as in non-profits, to as strong monetary incentives as in for profits. Conceptually, we thus covered the entire range of relevant incentive strengths. Nevertheless, it might be that our choice of experimental parameters still limited this range. In particular, the good cause payoff was set to always be £1 and the monetary bonus was set to be at most £1 per slider, as well. The relatively high return to the social task was chosen to ensure that this task was on a level playing field with the commercial task; but one may be legitimately worried that the monetary incentives used in our experiment could be perceived as overly weak. A wider range of bonus levels, exceeding the good cause payoff, could potentially lead to different findings for strong incentives and explain why we failed to find evidence for hypotheses 2 and 4. However, this should not affect our main conclusion that modest incentives seem to help social enterprises rather than hurt them.

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. One could thus be concerned with the external validity of our experiment, given the multitude of tasks social enterprise employees may perform, the number of dimensions on which employees may differ, as well as the nuanced presentation of different types of firms' objectives in real life. Nonetheless, we believe our experiment captures the fundamental tension between commercial and social tasks in social enterprises and employee effort allocation in a simple and transparent manner. While we attempted to alleviate external validity concerns by recruiting participants from a representative pool of potential employees and giving them realistic choices of company types and missions, field replications represent a natural and necessary extension of our study.

Future research Our experiment raises several questions that future research should address. First, social enterprise effectiveness may depend on coordinated action and team performance. If so, are team incentives better than individual incentives? Does their effectiveness vary with employee heterogeneity? These questions need to be addressed in future research. Second, the introduction of pay for performance is often accompanied by a justification, helping the practice gain legitimacy. Does the communication of incentives matter for their success? How does goal clarity improve that communication? How do social enterprises justify and implement an incentive change? Again, these questions deserve further exploration in subsequent work. Third, the presence of monetary incentives in social enterprises is likely to affect a broad range of institutional referents, beyond employees. For example, it is unclear how targeted communities perceive social enterprises that adopt monetary rewards. Are they more likely to be perceived as professional or are they downgraded for possibly courting mission drift?

Fourth, financial incentives may have other, potentially negative effects on aspects of hybrid organizations that our work has not considered. For instance, while we find that self-selected employees' inequality aversion is not affected by incentive strength (see Online Appendix E),

pay for commercial performance may reveal significant productivity differences and engender social comparisons, with potentially adverse effects. One can then ask: do such incentives create internal conflicts between employees with different social preferences? Do monetary rewards generate long run distortions? What are joiners' commitment and retention rates? For social enterprises to achieve their promise, it is crucial to answer these questions, in both stylized experimental settings and in the field. Fifth, social enterprises exhibit substantial heterogeneity in their ability to measure social impact, which may affect how pay for performance influences effort allocation. Analyses of the relationship between the extent to which social impact is measurable and the potential for revenue and mission drift in a multitasking framework thus represent additional research avenues that may prove valuable for understanding the tensions and challenges social enterprises face.

Finally, the nature of the experiment constrains our ability to describe social enterprises in detail; nonetheless, even when exactly the same available actions and reward structure (for both own and good cause payoffs) exist in social enterprises as in for-profits or non-profits, rather surprisingly, behavior differs markedly across organizational forms. Despite our avoidance of the explicit 'social enterprise' label, subjects clearly distinguished this organizational form from the alternatives, as evidenced by their effort choices. This result highlights the importance of future research on the perception of social enterprises as a potentially distinct category, whose prescriptions for action employees and wider audiences may ultimately internalize, thereby providing additional legitimacy for this organizational form.

Conclusion We have argued theoretically and provided experimental evidence that social enterprises can address revenue drift – an excessive focus on social impact at the expense of generating revenue – by deploying monetary rewards. Modest incentives redirect employee attention to commercial tasks and reinforce social enterprises' commitment to achieving their social mission via market-based mechanisms, without attracting less socially employees. Our

findings question a common view in social entrepreneurship that incentives are incongruent with social impact and/or attract the 'wrong' kind of employee. This study contributes to a growing literature on hybrid organizations' challenges of managing competing logics and opens up a whole vista of interesting questions regarding social enterprises' 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.

Andersson, O., Huysentruyt, M., Miettinen, T., and Stephan, U. (2017). Person–Organization Fit and Incentives: A Causal Test. *Management Science*, 63(1):73–96.

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.

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., 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.

Becchetti, L., Castriota, S., and Tortia, E. C. (2013). Productivity, Wages, and Intrinsic Motivations. *Small Business Economics*, 41(2):379–399.

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.

Benabou, R. and Tirole, J. (2016). Bonus Culture: Competitive Pay, Screening, and Multitasking. *Journal of Political Economy*, 124(2):305–370.

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.

Bitektine, A., Lucas, J. W., and Schilke, O. (2018). Institutions Under a Microscope: Experimental Methods in Institutional Theory. In Bryman, A. and Buchanan, D. A., editors, *Unconventional Methodology in Organization and Management Research*, pages 147–167. Oxford University Press, Oxford.

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.

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.

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.

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.

Croson, 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.

Davies, I. A. and Doherty, B. (2018). Balancing a Hybrid Business Model: The Search for Equilibrium at Cafédirect. *Journal of Business Ethics*, forthcoming.

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.

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.

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.

Eldar, O. (2017). The Role of Social Enterprise and Hybrid Organizations. *Columbia Business Law Review*, 92(1):78.

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. (2016). The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences. *Working Paper*.

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.

Galizzi, M. M. and Navarro-Martinez, D. (2018). On the External Validity of Social Preference Games: A Systematic Lab-Field Study. *Management Science*, forthcoming.

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.

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.

Grimes, M., 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., Williams, T. A., and Zhao, E. Y. (2018). Anchors Aweigh: The Sources, Variety, and Challenges of Mission Drift. *Academy of Management Review*, forthcoming.

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.

Hockerts, K. (2015). How Hybrid Organizations Turn Antagonistic Assets into Complementarities. *California Management Review*, 57(3):83–106.

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.

Hwang, H. and Powell, W. W. (2009). The Rationalization of Charity: The Influences of Professionalism in the Nonprofit Sector. *Administrative Science Quarterly*, 54(2):268–298.

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.

Kitzmueller, M. and Shimshack, J. (2012). Economic Perspectives on Corporate Social Responsibility. *Journal of Economic Literature*, 50(1):51–84.

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., 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.

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. (2018). Multiple Hypothesis Testing in Experimental Economics. *Experimental Economics*, forthcoming.

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 Entrepreneuring Models. *Journal of Business Ethics*, 111(3):353–373.

Margolis, J. D. and Walsh, J. P. (2003). Misery Loves Companies: Rethinking Social Initiatives. *Administrative Science Quarterly*, 48(2):268–305.

Marquis, C. and Park, A. (2014). Inside the Buy-One Give-One Model. *Stanford Social Innovation Review*, 12(1):28–33.

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.

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*.

Ocasio, W. (1997). Towards an Attention-Based View of the Firm. *Strategic Management Journal*, 18:187–206.

Osterloh, M. and Frey, B. S. (2000). Motivation, Knowledge Transfer, and Organizational Forms. *Organization Science*, 11(5):538–550.

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.

Pacheco, D. F., York, J. G., and Hargrave, T. J. (2014). The Coevolution of Industries, Social Movements, and Institutions: Wind Power in the United States. *Organization Science*, 25(6):1609–1632.

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.

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.

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.

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*, forthcoming.

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.

Staessens, M., Kerstens, P. J., Bruneel, J., and Cherchye, L. (2018). Data Envelopment Analysis and Social Enterprises: Analyzing Performance, Strategic Orientation, and Mission Drift. *Journal of Business Ethics*, forthcoming.

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.

Wolfolds, 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.

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

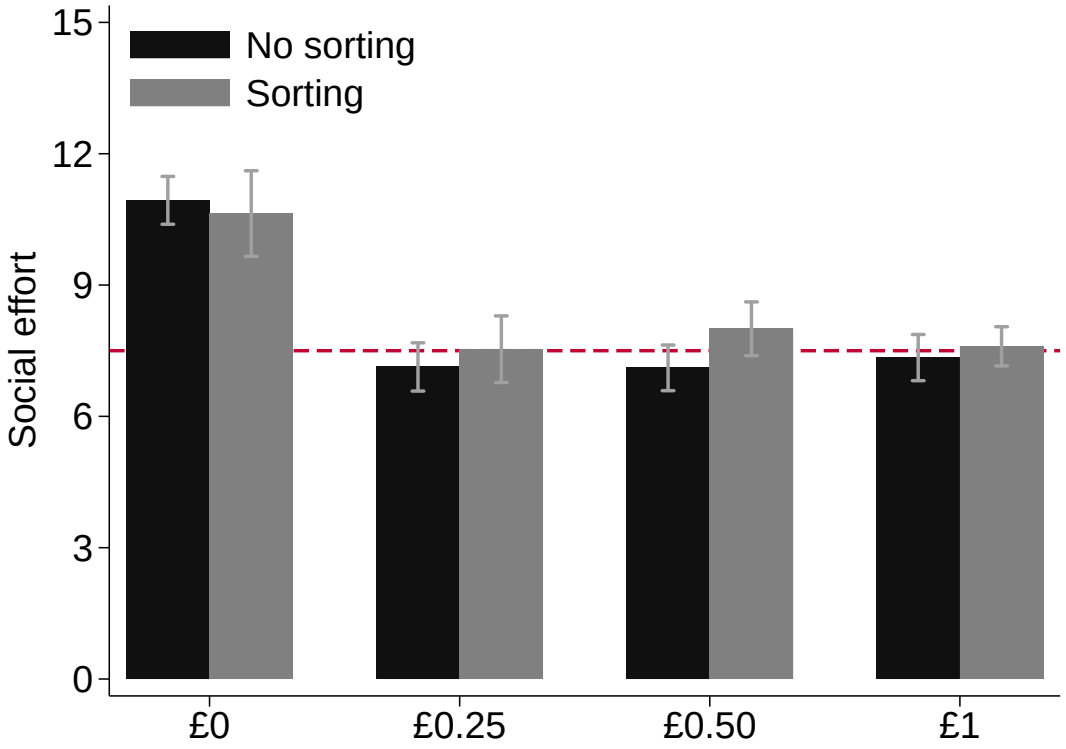

Figure 1: Social effort in social enterprises by treatment and sorting condition, with 95% confidence intervals; the dashed line represents a fully balanced effort allocation.

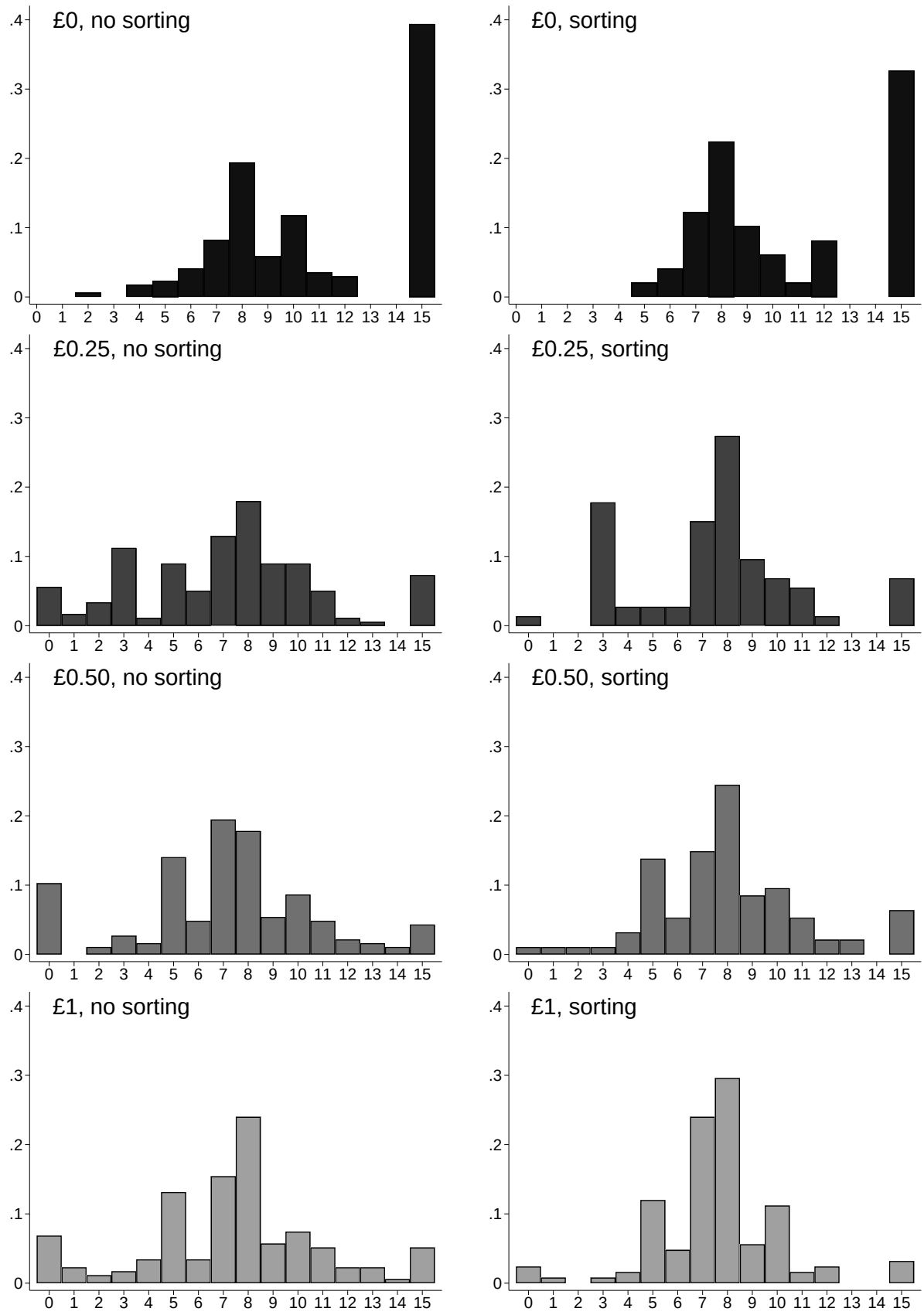

Figure 2: Distribution of social effort in social enterprises. The £0 distribution is different from the £0.25, £0.50, and £1 distributions (Kolmogorov-Smirnov $p < 0.001$, no sorting; $p < 0.01$, sorting). The latter distributions are not different from each other ($p > 0.1$).

Table 1: **Experimental Design**

	Actions and measures	Own payoff (C task)	Good cause payoff (S task)
Part 1	Choose good cause		
	Practice slider task		
	Comprehension check		
Part 2	FP contract	£1	£1
	NP contract	£0	£1
	SE contract	£0/£0.25/£0.50/£1	£1
Part 3	Choose preferred contract		
	FP contract (if selected)	£1	£1
	NP contract (if selected)	£0	£1
	SE contract (if selected)	£0/£0.25/£0.50/£1	£1
Part 4	Demographics		
	Social preferences		
	Compassion: sub-scale of public service motivation scale		
	Altruism: incentivized £10 dictator (giver)		
	Inequality aversion: hypothetical £10 ultimatum (receiver)		
	Hypothetical altruism: hypothetical £1,000 lottery		
	Willingness to share: without expecting anything in return		
	Prosocial behavior: observed prosocial behavior		
	Risk and time preferences		
	Attention check		
	Manipulation check		

Recruitment: Prolific Academic, a 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, ages 18-64, active labor force, prior approval rate > 90%
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 (link)

We randomize the order in which contracts are displayed in Part 2 and the order in which the options are presented in Part 3. The choice of good cause determines the charity that the good cause payoffs will be donated to, but also the organizational mission that the fictional (social) enterprises will pursue, 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 exposed to only one of these levels throughout. In addition, subjects never see the labels used (FP, SE, NP), and only infer the type of company. We restrict subjects' prior approval rate on the platform to be higher than 90% to ensure high-quality answers.

Table 2: **Social Effort, by Contract and Treatment**

	No sorting				Sorting			
	£0	£0.25	£0.50	£1	£0	£0.25	£0.50	£1
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Social effort levels across contracts								
FP	4.918	5.152	5.227	5.640	3.956	4.131	4.100	4.846
	(4.506)	(4.479)	(4.435)	(4.613)	(3.836)	(3.735)	(3.672)	(4.846)
N	170	178	185	175	90	84	70	26
SE	10.935	7.129	7.108	7.343	10.633	7.534	8.000	7.600
	(3.612)	(3.727)	(3.595)	(3.534)	(3.408)	(3.262)	(3.002)	(2.527)
N	170	178	185	175	49	73	94	125
NP	13.306	12.719	12.984	12.697	13.355	13.762	13.367	13.458
	(2.939)	(3.169)	(3.303)	(3.503)	(2.537)	(2.364)	(3.851)	(3.413)
N	170	178	185	175	31	21	21	24
B. SE effort *t*-tests of equality of means across treatments, *p*-values								
vs £0		0.000	0.000	0.000		0.000	0.000	0.000
vs £0.25			0.956	0.581			0.340	0.874
vs £0.50				0.533				0.286
C. SE effort *t*-tests of equality with FP and NP, *p*-values								
vs FP	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
vs NP	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
D. SE effort *t*-tests of equality with FP and NP average, sign and *p*-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 SE social effort with the average of FP and NP social effort to show where the SE lies on the continuum between FP and NP: '+' means SE is closer to NP than to FP, and '−' means SE is closer to FP than to NP. All p-values are two-sided.

Table 3: **Motivation and Contract Choice**

	Compassion (standardized)				Social motivation (factor)			
	£0	£0.25	£0.50	£1	£0	£0.25	£0.50	£1
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
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)
N	170	178	185	175	170	178	185	175
B. Motivation across contracts, with sorting								
FP	-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)
N	90	84	70	26	90	84	70	26
SE	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)
N	49	73	94	125	49	73	94	125
NP	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)
N	31	21	21	24	31	21	21	24
C. SE motivation *t*-tests of equality of means across treatments, *p*-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. SE motivation *t*-tests of equality with FP and NP, *p*-values								
vs FP	0.006	0.013	0.005	0.156	0.000	0.000	0.000	0.080
vs NP	0.854	0.547	0.449	0.745	0.782	0.777	0.736	0.014
E. SE motivation *t*-tests of equality with NP and FP average, sign and *p*-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 SE motivation with the average of FP and NP motivation to show where the SE lies on the continuum between FP and NP: '+' means SE is closer to NP than FP, and '−' means SE is closer to FP than NP. All *p*-values are two-sided.

Table 4: **Intensive and Extensive Margin Effects on SE Effort Allocation**

	Social effort units		Social effort share		Absolute balance	
	(1)	(2)	(3)	(4)	(5)	(6)
£0.25	-3.806***	-3.934***	-0.252***	-0.260***	-2.197***	-2.389***
	(0.394)	(0.397)	(0.026)	(0.026)	(0.587)	(0.587)
£0.50	-3.827***	-3.934***	-0.254***	-0.261***	-2.570***	-2.823***
	(0.383)	(0.389)	(0.026)	(0.026)	(0.589)	(0.591)
£1	-3.592***	-3.670***	-0.237***	-0.241***	-2.717***	-2.885***
	(0.385)	(0.392)	(0.026)	(0.026)	(0.591)	(0.600)
Sorting	-0.303	-0.240	-0.020	-0.017	-0.935	-0.856
	(0.471)	(0.464)	(0.031)	(0.031)	(0.838)	(0.822)
Sorting × £0.25	0.708	0.639	0.047	0.043	0.005	0.177
	(0.588)	(0.586)	(0.039)	(0.039)	(0.972)	(0.948)
Sorting × £0.50	1.195**	1.128**	0.081**	0.077**	0.047	0.005
	(0.577)	(0.572)	(0.038)	(0.038)	(0.930)	(0.913)
Sorting × £1	0.560	0.489	0.034	0.030	-0.778	-0.837
	(0.537)	(0.531)	(0.036)	(0.035)	(0.904)	(0.886)
Constant	10.935***	11.545***	0.729***	0.770***	7.894***	7.899***
	(0.277)	(0.631)	(0.018)	(0.042)	(0.467)	(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

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered at the subject level in parentheses. The baseline category is the £0 treatment when sorting is not possible. The dependent variable is *Social effort*, i.e. S task effort, in columns (1) and (2), social effort as a share of total effort, i.e. $S/(S+C)$, in columns (3) and (4), and absolute balance, i.e. $|S-C|$, in columns (5) and (6). Controls include age, gender, studentship, education, income, risk and time preferences, and choice of good cause.

ONLINE APPENDIX for "Striking a Balance: Revenue Drift, Incentives, and Effort Allocation in Social Enterprises"

Theodor Vladasel, Copenhagen Business School
Simon C. Parker, Ivey Business School
Randolf Sloof, Universiteit van Amsterdam
Mirjam van Praag, Vrije Universiteit Amsterdam

This online appendix includes additional material for our paper entitled "Striking a Balance: 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 checks on our experimental design, including descriptive statistics, social preference correlations, an analysis of good cause choice, and a randomization check, as well as discusses the results of our pilot experiment. Appendix C details the results of power calculations performed prior to running the experiment, including minimum detectable effect sizes. Appendix D shows that results are robust to tightening or relaxing the sample restrictions, Appendix E examines robustness with regards to alternative social preference measures, and Appendix F considers a set of alternative effort measures and regression analyses. Finally, Appendix G discusses the results of multiple hypothesis testing adjustments, reflecting the number of outcomes and treatments we consider, as well as heterogeneity analyses with regards to gender, previous social sector experience, and mission choice.

Contents

Appendix A Instructions

Welcome Thank you for taking part in this study designed to learn about how people make decisions.[17] The choices we ask you to make are based on methods and techniques from economics, business administration, and psychology. The experiment is expected to take around 20 minutes. At the top of your screen, you will be able to see what percentage of the questionnaire you have already completed.

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

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

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

Figure A.1: Slider example.

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

[17] A preview can be accessed here.

Part 1 – Good cause and practice

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

Non profit 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 FP 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.

Non profit 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.

Non profit 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 Appendix Table D.1 shows our findings are robust to tightening or relaxing these restrictions. Table B.1 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 out of 40 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.2 also 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.3, 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 SE 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.

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 with multiple hypothesis testing adjustments (Romano and Wolf, 2005; List et al., 2018). 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 for all available observations ($N = 796$). Overall, our randomization has been successful.

Table B.1: **Descriptive Statistics**

		Mean	Std. dev.	*N*	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 B.2: **Correlation Table: Social Preferences**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) Compassion	1.0000							
(2) Altruism	0.1362	1.0000						
(3) Inequality aversion	-0.1353	0.0746	1.0000					
(4) Hypothetical altruism	0.2020	0.2650	0.0312	1.0000				
(5) Willingness to share	0.3994	0.1729	0.0050	0.3415	1.0000			
(6) Prosocial behavior	0.2337	0.1413	-0.0021	0.2570	0.3049	1.0000		
(7) Risk taking	0.0215	0.0515	0.0328	0.0880	0.1484	0.1108	1.0000	
(8) Time preferences	0.1190	0.0519	-0.0867	0.1530	0.3215	0.1403	0.1833	1.0000

Pairwise correlations between social preference measures. Compassion is measured through the compassion subscale of the public service motivation scale, altruism is measured through a dictator game, inequality aversion is measured through an ultimatum game, hypothetical altruism is measured as the donation in a hypothetical lottery, willingness to share is measured as subjects' willingness to share with other without expecting anything in return, and prosocial behavior is measured by combining questions on i) past non-profit employment, ii) past social enterprise employment, iii) volunteering, iv) donations, or v) professional relations with social organizations. Risk taking and time preferences are assessed with self-reflection questions on willingness to take risks and willingness to give something up today for a reward tomorrow. All correlations larger than 0.07 (in absolute terms) are significant at 5%.

Table B.3: **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.4: **Choice of Good Cause: Multinomial Logit**

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

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

Table B.5: **Randomization Check**

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

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses. Each row presents results from a separate model, regressing the given characteristic on treatment dummies, with p-values from tests of joint significance. The omitted education categories are completed high school and other degrees; the omitted income category comprises those who prefer not to answer. Estimating a series of seemingly unrelated regressions produces similar results. Multiple hypothesis testing adjustments (Romano and Wolf, 2005; List et al., 2018) 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:

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

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 less noisy our estimates, 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 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 (SE) and smallest group (FP/NP) 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 SE and NP 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

[20] There is no *ex ante* reason to expect different variance in social enterprise social effort across treatments when sorting is not allowed; therefore, we opted for equal samples across treatments (List et al., 2011). The notes to Table 2 show that, *ex post*, the assumption of equal variance is valid.

[21] The FP and NP groups are smaller, but their mean difference is expected to be large, so power is retained.

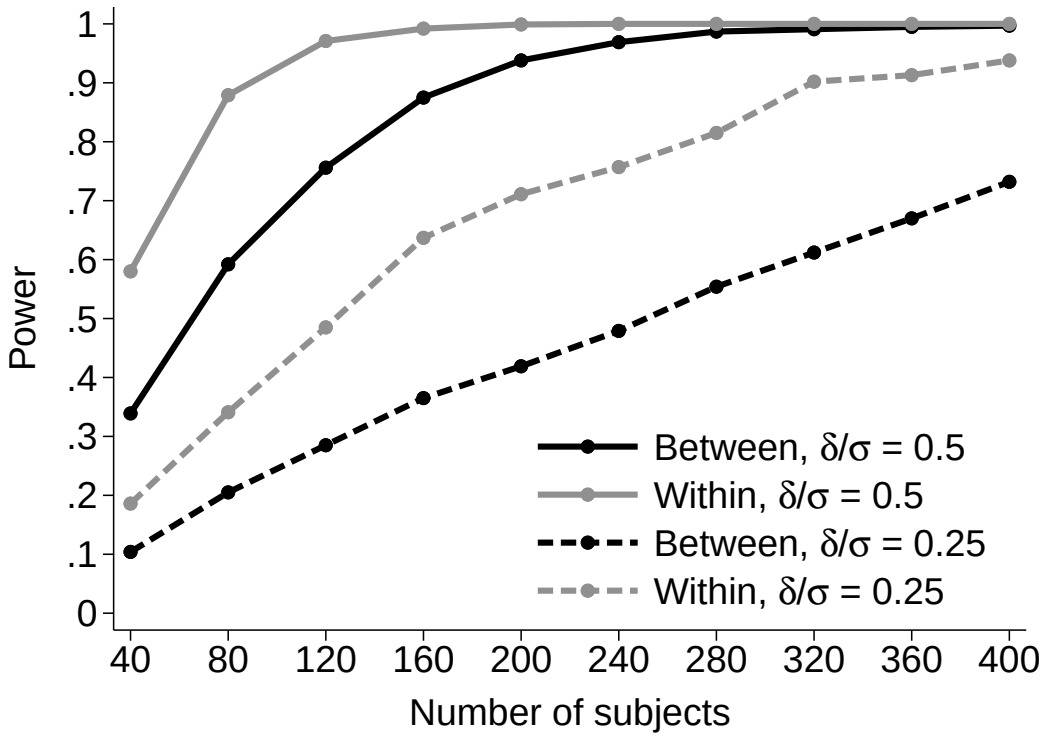

Figure C.1: Power levels for two-sided mean comparisons between- and within-individuals ($\alpha = 0.05, \sigma_a = \sigma_b = \sigma = 2, \delta = 1$ or $0.5, \delta/\sigma = 0.5$ or $0.25, H_0$: $\mu_a = \mu_b, H_a$: $\mu_a \neq \mu_b$).

the 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 fail to achieve 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., 2018). 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.

The conclusion of this section is that across a range of assumptions regarding sample sizes, means, standard deviation, and significance, our experiment is able to detect small changes in social effort – i.e. of at least half a standard deviations – even under the most conservative specifications. *Ex post*, it is important to note that while the standard errors 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] List et al. (2011) warn against performing power calculations with the actual data from the experiment, which is why we emphasize the *ex ante* calculations.

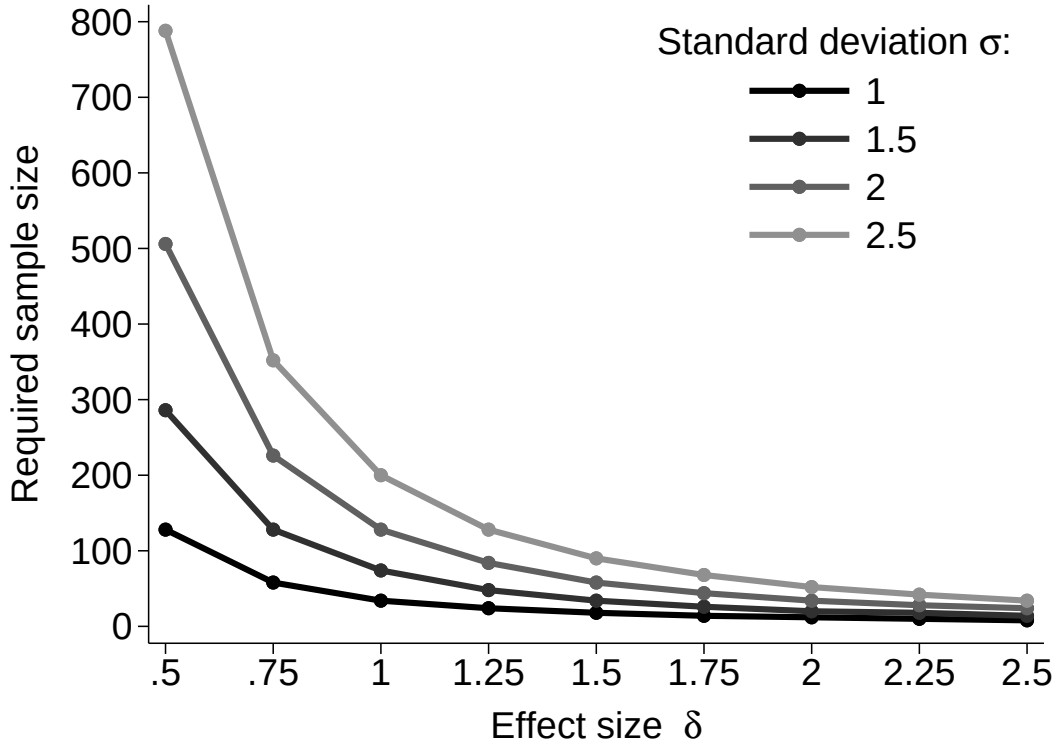

Figure C.2: Required sample size for detecting effect size δ with 80% power in two-sided mean comparison ($1-\beta=0.8, \alpha=0.05, \sigma_a=\sigma_b=\sigma, H_0$: $\mu_a=\mu_b, H_a$: $\mu_a\neq\mu_b$).

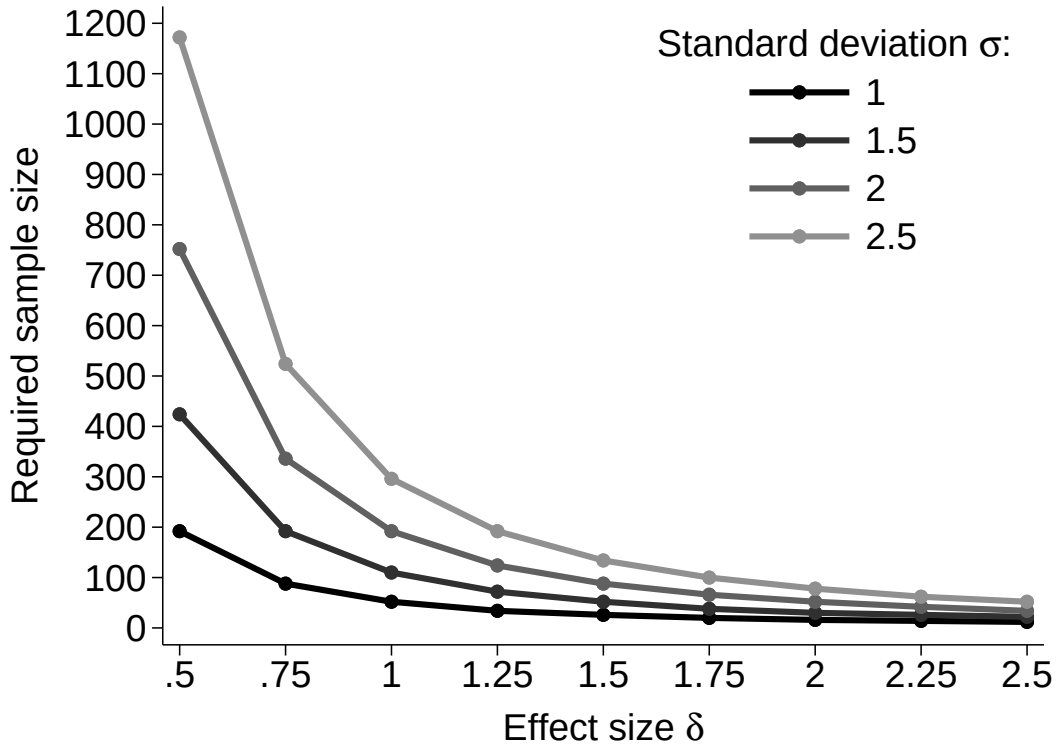

Figure C.3: Required sample size for detecting effect size δ with 80% power in two-sided mean comparison ($1-\beta=0.8, \alpha=0.01, \sigma_a=\sigma_b=\sigma, H_0$: $\mu_a=\mu_b, H_a$: $\mu_a\neq\mu_b$).

Table C.1: **Power Calculations**

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

Power calculations for mean comparisons. In panel A, the comparison is for 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 δ/σ in column (10) are 0.280, 0.524, and 0.616 in panels A, B, and, respectively, C.

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 subscale, 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 SE 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 SE 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 slider 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 (either FP, NP, or SE) 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 FP, NP, and respectively, SE contract. In addition, 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 SE contract, weighted by number of observations. Most data points lie along 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 SE 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 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 – summarized in Figure D.2 – display a similar pattern as the results we present in Figure 1: adverse specialization on the social task arises in the absence of pay for performance, while bonuses induce a more balanced effort allocation without reducing social motivation levels, regardless of the incentive steepness (although there is an elevated risk of mission drift 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 the 10 we used in the pilot experiment – in order to limit any scale similarity between the slider task and the dictator game.

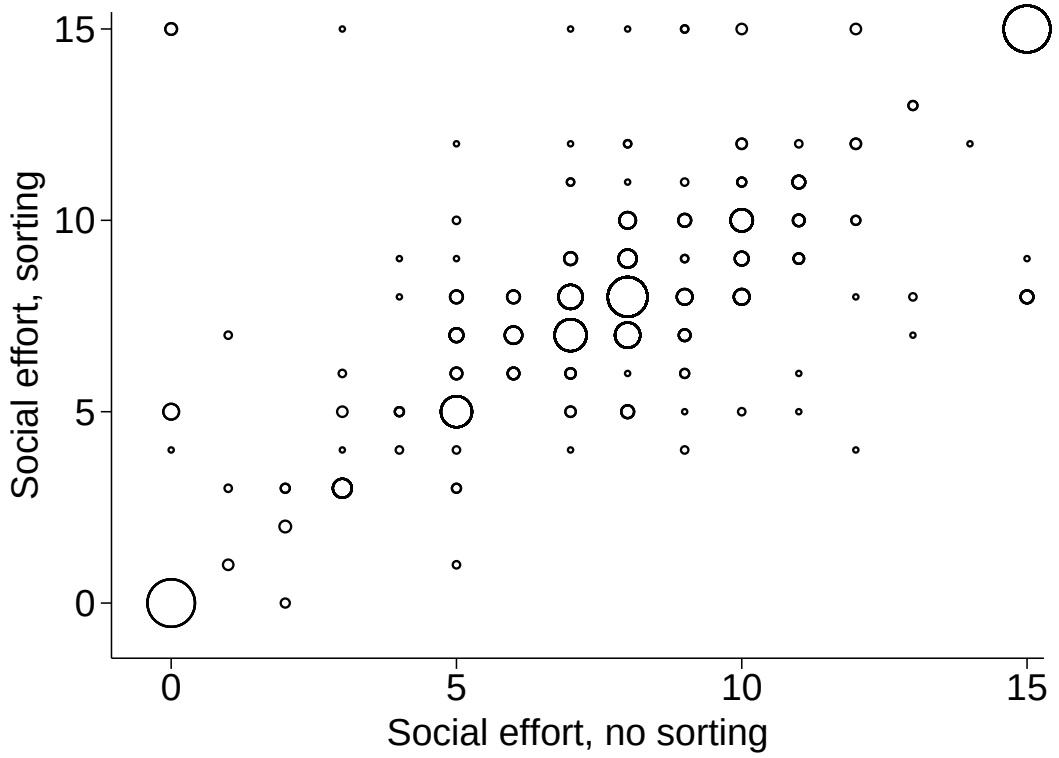

Figure D.1: Answer consistency across SE contracts with and without sorting, for subjects who performed the SE contract twice.

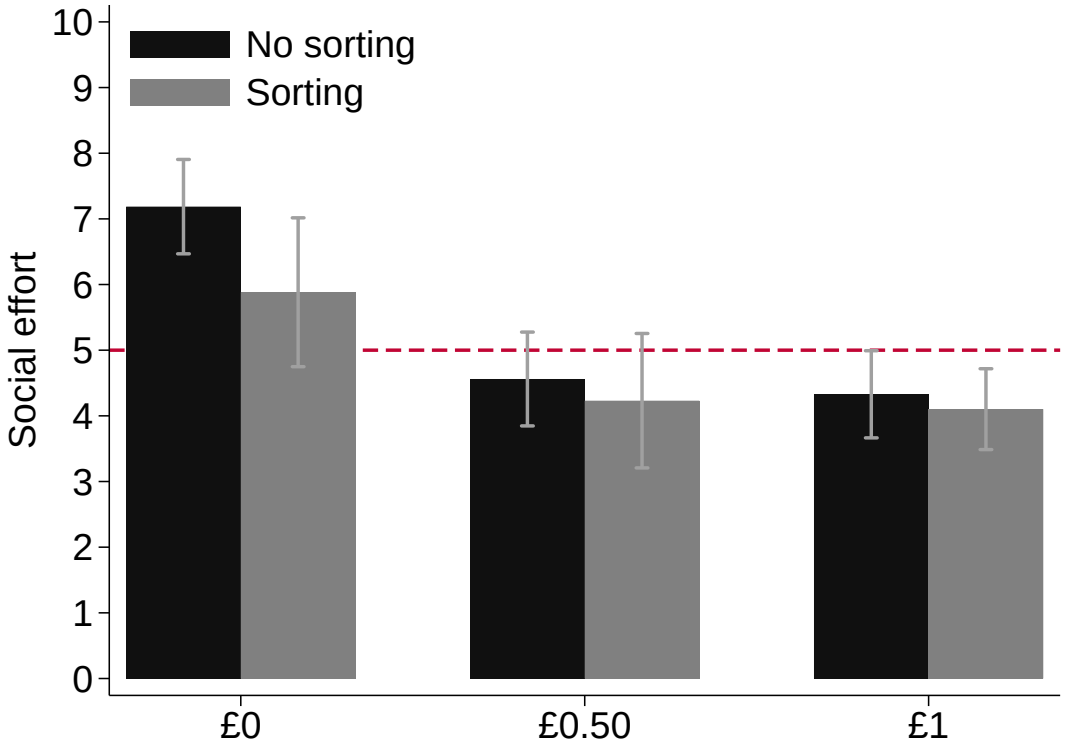

Figure D.2: Pilot data social effort in social enterprises by treatment, with 95% confidence intervals; the dashed line represents a fully balanced effort allocation.

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	7.099	7.088	7.338	10.350	7.390	7.951	7.504
	(3.599)	(3.672)	(3.762)	(3.741)	(3.416)	(3.140)	(3.054)	(2.572)
Share	0.733	0.473	0.472	0.489	0.690	0.492	0.530	0.500
	(0.239)	(0.244)	(0.250)	(0.249)	(0.227)	(0.209)	(0.203)	(0.171)
B. $NS = 0$ ($N = 623$, $N_{SE} = 307$)								
Units	10.913	7.238	7.141	7.463	10.302	7.405	7.940	7.459
	(3.676)	(3.703)	(3.776)	(3.774)	(3.447)	(3.025)	(3.019)	(2.620)
Share	0.727	0.482	0.476	0.497	0.686	0.493	0.529	0.497
	(0.245)	(0.246)	(0.251)	(0.251)	(0.229)	(0.201)	(0.201)	(0.174)
C. $NS \leq 10, F = F_p, 10 \leq t \leq 40$, pass attention check ($N = 686$, $N_{SE} = 332$)								
Units	10.928	7.088	7.165	7.323	10.687	7.362	8.000	7.590
	(3.612)	(3.760)	(3.587)	(3.526)	(3.421)	(3.180)	(3.018)	(2.557)
Share	0.728	0.474	0.478	0.490	0.712	0.491	0.535	0.506
	(0.240)	(0.249)	(0.239)	(0.234)	(0.228)	(0.210)	(0.200)	(0.170)
D. $IS \leq 10$, pass attention check ($N = 717$, $N_{SE} = 346$)								
Units	10.948	7.173	7.215	7.325	10.775	7.378	7.989	7.500
	(3.649)	(3.757)	(3.592)	(3.633)	(3.441)	(3.073)	(2.987)	(2.682)
Share	0.729	0.479	0.481	0.491	0.718	0.492	0.534	0.500
	(0.243)	(0.249)	(0.239)	(0.242)	(0.229)	(0.203)	(0.198)	(0.178)
E. $IS \leq 30, F = F_p, 10 \leq t \leq 40$, pass either check ($N = 722$, $N_{SE} = 349$)								
Units	10.794	7.088	7.112	7.222	10.274	7.426	7.978	7.500
	(3.714)	(3.743)	(3.659)	(3.599)	(3.800)	(3.333)	(2.992)	(2.632)
Share	0.727	0.480	0.474	0.484	0.706	0.496	0.539	0.503
	(0.239)	(0.250)	(0.243)	(0.239)	(0.225)	(0.222)	(0.204)	(0.170)
F. $IS \leq 30$ ($N = 767$, $N_{SE} = 372$)								
Units	10.700	7.149	7.158	7.272	10.321	7.566	7.959	7.407
	(3.827)	(3.775)	(3.635)	(3.664)	(3.785)	(3.298)	(2.949)	(2.735)
Share	0.724	0.488	0.478	0.487	0.709	0.505	0.538	0.497
	(0.242)	(0.254)	(0.242)	(0.244)	(0.226)	(0.219)	(0.201)	(0.177)
G. Full sample ($N = 796$, $N_{SE} = 388$)								
Units	10.239	6.984	6.901	7.040	9.419	7.388	7.774	7.208
	(4.296)	(3.868)	(3.790)	(3.802)	(4.550)	(3.457)	(3.101)	(2.935)
Share	0.717	0.493	0.479	0.489	0.717	0.505	0.547	0.500
	(0.252)	(0.258)	(0.246)	(0.245)	(0.247)	(0.247)	(0.209)	(0.182)

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

Revealed preference social motivation Social task effort in the FP 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 SE bonus is revealed to some subjects before they perform the FP 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 3, 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.2 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. Moreover, 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., 2018). Another way to analyze such shifts is to perform quantile regressions of social preference variables on treatment dummies. Table E.3 suggests that the *Social motivation* of individuals who select into the SE 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, 2018). 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.4. 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. (2018) multiple hypothesis testing adjustment renders all coefficients statistically

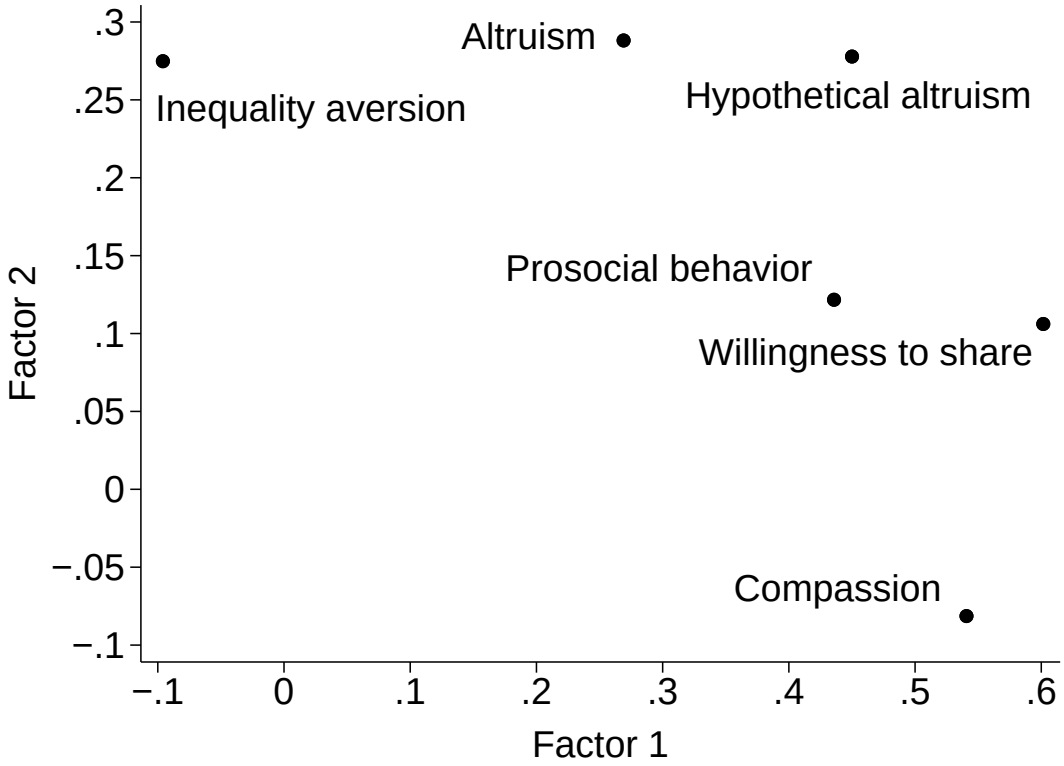

Figure E.1: Social preference loadings on the first two factors.

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 SE 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 SE bonus, we expect that other-regarding preferences are lowest for self-selected FP workers and highest for self-selected NP workers; SE workers are in between, with some differences across treatments. To see this, Table E.5 presents a regression analysis counterpart to the comparisons in Table 3, considering subjects make a single choice between the three contracts: SE motivation is different from FP motivation but not NP 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.6 shows that our expectation is met for SE and FP 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 SEs, although SE workers put more weight on the future relative to FP workers. While FP and NP workers are highly different in their social preferences, SE and NP 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 SEs are highly socially motivated, which may result in adverse specialization when pay for performance is not used.

[24] We have also checked that stronger incentives do not attract individuals with higher education, or with higher or lower income levels. They appear to attract individuals who took longer to complete the comprehension check, but multiple hypothesis testing adjustments eliminate the significant coefficients.

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	*Social motivation*		

Factor loadings for principal component analysis with an orthogonal varimax rotation; results are similar with oblique (non-orthogonal) rotations. As a 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 FP contract or use the 5 items that comprise *Prosocial behavior* individually.

Table E.2: **Social Preferences, Conditional on Social Enterprise Sorting**

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

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

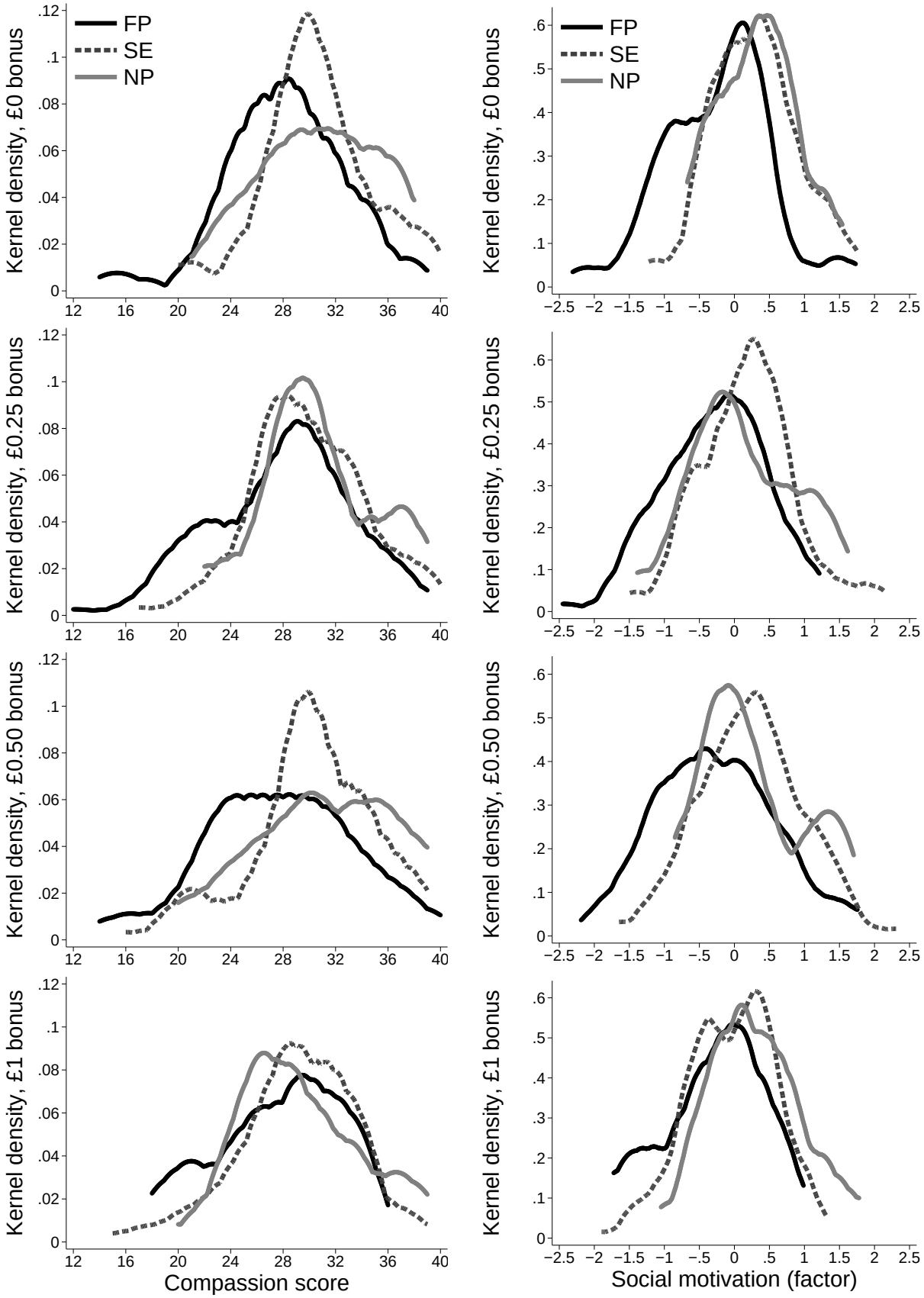

Figure E.2: Distribution of *Compassion* and *Social motivation*, by treatment and contract choice.

Table E.3: **Social Preferences in Social Enterprises: Quantile Regressions**

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

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. $N = 341$. Bootstrapped standard errors in parentheses (1000 replications). Columns present quantile regressions of *Compassion* (standardized) and *Social motivation* (factor) on treatment dummies for the 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} quantiles; the baseline category is the £0 treatment. Controls include age, gender, income, studentship, education, risk and time preferences, and choice of good cause.

Table E.4: **Social Preferences by Treatment, Conditional on Social Enterprise Sorting**

	Altruism (1)	Inequality aversion (2)	Hypothetical Altruism (3)	Willing to share (4)	Prosocial behavior (5)	Risk (6)	Time (7)
£0.25	-0.389	-0.316	14.214	-0.150	-0.291	-0.136	-0.076
	(0.308)	(0.384)	(29.871)	(0.375)	(0.208)	(0.437)	(0.378)
£0.50	-0.497*	-0.149	2.896	-0.039	-0.189	0.167	-0.109
	(0.294)	(0.366)	(28.500)	(0.358)	(0.198)	(0.417)	(0.361)
£1	-0.762***	-0.199	-43.925	-0.810**	-0.295	-0.043	-0.555
	(0.281)	(0.350)	(0.262)	(0.343)	(0.189)	(0.399)	(0.345)
Test of joint significance p-value:							
	0.051	0.870	0.051	0.014	0.426	0.862	0.217
R^2	0.023	0.002	0.023	0.031	0.008	0.002	0.013

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

Table E.5: **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
FP	-0.472***	-0.418***	-0.523***	-0.289	-0.543***	-0.494***	-0.509***	-0.248
	(0.163)	(0.159)	(0.164)	(0.209)	(0.124)	(0.119)	(0.128)	(0.159)
NP	0.002	0.109	0.186	0.076	0.020	-0.078	0.063	0.353**
	(0.211)	(0.223)	(0.262)	(0.214)	(0.143)	(0.188)	(0.179)	(0.150)
R^2	0.071	0.052	0.107	0.018	0.137	0.099	0.124	0.061
N	170	178	185	175	170	178	185	175

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors in parentheses. We regress *Compassion* or *Social Motivation* on subjects' choice of contract, using the SE motivation means in Panel B of Table 3 as a baseline.

Table E.6: **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)
FP	28.873***	3.716***	2.695***	96.148***	6.015***	1.307**	-0.136*	5.086***	6.209***
	(0.496)	(0.248)	(0.236)	(13.888)	(0.264)	(0.137)	(0.081)	(0.264)	(0.216)
SE	30.977***	4.822***	2.496***	143.172***	7.169***	1.534***	0.307***	5.176***	6.643***
	(0.522)	(0.237)	(0.247)	(18.353)	(0.262)	(0.149)	(0.083)	(0.280)	(0.230)
NP	31.514***	4.913***	2.685***	192.515***	7.371***	1.527***	0.427***	5.044***	6.607***
	(0.647)	(0.251)	(0.313)	(22.131)	(0.314)	(0.153)	(0.094)	(0.341)	(0.305)
Test of equality p-values:									
FP vs SE	0.000	0.000	0.312	0.001	0.000	0.031	0.000	0.675	0.020
FP vs NP	0.000	0.000	0.971	0.000	0.000	0.086	0.000	0.894	0.157
NP vs SE	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 2 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 SE. 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 FP, NP, and SE contracts. Comparisons across contracts must then adjust for the paired nature of the test. More specifically, the results in Table 2 suggest the SE contract is always different from the NP and FP 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 SE and NP contracts, using Part 2 observations only and the FP contract as a baseline. The results in Table F.2 confirm our results: social effort is higher in the SE and NP contracts relative to the FP 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*, although these results are not significant (potentially due to small sample size). However, these individuals have self-selected into the SE contract and are more motivated than those who self-selected into the FP contract, as per Table E.5. Figure F.1 displays a scatter plot of SE

[25] Note that we do not find differences in SE total effort across treatments, regardless of the sample we use.

[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 (see also Lazear et al., 2012).

Social effort in the £0 treatment and *Social motivation*. Those who self-select into the SE (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 (available upon request).

Measures of imbalance In our main analysis, we focus on units of *Social effort* as the most straightforward measure of effort allocation, and compare SE social effort with a fully balanced effort allocation (i.e. 7.5 units) and with FP and NP 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 ($S - C$) 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 ($|S - C|$). 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 4 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, SE 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 SE that uses no bonus. Table F.6 replicates the analysis in Table 4 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.

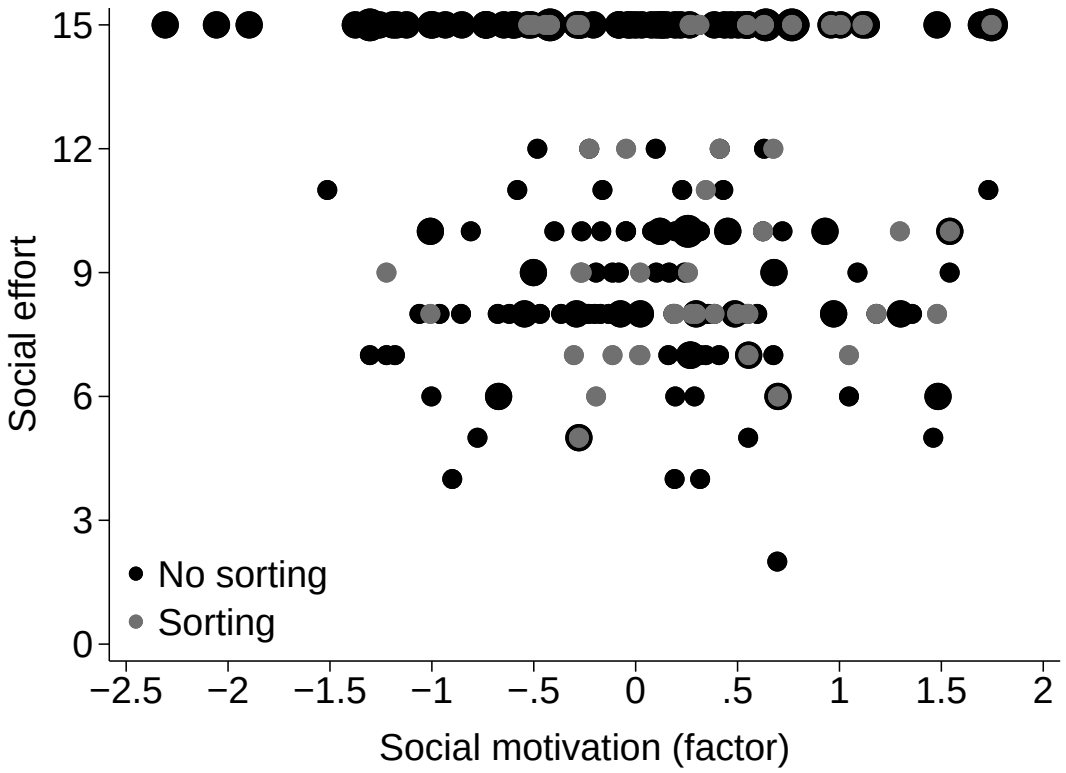

Figure F.1: Social effort in the SE contract, £0 treatment.

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

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

Standard errors in parentheses. Within each column the FP, NP, and SE 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 SE is closer to NP than FP, and '−' means SE is closer to FP than NP.

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				
SE	6.018***	1.978***	1.881***	1.703***
	(0.426)	(0.331)	(0.319)	(0.336)
NP	8.388***	7.567***	7.757***	7.057***
	(0.455)	(0.433)	(0.425)	(0.443)
B. Share of social effort				
SE	0.401***	0.133***	0.126***	0.116***
	(0.028)	(0.022)	(0.021)	(0.022)
NP	0.559***	0.510***	0.521***	0.478***
	(0.030)	(0.028)	(0.028)	(0.029)
Observations	510	534	555	525
Subjects	170	178	185	175

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses. The baseline category consists of social effort in the FP contract. Social effort is different in the SE and NP 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	0.037		0.235**	0.114*	
	(0.047)	(0.040)		(0.085)	(0.080)	
Compassion^2		0.095***			0.150***	
		(0.025)			(0.050)	
Compassion $\geq 90^{\text{th}}$ pct.			0.363***			0.946***
			(0.134)			(0.148)
R^2	0.093	0.154	0.136	0.579	0.657	0.740
N	170	170	170	49	49	49
B. Social motivation						
Social motivation	-0.186	-0.076		0.241	0.145	
	(0.057)	(0.056)		(0.176)	(0.183)	
Social motivation^2		0.085*			0.146	
		(0.047)			(0.163)	
Social motivation $\geq 90^{\text{th}}$ pct.			-0.021			0.573
			(0.139)			(0.372)
R^2	0.106	0.122	0.091	0.531	0.546	0.566
N	170	170	170	49	49	49

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors in parentheses. Results from linear regressions of dummies for exerting only social effort in the SE contract 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 $\lvert S-C\rvert$		Absolute imbalance $S-C$	
	No sorting	Sorting	No sorting	Sorting
	(1)	(2)	(3)	(4)
A. Differences in balance across treatments				
£0 (benchmark)	7.894	6.959	6.871	6.265
	(0.349)	(0.623)	(0.553)	(0.844)
£0.25	-2.197***	-2.192***	-7.522***	-6.128***
	(0.551)	(0.805)	(0.774)	(1.091)
£0.50	-2.570***	-2.523***	-7.632***	-5.212***
	(0.546)	(0.768)	(0.767)	(1.040)
£1	-2.717***	-3.495***	-7.111***	-6.065***
	(0.553)	(0.735)	(0.777)	(0.995)
R^2	0.043	0.064	0.163	0.111
N	708	341	708	341
B. t-tests of equality p-values				
£0.25 vs £0.50	0.490	0.627	0.884	0.321
£0.25 vs £1	0.343	0.043	0.592	0.942
£0.50 vs £1	0.786	0.103	0.493	0.291
C. Compassion				
$\geq 90^{\text{th}}$ pct.	4.590***	9.210***	5.805***	10.885***
	(1.449)	(1.986)	(1.710)	(2.432)
R^2	0.153	0.701	0.175	0.677
N	170	49	170	49
D. Social motivation				
$\geq 90^{\text{th}}$ pct.	-0.202	5.293	-0.461	6.849
	(1.706)	(3.885)	(2.007)	(4.522)
R^2	0.107	0.597	0.123	0.569
N	170	49	170	49

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

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

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered at the subject level in parentheses. The baseline category is the £0 treatment when sorting is not possible. See Table 4 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***	-3.850***	-0.248***	-0.255***	-2.494***	-2.694***
	(0.318)	(0.322)	(0.021)	(0.021)	(0.510)	(0.513)
Sorting	-0.303	-0.238	-0.020	-0.016	-0.935	-0.875
	(0.470)	(0.463)	(0.031)	(0.031)	(0.837)	(0.822)
Sorting × Bonus	0.824	0.756	0.054	0.050	-0.362	-0.333
	(0.501)	(0.495)	(0.033)	(0.033)	(0.868)	(0.851)
Constant	10.935***	11.544***	0.729***	0.769***	7.894***	7.860***
	(0.277)	(0.630)	(0.018)	(0.042)	(0.466)	(0.923)
Controls	No	Yes	No	Yes	No	Yes
N	1,049	1,049	1,049	1,049	1,049	1,049
R^2	0.151	0.175	0.150	0.173	0.062	0.108

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered at the subject level in parentheses. The baseline category is the £0 treatment when sorting is not possible (i.e. no bonus); *Bonus* captures all other treatments with a positive incentive. See Table 4 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. (2018) 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. (2018) 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 show that controlling for gender – our stratifying variable – does not affect our overall results.

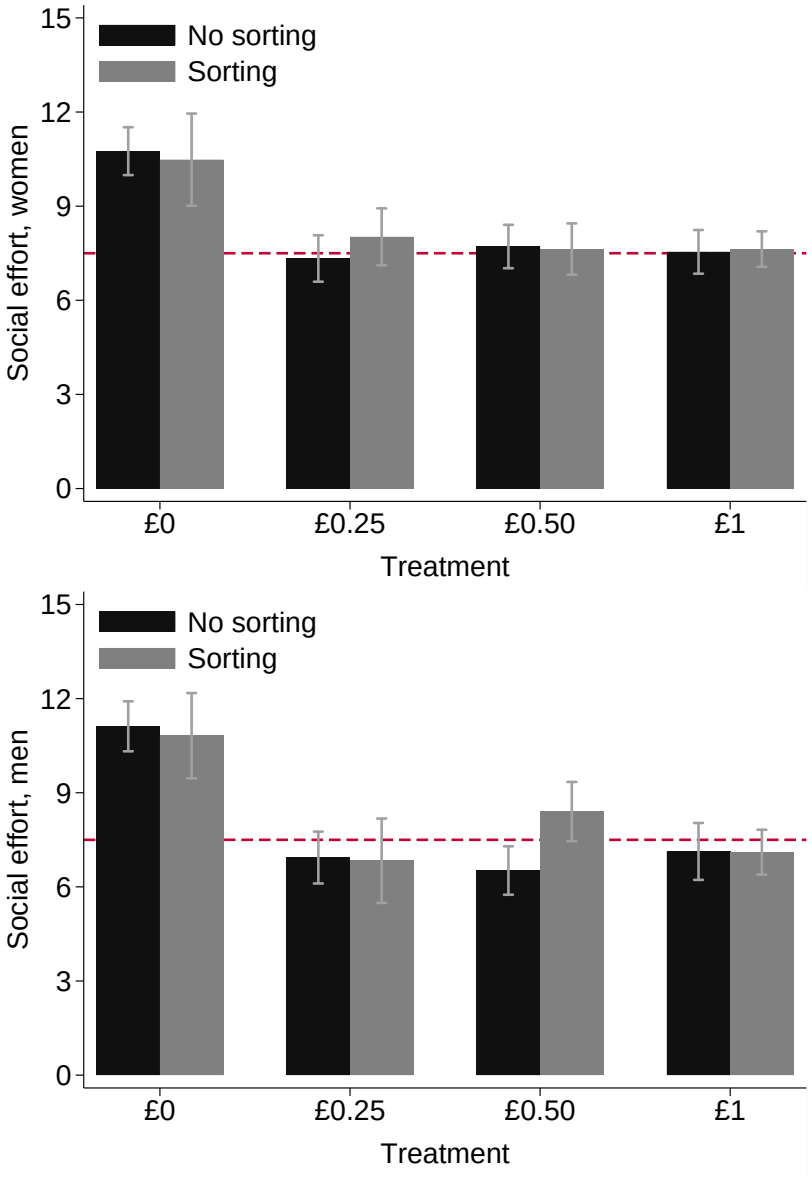

Figure G.1: Social effort in social enterprises by treatment for women (top) and men (bottom), with 95% confidence intervals; the dashed line represents a fully balanced effort allocation.

Mission heterogeneity Because social preferences may be weakly correlated with good cause choice (Appendix 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 SEs, Mair et al., 2012), also hints at the validity of our results for other types of missions.

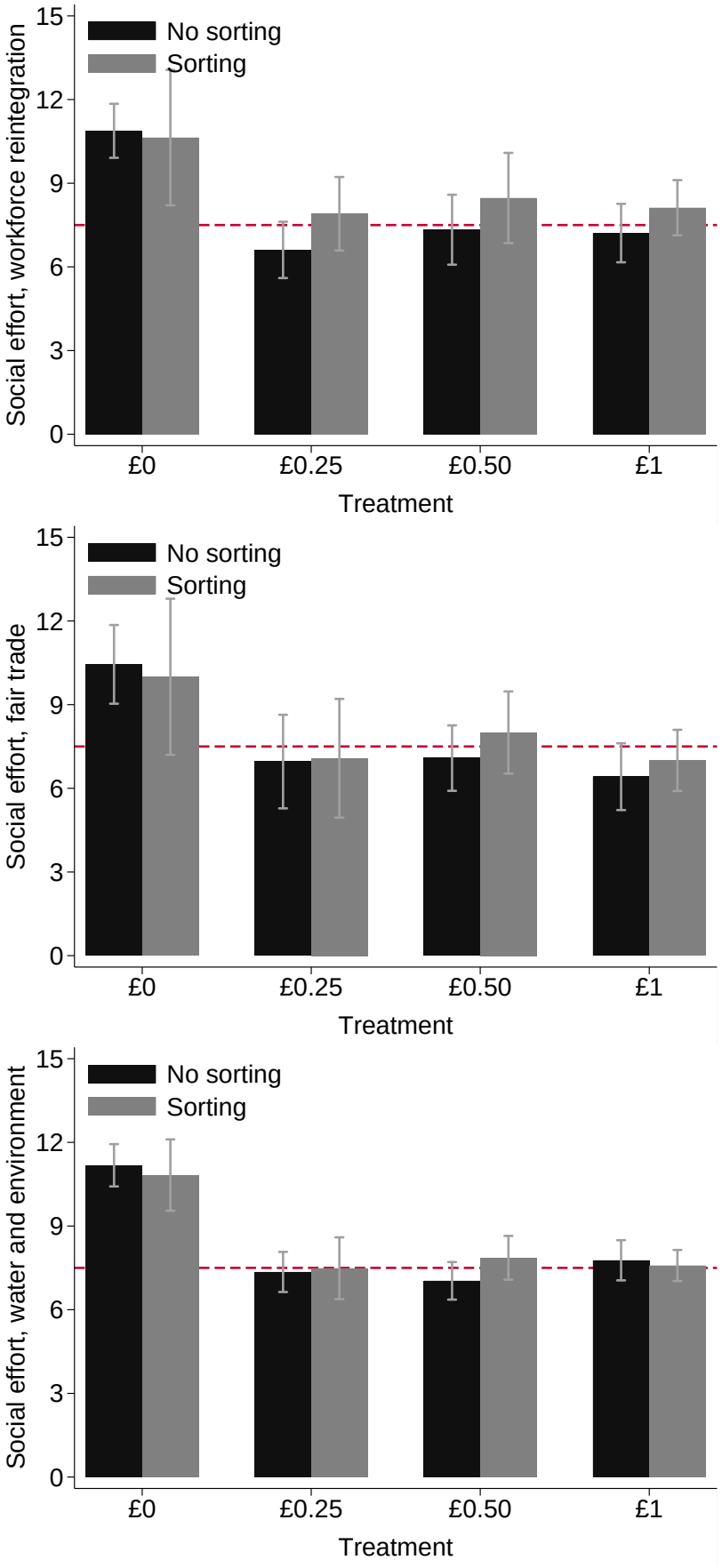

Figure G.2: Social effort in social enterprises by treatment and mission, with 95% confidence intervals; the dashed line represents a fully balanced effort allocation.

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

Comparison			p-value			
Group 1	Group 2	Difference	Unadjusted	Adjusted	Bonferroni	Holm
(1)	(2)	(3)	(4)	(5)	(6)	(7)
A1. Units of SE 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 SE 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 SE 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 SE 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. (2018), 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. (2018) multiplicity-adjusted p-value, and p-values from the application of conservative Bonferroni and Holm corrections.

Table G.2: **Gender Differences in SE Effort Allocation**

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

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

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

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

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