Efficiency wages with motivated agents

Games and Economic Behavior 145 (2024) 66–83

Contents lists available at ScienceDirect

Games and Economic Behavior

journal homepage: www.elsevier.com/locate/geb

Efficiency wages with motivated agents ☆

Jesper Armouti-Hansen [a], Lea Cassar [b,*], Anna Deréky [c], Florian Engl [d]

[a] Institute for Applied Microeconomics, University of Bonn, 53113 Bonn, Germany
[b] Chair of Empirical Economics, University of Regensburg, CEPR and CESifo affiliate, Universitätsstrasse 31, 93040 Regensburg, Germany
[c] Chair of Technology and Innovation Management, Department of Management, Technology, and Economics, ETH Zurich, 8092 Zurich, Switzerland
[d] Chair of Economic Psychology and Ethics, University of Regensburg, CESifo affiliate, Universitätsstrasse 31, 93040 Regensburg, Germany

ARTICLE INFO

JEL classification:
D23
M52

Keywords:
Prosocial motivation
Gift exchange
Biased beliefs
Efficiency wages

ABSTRACT

Many jobs serve a social purpose beyond profit maximization, contributing positively to society. This paper uses a modified principal-agent gift-exchange game with positive externality (prosocial treatment) to study how workers' prosocial motivation interacts with the use of efficiency wages in stimulating effort. We find that prosocial motivation and efficiency wages are independent in stimulating effort: compared to a standard gift-exchange game (GE treatment), the presence of the externality shifts the agents' effort choice function upwards without affecting its slope. Thus, if principals were profit-maximizers, wage offers should be the same in both treatments. However, principals offer higher wages in the prosocial treatment. We show that this is due to principals in the GE treatment highly underestimating agents' reciprocity and thereby offering wages below the profit-maximizing level. Results from robustness-checks further suggest that our findings are unlikely to be driven by a simple efficiency effect.

Recent empirical evidence shows that workers' motivation is often driven by different nonfinancial motives in addition to financial compensation, such as the willingness to contribute to a social purpose, as well as social preferences induced concerns (e.g., of reciprocity and fairness) towards the employer and colleagues (for a recent review on this topic see Cassar and Meier (2018)). However, the implications of these multiple non-monetary motives for incentive theory and Human Resource Management (HRM) are still unclear. To date, most economic research has either studied one of these two non-monetary aspects in isolation or ignored non-financial motives altogether. Are workers' prosocial motivation and reciprocity concerns complementary, substitutes or independent? How does the shape of this relationship affect wage contracts? Do principals correctly predict workers' preferences and thus offer the right wage? The answers to these questions can provide useful guidance for the design of compensation packages in social enterprises and organizations with social purposes. They can also provide new insights to the debate about the existence of a wage differential between profit and non-profit organizations.

This paper takes the first steps in addressing these questions by studying contracting in a laboratory setting where agents can be motivated both by the social impact of their job and by social preferences towards the principal. More specifically, we use experimental tools to investigate whether and how agents' prosocial motivation to contribute to the social good affects the emergence of a specific type of efficiency wage, namely, those wages that are set above the competitive level with the aim of motivating effort *by*

☆ **Funding**: Financial support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2126/1-390838866 is gratefully acknowledged.

* Corresponding author.
E-mail addresses: armoutihansen@uni-bonn.de (J. Armouti-Hansen), lea.cassar@ur.de (L. Cassar), adereky@ethz.ch (A. Deréky), florian.engl@ur.de (F. Engl).

https://doi.org/10.1016/j.geb.2024.03.001
Received 15 September 2021

Available online 16 March 2024

appealing to workers' sense of fairness and reciprocity.[1] The social impact of the job and efficiency wages appeal to two fundamental – but clearly distinct – aspects of workers' intrinsic motivation. The social impact refers to the effect that the workers' effort has on third parties and society more generally, and may motivate workers who care about the social cause underlying the job (e.g., Murdock, 2002; Benabou and Tirole, 2003; Besley and Ghatak, 2005; Delfgaauw and Dur, 2007, 2008; Ashraf et al., 2014; Prendergast, 2008; Besley and Ghatak, 2018; Cassar and Armouti-Hansen, 2020). Examples include doctors who are committed to saving lives, researchers to advancing knowledge, teachers to transferring knowledge and values to students, journalists to reporting information to the world, engineers to promoting green technology and so on. Efficiency wages relate to the relationship between the employer and employees, who can be motivated to exert effort by reciprocity and fairness concerns (e.g., Fehr et al., 1993, 1998a; Levine, 1998; Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Charness and Rabin, 2002; Brown et al., 2004; Fehr and Falk, 1999; Gneezy and List, 2006).

Our laboratory experiment consists of a modified version of the (one-shot) principal-agent gift-exchange game, in which the agent's effort, in addition to generating revenues to the principal, also generates a positive externality on society in the form of a donation to a charity chosen by the agent. As in the standard gift-exchange game, the principal offers the agent a fixed wage contract, the agent decides whether to accept the contract or not, and conditional on accepting, decides how much effort to exert. Payoffs realize accordingly and the game ends. In this setting the agent may be driven to exert positive effort by his motivation to reciprocate the principal's high wage offer and/or by his motivation to contribute to the charitable donation. In order to test the effect of the agents' prosocial motivation on the realized effort and wage contract, we then compare the participants' behavior in this modified version of the gift-exchange game (henceforth, prosocial treatment) to the behavior of other participants in a standard gift-exchange game without donation, and thus without positive externality (henceforth, GE treatment). In particular, in each treatment we use the strategy method to elicit agents' effort level for each possible wage offer, which allows us to construct an optimal effort function for each agent.[2]

Our experimental findings show that the positive externality decreases the agents' minimum acceptable wage offer and shifts the optimal effort choice function upwards without affecting its slope. This means that from the point of view of the agents' preferences, prosocial motivation and reciprocity concerns are independent in stimulating effort. According to our theoretical predictions, if principals were profit-maximizers, we should observe no difference in wage offers between the prosocial and GE treatment. Surprisingly, however, we observe that principals in the prosocial treatment offer a significantly higher wage than principals in the GE treatment. After ruling out other explanations (including that principals are not profit-maximizers and want to support the charity) we show that the effect is clearly driven by the principals' distorted beliefs about the agents' effort function in the GE treatment. More specifically, in the GE treatment principals highly underestimate the agents' reciprocity (i.e., the slope of the effort-wage function) and, in turn, offer a wage that is significantly lower than the profit-maximizing wage. Interestingly, this is not the case for the principals in the prosocial treatment: they correctly predict the slope of the effort-wage function, and they offer a wage that is not significantly different from the profit-maximizing wage. In other words, we show that in both treatments principals act as profit-maximizers based on their beliefs. However, given that in the GE treatment principals' beliefs are too pessimistic, they end up offering wages that are too low.

Finally, as robustness-checks, we run two additional treatments with the aim of investigating whether these results are purely driven by a simple "efficiency effect", that is by the simple fact the total surplus is higher in the prosocial than in the GE treatment, or whether it is the presence of the positive externality (the fact that the beneficiary of the extra-surplus is a third-party) which drives the results. In the first new treatment, the beneficiary of the donation is no longer the agent's chosen charity but a more "neutral" recipient, namely, another student from a different experiment or a charity to whom the agent allocated the least number of points in a series of dictator games played at the beginning of the experiment. In the second new treatment, we keep the same overall efficiency as in the prosocial treatment but remove the externality completely by letting the principals receive the extra-surplus generated by the agents' effort, thereby mimicking an equally efficient standard enterprise. Taken together, the findings from these two additional treatments suggest that our results are not purely driven by a simple efficiency effect and that the presence of an externality (i.e., of a third party benefiting from the agents' effort) is necessary to generate our results. However, consistent with previous laboratory evidence (Cassar, 2019), the specific third-party who benefits from the externality does not seem to play a significant role.

This paper makes several contributions to the literature. Most broadly, it contributes to the economics literatures that study workers' preferences for monetary and non-monetary job attributes and their implications for performance and contracting. This includes literature on prosocial motivation (e.g., Bénabou and Tirole, 2006; Ariely et al., 2009), on fairness and identity (e.g., Akerlof and Kranton, 2008; Fehr et al., 2009), on awards and recognition (e.g., Kosfeld and Neckermann, 2011; Gallus and Frey, 2016), and on autonomy and alternative work arrangements (e.g., Mas and Pallais, 2017). By being grounded on richer theories of human's motivation than the traditional Homo Oeconomicus view, all these literatures make important contributions to our understanding of behavior in organizations.

[1] There are, of course, also other microfoundations for why managers may want to pay efficiency wages. The most notable one is the payment of efficiency wages to increase the cost of job loss and thus to make the threat of firing in case of shirking more effective (Shapiro and Stiglitz, 1984). For the sake of exposition, throughout the paper we will refer to "efficiency wages" more generally, but it must be clear that we refer to efficiency wages that are microfounded on social preferences.

[2] One disadvantage of the strategy method ist that it can lead to "cold" decisions by pushing the workers to think strategically. However, previous evidence comparing decisions elicited using the strategy method vs the direct response method find no qualitative difference between the two (Brandts and Charness, 2011). In fact, no case was found where a significant effect was detected by the strategy method but disappeared under the direct response method (Charness and Kuhn, 2011). Thus, if anything, the strategy method is more conservative and is likely to provide a lower bound for the size of the effects. Furthermore, there is no reason to believe that the strategy method would affect the two treatments differently and, therefore, invalidate our results.

However, one limitation of these literatures is that they often remain disconnected. In fact, while different non-financial motives are likely to coexist in the workplace, economists typically study them in isolation, thereby overlooking their potential interaction. An early exception is the study by Ichniowski et al. (1997), which showed in an industrial context that the combination of incentive pay and a flexible job assignment increases productivity, which implies that for one context, at least, the complementarities between monetary and nonmonetary incentives are important. More recently, Bartling et al. (2012) show in an experimental study that such complementarities can endogenously lead to two different types of jobs: "'bad' jobs with low discretion, low wages, and little rent-sharing, and 'good' jobs with high discretion, high wages, and substantial rent-sharing" (p. 834). Within this experimental setting, low wages cannot be offset by non-monetary attributes - in this case a contract with full discretion - if they violate fairness norms. Additionally, Kvaløy et al. (2015) found a positive correlation between performance pay and motivational talk in a field experiment. Motivational talk is defined here as words that may evoke agents' intrinsic motivation. Furthermore, they find that performance pay without being accompanied by motivation talk actually decreased effort. Finally, the experimental work by Kosfeld et al. (2017) shows that while job purpose and monetary incentives are independent in motivating effort, purpose and recognition interact negatively, which suggests that they might operate through the same channel of image-seeking. We add to this literature by providing evidence that prosocial motivation and wage reciprocity are independent in motivating agents' effort and, therefore, that in these settings workers' preferences can be represented by utility functions which are additively separable in the social purpose and wage-reciprocity.

Within these broad streams of literature, we contribute in particular to the expanding literature on performance and contracting in mission-oriented organizations by revealing a new hidden benefit of the social mission.[3] While the focus of many of these studies is on how the social mission allows economizing on monetary incentives (Besley and Ghatak, 2005; Cassar, 2019), we show that when contracts are incomplete, the presence of the mission (in the form of a positive externality) fosters the emergence of efficiency wages. In other words, while monetary incentives and an organization's social purpose can be used as substitutes in incentivizing effort, we show that because of the principals' biased beliefs, efficiency wages and an organization's social purpose are complements in motivating effort. Furthermore, our results suggest that the contribution of a social mission to the creation of more trusting and cooperative environments is likely to go beyond the self-selection effects documented by previous studies (Fehrler and Kosfeld, 2013, 2014; Friebel et al., 2019).[4]

Our results also provide new insights for the debate about the existence of a wage differential between the profit and non-profit sector, whose evidence is mixed. On the one hand, wages offered by non-profit organizations are expected to be lower because of the labor donation argument. Consistent with this argument, Handy and Katz (1998) finds evidence of a negative wage premium in the non-profit sector. On the other hand, different evidence has shown that such wage differentials only exist for certain positions or certain industries, that they are rather small and that their existence depends on competitive forces, such as the share of meaning-driven workers in the market (Preston, 1988; Leete, 2001; Ruhm and Borkoski, 2003; Jones, 2015). Consistent with the property right theory argument that managers in non-profit organizations are accountable to no owner and thus may pay higher rent, Mocan and Tekin (2003) finds a positive wage premium in non-profit organizations compared to their for-profit counterparts. Similarly, Børsting and Thomsen (2017) finds that foundation-owned companies (i.e., companies with non-profit ownership) pay higher wages and have higher retention. We contribute to this literature by showing that this latter evidence is consistent with our results that efficiency wages are more likely to emerge in mission-oriented organizations. Hence, we identify a new channel (to be tested in the field) that can affect the wage differential between the profit and non-profit sector.

Our results also contribute to the ongoing debate about the relevance of reciprocity in motivating effort provision when contracts are incomplete (e.g., Fehr et al., 1993, 1996, 1997, 1998b; Fehr and Gächter, 1998; Fehr and Falk, 1999; Fehr and Gächter, 2000; Fehr and Falk, 2002; Brown et al., 2004; List, 2009). Our results provide a new explanation for why efficiency levels in a one-shot gift-exchange game are typically lower than in a repeated setting: Contrary to the argument that social preferences play no role and that reputation is all that matters for motivating effort when contracts are incomplete (List, 2009), we show that agents exhibit social preferences which, if they had been predicted correctly by the principals, would have led to a higher level of efficiency even in a one-shot setting with no scope for reputation. Hence, our findings provide strong support in favor of the role of social preferences as a motivator for effort, in the laboratory. In the field, however, the evidence for gift-exchange has been rather mixed (see e.g., Gneezy and List, 2006; Hennig-Schmidt et al., 2010; Esteves-Sorenson, 2017) and more recently DellaVigna et al. (2022), which finds evidence for moderate levels of reciprocity in the field). Given our goal of studying contracting with motivated agents, it was important to test, in addition to agents' behavior, also principals' offered wage contracts. Hence, for the purpose of our study, the laboratory setting – despite its limitations in terms of external validity compared to field experiments – was the most appropriate choice.

Finally, our paper provides some novel insights to the literature on belief formation in prosocial contexts (Fehr et al., 2005; Fehr, 2009; Schwerter and Zimmermann, 2020). To the best of our knowledge, this is one of the very few studies to elicit principals' beliefs

[3] In describing our study, we refrain from using the term "mission" and instead use terms such as "social impact", "job prosociality", and "positive externality". This shift emphasizes that, unlike the seminal paper by (Besley and Ghatak, 2005), our contribution does not focus on the effects of varying the quality of mission-matching. We thank an anonymous referee for bringing this distinction to our attention.

[4] A recent literature on corporate social responsibility (CSR) also shows that CSR can sometimes backfire and reduce workers' effort (List and Momeni, 2021; Cassar and Meier, 2020). We discuss how the present article differs from these studies in Section 6.

of returned effort in a gift-exchange game, and we are one of the first to do so in an incentivized manner.[5] Our results suggest that employers are likely to underestimate workers' reciprocity in the work context, which is consistent with the recent finding by Caria and Falco (2022) that small employers in urban Ghana systematically underestimate workers' reciprocity in a real-effort task. We contribute to this related literature by showing that in jobs with a prosocial purpose, such a bias is less likely to emerge. This may be due to employers trusting more the workers to reciprocate higher wages, when their reciprocity also benefit society.

The remainder of the paper is organized as follows. Section 1 derives the predictions based on a theoretical benchmark. Section 2 describes our experimental design. Section 3 presents the main experimental findings. Section 4 investigates the behavioral mechanisms underlying our findings. Section 5 present the results from two robustness checks. Section 6 discusses the implications, limitations, and potential extensions of our study. Section 7 concludes the paper.

1. Theoretical framework

In this section we generate our theoretical predictions by analysing the effect of an exogenous increase in the social impact of the agent's job on (i) the minimum wage offer that the agent would accept, (ii) the agent's optimal effort provision, and (iii) the principal's wage offer, in a one-shot principal-agent framework with incomplete contracts. This model should be taken as a theoretical benchmark to which we will then compare our experimental findings.

Let $m \in \mathbb{R}_+$ capture the social impact of the agent's job,[6] $w \in \mathbb{R}_+$ the wage that the principal can offer to the agent, and $e \in \mathbb{R}_+$ the effort choice of the agent. A given contract in this setting is simply given by the wage w, whereas the environment in which an agent may be employed is given by the tuple (w, m). The total realized social impact in the environment (w, m) is assumed to be increasing in the agent's effort choice e. This is a key factor in mapping the model to the experimental design. Last, when offering a contract to the agent, the principal must take into consideration the outside option of the agent in the form of a transfer valuated at τ for any agent.

Let $U(w, m, e; \theta)$ be the agent's utility function. We assume that by entering a contract, the agent derives an extrinsic benefit depending on the wage received and the effort exerted. In addition to the wage and effort, the agent derives a potential intrinsic benefit from the social impact of the job. We assume that these two terms are additively separable such that his utility may be specified as

$$U(w, m, e; \theta) = w - \frac{c}{2}e^2 + M(w, m, e; \theta) \tag{1}$$

where the two first terms are the standard way of representing the extrinsic benefit, with $\frac{c}{2}e^2$ being an increasing and convex function capturing the disutility of exerting effort.[7] The last term, $M(w, m, e; \theta) \geq 0$, is the intrinsic benefit of an agent endowed with θ working in the present environment (w, m) and exerting effort e. The magnitude with which the intrinsic benefit impacts the agent's utility is governed by θ, which is distributed according to F_θ, and has well-defined first two moments. As in Cassar and Meier (2018), θ is a vector of parameters indicating the weights assigned to different intrinsic benefits. Thus, any heterogeneity in the utility of agents stems from this vector of parameters. Prosocial motivation and/or social preferences towards the principal stem from the function M. Social preferences toward the principal may consist of distributional concerns (as in Fehr and Schmidt, 1999; Charness and Rabin, 2002) as well as reciprocity (as in Rabin, 1993; Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2005) where the agent cares more about the principal if he judges an action taken by her as kind. Additionally, prosocial motivation may, in addition to pure altruism, contain impure motives such as warm glow (as in Andreoni, 1989, 1990). The meaning of - and the assumptions imposed on - the (cross-) partial derivatives of M are as follows:

- The marginal intrinsic utility of increasing the job's social impact is captured by the sign and magnitude of the partial derivative M_m and prosocial motivation is captured by the sign and magnitude of the cross derivative M_{em}, both assumed to be non-negative for all w, m, e, θ and positive in expectation for all w, m, e. The latter assumption corresponds to assuming that no worker is "antisocial." Furthermore, to simplify the comparative statics, we assume that $M_{eem} = 0$ for all w, m, e, θ.
- The marginal intrinsic utility of exerting effort is given by M_e, which is assumed to be (i) non-negative for all w, m, e, θ; (ii) positive in expectation for all m, e and $w > 0$; (iii) weakly decreasing for all w, m, e, θ and (iv) strictly decreasing in e in expectation for all w, m, e. Thus, given an environment with a positive social impact and positive wage, the average worker derives an intrinsic benefit by exerting effort. We will also assume that $M_{eee} = 0$.[8]
- The marginal intrinsic utility of increasing the wage is captured by the sign and magnitude of the partial derivative M_w, and reciprocity is captured by the sign and magnitude of the cross derivative M_{ew}, both assumed to be non-negative for all w, m, e, θ

[5] Brown et al. (2004) also elicits the principals' beliefs in a repeated gift-exchange game; however, their elicitation was not incentivized and was not based on the strategy method. In an unpublished paper, Guido et al. (2020) elicit beliefs in an incentivized manner with the strategy method in a repeated gift-exchange game looking at endowment shocks and information asymmetry, but without a baseline condition excluding shocks.

[6] That is, m measures how prosocial the agent's job is and if $m' > m''$ then a job with m' is more prosocial than a job with m''.

[7] Note that whereas this functional form does not completely capture the effort costs as listed in Table 1, by setting $c \approx 0.38$, the costs of effort in the experiment are very well approximated ($R^2 \approx 0.98$). Hence, this functional form provides a simple and effective mapping to the experiment. Although the comparative statics would increase in complexity, our results are, however, robust to a more general cost function, $c(e)$ with $c'(e) > 0$ for $e > 0$, $c'(0) \geq 0$, $c''(e) > 0$ and $c'''(e) \geq 0$ for all e.

[8] This assumption simplifies the comparative statics and the intuition behind our results significantly. However, as we show in the Online Appendix, our results do not significantly change by the weaker assumption, $M_{eee} \leq 0$.

and positive in expectation for all w, m, e. Furthermore, as a technicality, we assume that $M_{eew} = 0$ and $M_{eww} \leq 0$ for all w, m, e, θ.

- The potential additional (de)motivation arising from the interaction between reciprocity and prosocial motivation is captured by the sign and magnitude of M_{ewm}. As the subsequent analysis will show, this will be the term of main interest.

As the scope of the model is to provide a theoretical benchmark, the principal is assumed to be a risk-neutral profit-maximizer. For this, denote the expected effort choice of the agent given (w, m) by $\bar{e}^*(w, m)$, the expected lowest acceptable wage to the agent given m by $\underline{\bar{w}}_m(m, e^*)$, and the agent's expected intrinsic utility given (w, m) by $\bar{M}(w, m, e^*)$, where expectation is taken over θ.[9] Thus, the principal offers the wage w which maximizes her expected profits given by $r\bar{e}^*(w, m) - w$, with $r > 0$. In the analysis, we assume that the principal cannot observe the agent's type vector θ but that the distribution F_θ and the functional form of M are common knowledge. Note that because the intrinsic benefit is assumed to be increasing in m, it immediately follows that the lowest acceptable wage, that is, the wage that binds the participation constraint, decreases by an exogenous increase in the job's social impact:

Prediction 1. The expected lowest acceptable wage to the agents decreases with the job's social impact. That is, $\underline{\bar{w}}_m(m, e^*) < 0$.[10]

In addition, since the agent chooses the effort level which equalizes his marginal intrinsic benefit and his marginal cost of exerting effort, it follows that an exogenous increase in the job's social impact increases the effort choice, if the participation constraint is satisfied:

Prediction 2. For any given wage w, the expected optimal effort choice of agents increases with the job's social impact. That is, $\bar{e}^*_m(w, m) > 0$.[11]

Finally, notice that a change in the profit-maximizing wage $w^*(m, e^*)$, by an exogenous increase in the job's social impact, is fully determined by the change in the agent's effort-wage slope, $e^*_{wm}(w, m; \theta)$. As the type of the agent is unknown to the principal, the change in the expected profit-maximizing wage, $\bar{w}^*(m, e^*)$, is determined by the expected change in the agent's effort-wage slope, $\bar{e}^*_{wm}(w, m)$. As any reaction to the social impact comes from the agent's intrinsic benefit function M, it follows that the expected profit-maximizing wage increases if and only if expected additional motivation from the interaction between wages and prosocial motivation, $\bar{M}_{ewm}(w, m, e^*)$, is positive:

Prediction 3. In the interior, the profit-maximizing wage offer, w^*, increases (decreases) if the expected additional motivation from the interaction between wages and prosocial motivation, $\bar{M}_{ewm}(w, m, e^*)$, is positive (negative).[12]

Fig. 1 illustrates three possible examples of the profit-maximizing wage offer before and after an increase in the job's social impact, depending on the agent's expected optimal effort choice as a function of the wage. In Case 1, the slope of the expected effort choices remains constant, leading to the profit-maximizing wage being identical before and after an increase in the social impact. In Case 2, the slope of the expected effort choices increases, which in turn leads to an increase in the profit-maximizing wage offer. Finally, in Case 3, there is a flattening in the expected effort slope and hence a decrease in the profit-maximizing wage offer. In Section 1.3. of the Online Appendix, we also provide examples of explicit functional forms of the intrinsic utility function M and provide behavioral explanations of their meaning. Thus, all three situations are rationalizable rendering a priori prediction vacuous.

We conclude our theoretical exposition as follows: Although we are able to make reasonable predictions on (i) change in the minimum acceptable wage offer and (ii) change in the effort exerted for any given wage offer following an increase in the job's social impact, (iii) it *is unclear a priori what happens to the slope of the agent's optimal effort function* and thus to the profit-maximizing wage offer following an increase in the social impact What we can say is that it will depend on the interaction between the agent's social preferences towards the principal and his prosocial motivation.

2. Experimental design

The objective of the experiment is to investigate empirically the relationship between the job's social impact and efficiency wage. It was designed specifically to test our theoretical predictions on how the job's social impact affects (i) agents' acceptance rates, (ii) agents' effort provisions and (iii) the principals' wage offers.

We collected experimental data in six separate sessions. At the beginning of each session, participants were informed that the experiment comprised two stages, and were told also that their decisions in one stage would be irrelevant for the other stage except for their chosen charity (see below). In each session, a single treatment variation was implemented. Participants were not given details about the second stage until they had completed the first stage. In the first stage, we elicited participants' prosocial/charitable

[9] Specifically, $\bar{e}^*(w, m) = E_\theta[e^*(w, m; \theta)]$, $\underline{\bar{w}}_m(m, e^*) = E_\theta[\underline{w}(m, e^*(\underline{w}, m; \theta); \theta)]$ and $\bar{M}(w, m, e^*) = E_\theta[M(w, m, e^*(w, m; \theta); \theta)]$.

[10] Please see Section 1.1 in the Online Appendix for the derivation.

[11] Please see Section 1.1 in the Online Appendix for the derivation.

[12] Please see Section 1.2 in the Online Appendix for the derivation.

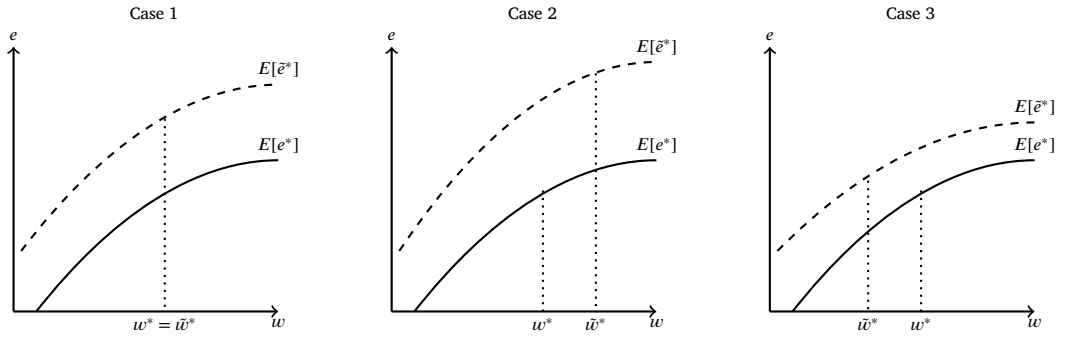

Fig. 1. Examples of expected effort wage slopes and optimal wage offers for $E[e^*_{iwm}] = 0$ in Case 1, $E[e^*_{iwm}] > 0$ in Case 2 and $E[e^*_{iwm}] < 0$ in Case 3.
Note: $E[e^*]$: expected slope before an increase of the job's social impact; $E[\tilde{e}^*]$: expected slope after an increase of the social impact. w^*: optimal wage before an increase of the social impact; $\tilde{w}^*$: optimal wage after an increase of the social impact.

motivation and fairness concerns using a dictator game with a donation to a charity and an ultimatum game, respectively. In the second stage, the main part of the experiment, we implemented a stylized version of a contractual setting using a principal-agent gift-exchange game, with or without social impact. The social impact of the job was implemented by letting the agent's effort generate a positive externality in the form of a donation to a charity of the agent's choice. In each session, participants completed both stages. Which stage counted towards the payment was determined randomly. The selected stage was the same for all participants in a given session. Individual payoffs and earnings were revealed only after both stages were completed. All participants were asked to choose their preferred charitable organization from a list of 12 charities (see Online Appendix). They were informed that all donations they generated in the stage chosen for payment would be paid to the organization of their choice.

2.1. Dictator and ultimatum game

We elicited participants' preferences before the main experiment, using a dictator game with a donation to charity, and an ultimatum game. In the dictator game, all participants were asked to divide 100 points (in multiples of 10) between themselves and their chosen charitable organization. In the ultimatum game, participants were randomly assigned to the roles of either proposer or responder and were randomly matched in pairs. The proposer received an endowment of 100 points and was asked to propose a split of these points (in multiples of 10) between himself/herself and the responder. Beforehand, the responder was asked to indicate acceptance or not for each possible split. If the proposed allocation was accepted, the players received the corresponding amounts. If the proposal was rejected, neither player of the pair received anything. The resulting individual payoffs were revealed only at the end of the experiment.

2.2. Gift-exchange game and treatment variation

We used the same random assignment of roles in the ultimatum game to divide participants between principals and agents. In other words, those in the role of the proposers were now principals, while those in the role of the responders were now agents. Principals and agents were then matched randomly in pairs, and depending on the session, were allocated to one of the two treatments: GE treatment or prosocial treatment. Those in the GE treatment played a standard gift-exchange game where the agent decided whether or not to accept the wage contract offered by the principal, and conditional on acceptance what level of effort to exert. The only difference in the prosocial treatment was that the agent's effort generated a donation to the charity that was chosen at the beginning of the experiment. The donation was an externality paid by the experimenter.

Timings and payoffs in both treatments were as follows. The principal chose a lump-sum wage offer w. Meanwhile, the agent chose, for each possible wage offer, whether he would accept the contract and, conditional on accepting, a costly level of effort e.[13] As in Fehr et al. (1993), the set of possible wages is given by $w \in \{1, 3, 5, 10, 15, \ldots, 65, 75, 85, 95\}$, and the set of possible effort levels is given by $e \in \{1, \ldots, 10\}$. In the case of rejection of a given wage offer, we code $e = 0$. Both parties received an initial endowment of 100 points to ensure a non-negative return for all participants. There was an additional 5-point outside option available to the agent, should the contract not be concluded: i.e. if the principal offered a wage which the agent did not accept. There was no such outside option for the principal. Thus, in the case that the agent rejected the contract, the monetary payoffs to the principal and agent were respectively 100 and 105 points. If the contract was accepted by the agent, then the principal's and the agent's monetary payoffs from concluding the contract were respectively

$$\begin{aligned} \Pi_{P|e>0} &= 10e - w + 100 \\ \Pi_{A|e>0} &= w - c(e) + 100 \end{aligned} \tag{2}$$

[13] The use of the strategy method allows us to reconstruct an optimal effort function of the wage for each agent, which is essential to testing how the slope of this function varies across treatments.

Table 1
Effort and corresponding costs of effort.

e	1	2	3	4	5	6	7	8	9	10
$c(e)$	0	1	2	4	6	8	10	12	15	18

where the first term of the principal's monetary payoff constituted her revenue based on the agent's chosen effort level multiplied by 10, and the last term her endowment. If the contract was accepted, the principal was bound to pay the wage to the agent. The agent's monetary payoffs consisted of the wage offered and his endowment minus the cost of the chosen effort level. Table 1 presents the costs of each effort level, which are the same as in Fehr et al. (1993) and Brown et al. (2004). Since the marginal cost is increasing with the effort, it corresponds to a large extend to the quadratic cost function considered in the examples in section 1.

Additionally, in the prosocial treatment, if the principal-agent pair concluded a contract, the charitable organization received a donation:

$$\Pi_C = 25e \tag{3}$$

Thus, the charitable organization received the agent's chosen effort level multiplied by 25.

Note that if both parties were maximizing their monetary payoffs, the payoffs corresponding to the unique Nash equilibrium were $(105, 105)$. This corresponds to the principal offering a wage that matches the outside option, i.e. $w = 5$, making the agent indifferent between rejecting the contract and receiving 105, or accepting it and exerting the minimum effort $e = 1$, which comes at a cost of 0. This applies because the agent's best response was to reject the contract for any wage offer below the outside option, and to accept the contract by exerting minimum effort level for any wage offer above the outside option. Thus, a wage offer of 5 maximized the principal's payoff.

2.3. Elicitation of principals' beliefs

Following the principals' wage choices, we elicited their beliefs about the agents' effort responses. The use of beliefs elicitation to explain behavior has, however, its limitations. First, the casuality is not clear. Second, subjects might engage in hedging behavior or report biased beliefs to justify their own actions. We tried to minimize these limitations in two ways. First, to avoid priming the participants, the instructions did not mention that we would elicit principals' beliefs. This may also help reduce the risk of planned hedging behavior as principals did not expect to be asked about their beliefs. Second, and more importantly, the beliefs elicitation was incentivized. Hence, it was costly for the subjects to report wrong beliefs just to be consistent with their previous actions. Specifically, for each possible wage offer, we asked principals to guess whether the matched agents would accept the contract, and conditional on accepting, what effort level would most likely be chosen. Principals received 0.5 points for correctly guessing (i) each wage that was not accepted, and (ii) each accepted wage and chosen effort by the agent. Thus, we elicited the modal effort choice for each possible wage offer.[14] After eliciting principals' beliefs about agents' effort, we asked principals to guess the wage that, given her beliefs, maximizes profit. In the empirical analysis, we will refer to this variable as the "guessed profit-maximizing wage". The question we posed was: "Based on your guesses of what wages the worker would accept and how much effort he/she would put, what wage level do you think gives you the highest income?". The principal received 0.5 points for a correct guess.

2.4. Procedural details

The six laboratory sessions were conducted at the University of Cologne in September 2016. In total, 190 students participated in this between-subject design experiment, none of whom participated in more than one session. In five of the six sessions, 32 subjects participated, whereas 30 subjects participated in the remaining session. Among the 190 participants, 94 were assigned to the GE treatment and 96 were assigned to the prosocial treatment. The experiment was programmed in and used z-Tree software (Fischbacher, 1999).[15] Participants in the experiments received points with a conversion rate of 1/12. Average earnings were 13.72 euro with a standard deviation of 2.14 and a minimum earning of 5.66 euro.[16]

[14] By imposing the assumption that the mode coincides with the mean (e.g., as with a symmetric unimodal distribution), we have elicited the principals' conditional expectation in a risk-robust manner (Hurley and Shogren, 2005).

[15] For instructions, please see section 3. Supplemental Material in the Online Appendix. We are grateful to a referee for spotting a mistake in the instructions in which we told the students "You will receive a copy of the transfer receipt of all the donations generated in this session by email. Due to the upcoming **holidays' season**, please allow approximately one month for us to make the donation." The "holidays' season" was mistakenly copied from the instructions of an earlier experiment which was run in December and which's instructions we partly reused. We forgot to delete this part. However, we have no reason to believe that this affected subjects' behavior in the experiment in a systematic way. Furthermore, no students raised concerns about this point during the sessions.

[16] Summary statistics of subjects' *age, gender* and *study subject* across treatments is given in Table A.1 in the Online Appendix.

Table 2
Regressions (Wage acceptance).

	(1)	(2)	(3)	(4)	(5)	(6)
Wage<5	−0.707***	−0.759***	−0.759***	−5.417***	−6.474***	−6.516***
	(0.033)	(0.040)	(0.040)	(0.402)	(0.756)	(0.753)
Prosocial treatment		0.031	0.035		0.415	0.477
		(0.041)	(0.040)		(0.426)	(0.418)
Wage<5∗Prosocial treatment		0.103	0.103		1.667*	1.714*
		(0.065)	(0.065)		(0.908)	(0.904)
Constant	0.818***	0.802***	0.757***	2.237***	2.012***	1.333
	(0.021)	(0.033)	(0.147)	(0.253)	(0.284)	(1.386)
Controls	No	No	Yes	No	No	Yes
(Pseudo) R^2	0.256	0.260	0.263	0.269	0.273	0.274
Observations	1710	1710	1710	1710	1710	1710

Note: Random effects linear regressions in Columns (1) to (3). Random effects logistic regression in Columns (4) to (6). Controls include gender, age and study subject. All regressions are estimated with standard errors clustered on the individual level. Significance levels: *p<0.1; **p<0.05; ***p<0.01.

3. Results

3.1. Agents' acceptance rates and effort choices

We start by looking at the agents' wage acceptance rates across treatments. According to our theoretical model, the average minimum acceptable wage should decrease in the prosocial treatment compared to the GE treatment. Fig. A.1 summarizes the percentage of agents accepting the offer for each possible wage across the two treatments. We find no significant treatment differences in acceptance rates for the vast majority of wages above the agents' outside option, of 5. Additionally, the acceptance rate increases with the wage. However, at wage offers below the outside option of 5, we find marginally significant treatment differences. Below wage offers of 5, approximately 4 percent of the agents in the GE treatment accepted a wage offer (for both wages equal to 1 and 3) compared to approximately 17 and 19 percent in the prosocial treatment (respectively for wages equal to 1 and 3). This difference is marginally significant for both wage levels using the two-sample Wilcoxon rank-sum test ($p = 0.07$).[17] Thus, consistent with the theory, the presence of the positive externality reduces the agent's average minimum acceptable wage.

The regressions in Table 2 provide further (weak) evidence that agents are more likely to accept a wage below their outside option in the prosocial treatment than in the GE treatment. The interaction term between the dummy variable "prosocial treatment" and "a wage below 5" is positive and marginally significant in the logistic regressions.[18]

In summary, the presence of a positive externality leads more agents to accept a contract with which they are losing money, as compared to the outside option. However, the effect is rather weak. We conclude:

Result 1. In line with Prediction 1, the average minimum acceptable wage offer is marginally smaller in the prosocial treatment compared to the GE treatment.

After investigating agents' acceptance rates, we characterize their effort choices. Fig. 2 shows the mean effort choice for each potential wage offer in both treatments. Since we observe the chosen effort levels for all agents at every wage level, this graph provides an approximation of the agents' optimal effort function in each treatment. It shows that the externality shifted the agents' optimal effort function upwards. For each possible wage offer, the average effort is higher in the prosocial treatment than in the GE treatment (the difference is significant or marginally significant for most wages below 55).

Next, we investigate whether the positive externality also affects the slope of the agents' optimal effort function. We test for treatment differences investigating the change in effort with increasing wage. For example, we start by testing whether the change in effort level following an increase in the wage from 1 to 3 is different across treatments, and conduct the same analysis for all remaining wage differentials. We find no significant differences over most of the wage intervals, except at wages 3, 45 and 55. As expected, and based on our previous results, in the prosocial treatment agents exert more effort than in the GE treatment in response to an increase in the wage from 1 to 3 ($p = 0.08$). This is because wages 1 and 3 are below the outside option and very few agents are motivated to accept such offers in the GE treatment. However, in the case of middle-level wages such as 45 and 55, the agents' effort responses to an increase in the wage are significantly lower in the prosocial treatment compared to the GE treatment, suggesting that at these points the agents' reaction is flatter in the prosocial treatment than in the GE treatment ($p = 0.02$ for both wage levels).[19]

[17] For the rest of the analysis, we continue to use the Wilcoxon rank-sum test, unless otherwise specified.

[18] Results do not change if we add more controls, such as behavior in the dictator and ultimatum games, which were run prior to the main experiment, or answers to the ex-post questionnaire.

[19] See Fig. A.5 and Fig. A.6 for the individual effort choice function of each agent in the GE and prosocial treatment, respectively. As can be seen, there are 7 non-reciprocal agents (i.e., agents whose effort-wage curve is basically flat) in the GE treatment and 6 in the prosocial treatment.

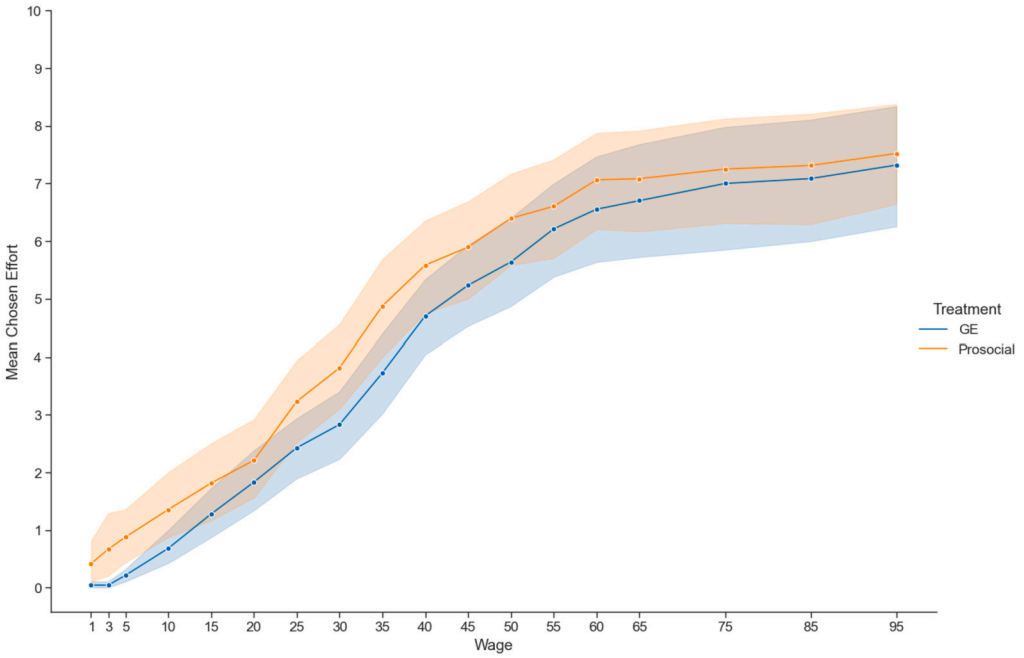

Fig. 2. Mean chosen effort for each wage across treatments.

Table 3
Regressions (Effort).

	(1)	(2)	(3)	(4)	(5)	(6)
Wage	0.089***	0.091***	0.091***	0.138***	0.145***	0.145***
	(0.005)	(0.007)	(0.007)	(0.008)	(0.011)	(0.011)
Prosocial treatment		0.749**	0.634*		1.431**	1.340*
		(0.302)	(0.374)		(0.626)	(0.707)
Wage∗Prosocial treatment		−0.004	−0.004		−0.013	−0.014
		(0.009)	(0.009)		(0.016)	(0.016)
Constant	0.620***	0.242*	−2.993**	−1.826***	−2.552***	−7.608***
	(0.157)	(0.142)	(1.239)	(0.360)	(0.413)	(1.946)
Controls	No	No	Yes	No	No	Yes
(Psuedo) R^2	0.459	0.466	0.491	0.138	0.141	0.150
Observations	1710	1710	1710	1710	1710	1710

Note: Random effects linear regressions in Columns (1) to (3). Random effects tobit regression in Columns (4) to (6). Controls include gender, age and study subject. All regressions are estimated with standard errors clustered on the individual level. Significance levels: *p<0.1; **p<0.05; ***p<0.01.

We complement our non-parametric analysis by running OLS and Tobit regressions of the agents' effort choices on the wage level, with a treatment dummy and the wage-treatment interaction (see Table 3). The coefficient of the prosocial treatment dummy is positive and either significant or marginally significant when adding controls, which confirms that the externality shifts the agents' optimal effort function upwards.[20] On the contrary, we find no change in the slope. The coefficient of the interaction between the prosocial treatment dummy and the wage is very close to 0 and largely insignificant.[21] These results suggest that the externality shifts the agents' effort function upwards without affecting its slope. We conclude:

Result 2. In line with Prediction 2, effort provision tends to be larger for any given wage in the prosocial treatment compared to the GE treatment. Furthermore, we find no difference in the slope of the optimal effort function in the prosocial treatment compared to the GE treatment. Hence, social impact and wages are independent in motivating agents' effort.

[20] Results do not change if we add more controls, such as behavior in the dictator and ultimatum games, which were run prior to the main experiment, or answers to the ex-post questionnaire.

[21] Note that if we only focus on the effort choices conditionally on accepting the contract, the results strengthen, and the negative interaction term also becomes significant. However, given that this regression does not control for selection into the contract, we chose to exclude it from the analysis.

Table 4
Regressions (Offered wage).

	(1)	(2)	(3)	(4)
Prosocial treatment	10.196**	10.187**	14.593***	13.218**
	(3.952)	(3.958)	(5.668)	(5.669)
Charitable motivation		0.098	0.250**	0.145
		(0.110)	(0.124)	(0.149)
Prosocial treatment*Charitable motivation			−0.287	−0.182
			(0.204)	(0.203)
Constant	25.574***	24.072***	21.740***	21.360
	(2.443)	(3.161)	(3.589)	(13.294)
Controls	No	No	No	Yes
R^2	0.066	0.073	0.088	0.115
Observations	95	95	95	95

Ordinary least-squares (OLS) Regressions with robust standard errors in columns (1) to (4). "Charitable motivation" is measured by the numbers of points that the principals allocated to the charity in the dictator game played at the beginning of the experiment. Controls include gender, age and study subject. Significance levels: *p<0.1; **p<0.05; ***p<0.01.

3.2. Principals' behavior

Next, we look at the behavior of the principals. Note that Result 2 suggests that we are in Case 1 of our model; that is, the agents' utility function is additively separable in the job's social impact and wage. Our theory predicts that in this case the optimal wage is independent of the job's social impact, i.e., of the externality. Hence, in our experiment, if principals were profit-maximizers, we should observe no significant differences in the wage offers across treatments.

However, contrary to our predictions, in the GE treatment principals offer on average a wage of 26 points, while in the prosocial treatment the average offered wage is about 36 points, hence 10 points higher. The difference is statistically significant ($p = 0.01$). OLS regressions in Table 4 provide similar results.[22] Thus, we conclude:

Result 3. Contrary to our predictions, principals in the GE treatment offer significantly lower wages than principals in the prosocial treatment.

The next question arises: who is getting it wrong? Are the principals in the GE treatment or in the prosocial treatment (or both) sacrificing profits? Surprised by Result 3, we investigate the potential mechanism(s) underlying the principals' behavior. We discuss and test two possible behavioral channels for this result in Section 4 below.

4. Principals' behavioral channels

There are two potential channels that can explain the principals' behavior: (i) principals care about contributing (indirectly) to the charitable donation and thus offer a higher wage in the prosocial treatment in order to boost the agent's effort; or (ii) principals act as profit-maximizers based on biased beliefs about the agents' effort response. These potential explanations are not mutually exclusive, and our results might be due to the combination of both of them. We subsequently show that the latter channel is the only possible explanation.

4.1. Charitable motivation and profit-maximizing wage

First, if the principals were motivated by the charitable motives, our measure of "charitable motivation" elicited at the beginning of the experiment through a dictator game with charity should enter significantly in the principals' wage-setting. In particular, we should observe a positive interaction between the prosocial treatment and the principals' charitable motivation. Results from OLS regressions are reported in columns 3 and 4 of Table 4. Since the interaction term is insignificant and the point estimate is slightly negative, we can conclude that the principals' charitable motivation does not lead to higher wage offers in the prosocial treatment.

Second, if the principals' charitable motivation were the main mechanism underlying Result 3, we should observe principals maximizing profits in the GE treatment, while sacrificing profits in the prosocial treatment. To test this, we compare the wage offered by the principals to what they believed was the profit-maximizing wage, which we elicited at the end of the experiment in an incentivized manner and denoted as the principal's "guessed profit-maximizing wage". If the principals were motivated to sacrifice profit to benefit the charity in the prosocial treatment, they should offer a wage that is significantly higher than what they believe to be the profit-maximizing wage. However, Fig. 3 shows that this is not case. In both treatments, principals offer wages that are not significantly higher than their guessed profit-maximizing wage (Wilcoxon signed-rank test $p > 0.42$ in both treatments).

[22] Results do not change if we add more controls, such as behavior in the dictator and ultimatum games, which were run prior to the main experiment, or answers to the ex-post questionnaire.

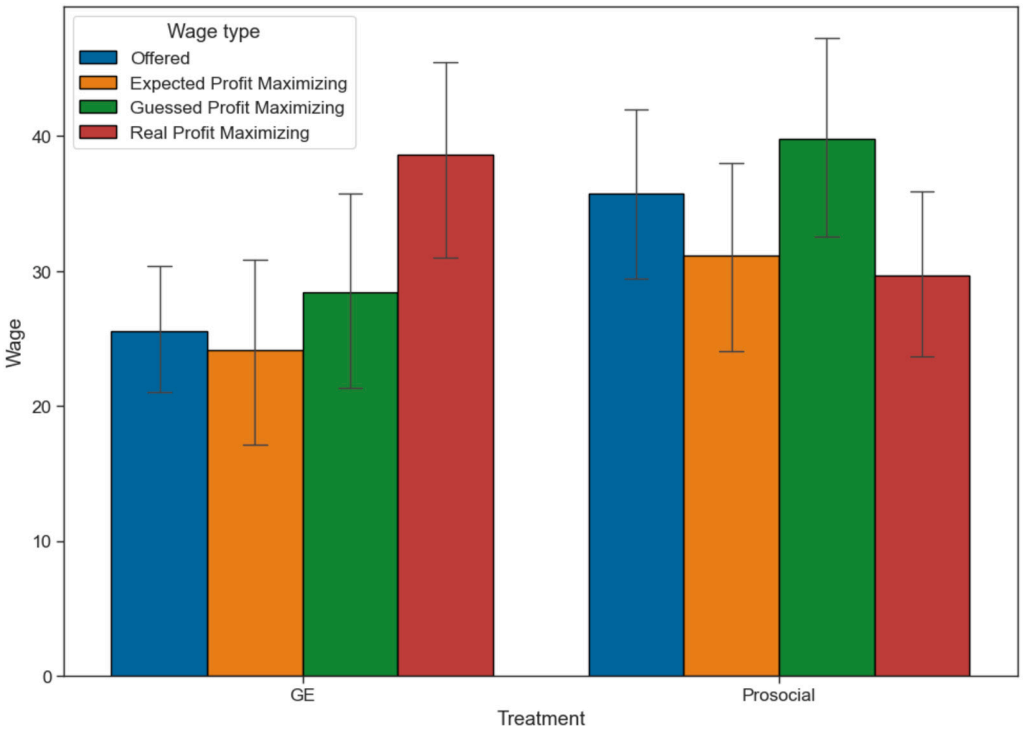

Fig. 3. Mean offered wage and mean expected, guessed, and average real profit-maximizing wage across treatments.

4.2. Biased beliefs

Finally, we investigate the second possible explanation, which is that principals are profit-maximizers but have biased beliefs about the agents' optimal effort response. Fig. 4 compares the difference between the average effort chosen by the agent and the average effort expected by the principal for each treatment and wage. We see that in both treatments and for most wage levels, principals underestimate the agents' effort: The expected effort was always equal to or lower than the chosen effort.

However, as our theory clearly shows, what matters for determining the profit-maximizing wage is not the absolute value of the exerted effort but the slope of the optimal effort function. The left-hand panel of Fig. 4 clearly shows that in the GE treatment, the difference between the real and expected effort increases significantly with the wage. Principals are good at predicting the effort for low wage levels but increasingly underestimate the effort response to an increase in the wage – to the point that at a wage of 95 the expected effort is approximately 60 percent lower than the real effort. In other words, principals underestimate the role of wages in motivating effort, i.e., agents' reciprocity. This implies that in the GE treatment the principals' expected optimal effort function is flatter than the real average optimal effort function. This result is also confirmed by rank-sum tests, which compare the variations in the chosen effort to the variations in the expected effort following an increase in the wage: Significant differences in the slope emerge at wage levels equal to 15, 35, 40, 45, 50 and 60.

We obtain similar findings from the regression analysis. Columns 1 and 2 in Table A.2 present the respective chosen effort and expected effort regressions on wage levels in the GE treatment. The wage coefficient in column 1 is almost twice as large as the wage coefficient in Column 2, suggesting that the linear approximation of the agents' optimal effort function is almost twice as steep as the principals' average expected effort function. Thus, it is not surprising that principals in the GE treatment offer a wage that is too low compared to the profit-maximizing wage.[23]

Additionally we calculate for each principal-agent pair the profit-maximizing wage based on the principal's elicited beliefs (henceforth, "expected profit-maximizing wage"),[24] which are depicted in Fig. 3. It shows that in the GE treatment the expected profit-maximizing wage is neither significantly different from the offered wage (Wilcoxon signed-rank test, $p = 0.73$), nor from the guessed profit-maximizing wage (Wilcoxon signed-rank test, $p = 0.45$). However, all these three wages are significantly lower than the real profit-maximizing wage in the GE treatment (Wilcoxon signed-rank tests, $p = 0.01$ compared to the offered wage and the

[23] Naturally, one criticism here may be that expected efforts are not the main driver of principals' wage-setting. We show in Fig. A.4 in the Online Appendix that this is indeed the case. In particular, the figure shows that principals with more optimistic beliefs offer higher wages in both treatments.

[24] Note that while the guessed profit-maximizing wage is the optimal wage guessed by the principal, the expected profit-maximizing wage is the wage that gives the highest profit given the principal's expressed beliefs about acceptance rate and effort response. While the guessed profit-maximizing wage is reported directly, the believed profit-maximizing wage is calculated based on principals' report of expected returned effort for each wage level.

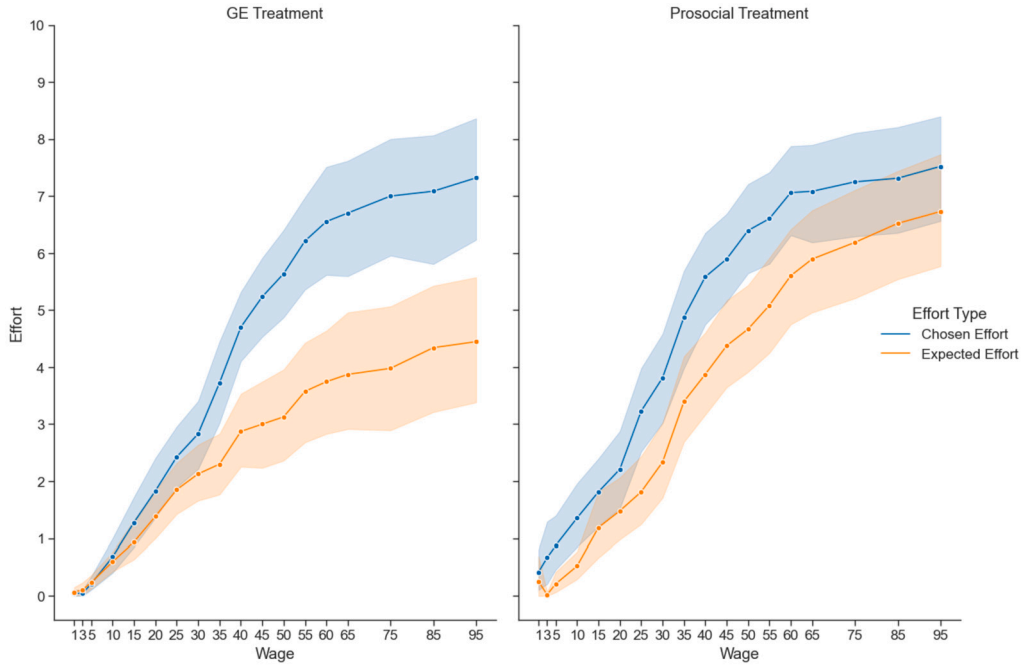

Fig. 4. Mean chosen and expected effort for each wage across treatments.

expected profit-maximizing wage and $p = 0.02$ compared to the guessed profit-maximizing wage). This suggests that in the GE treatment principals consciously maximize profits based on their beliefs but these beliefs are wrong.

Remarkably, we do not find the same belief distortion in the prosocial treatment. The right-hand panel in Fig. 4 clearly shows that in the prosocial treatment the difference between the real and the expected effort is not strictly increasing with the wage. Rank-sum tests comparing the changes in chosen effort to the changes in expected effort following an increase in the wage, reveal a significant difference only at wages 3, 25, 65 and 85. Furthermore, among these four wage levels, the difference is positive for the first two and negative for the last two. The regression analysis provides similar findings. OLS regressions in Columns 3 and 4 of Table A.2 present, respectively, regressions for the chosen effort and the expected effort on the wage level in the prosocial treatment. In contrast to the GE treatment, the coefficient of the wage in Column 3 is almost the same as the coefficient of the wage in Column 4, suggesting that the linear approximation of the agents' real effort function is as steep as the principals' expected effort function. Lastly, also in the prosocial treatment, the expected profit-maximizing wage is neither significantly different from the offered wage (Wilcoxon signed-rank test $p > 0.33$), nor from the guessed profit-maximizing wage (Wilcoxon signed-rank test $p > 0.15$). However, differently from the GE treatment, the profit-maximizing wage in the prosocial treatment is not statistically different from the offered wage (Wilcoxon signed-rank test $p > 0.21$), nor from the expected profit-maximizing wage (Wilcoxon signed-rank test $p > 0.18$) and only marginally so from the guessed profit-maximizing wage (Wilcoxon signed-rank test $p = 0.07$). Hence, we conclude:

Result 4. In both treatments, principals act on average as profit-maximizers based on their effort-beliefs. The reason why the offered wage is lower in the GE treatment than in the prosocial treatment is that principals in the GE treatment have biased beliefs. This bias is such that they highly underestimate the role of reciprocity in motivating effort provision.

5. Robustness checks: efficiency or externality?

In this section we report the results of two robustness checks which serve to rule out that our results are purely driven by a "simple efficiency effect", i.e., by the simple fact that the total efficiency is higher in the prosocial treatment than in the GE treatment. Note that by definition, a job's social impact, by generating a positive externality for a third-party, necessarily implies higher efficiency. Hence, the question is whether our results are driven by the presence of the externality, i.e., by the fact that the extra surplus benefits a *third party*, or by pure efficiency, namely, by the size of the total surplus independent of who benefits from it. To answer this question, we ran two additional treatments, each of which has some limitations, but which, taken together, hopefully can give a coherent picture of the role of efficiency vs. externality in driving our results.

5.1. Treatment with a neutral third-party

In the first new treatment (henceforth, "neutral treatment"), we replace the charity chosen by the agents with a "neutral third-party" as the new recipient of the donation while keeping everything else constant.[25] The neutral third-party was selected as follows: In the first stage of the experiment, the agents played five standard dictator games, in each of which they had to allocate points between themselves and a student from another experiment or between themselves and each of four other relatively unknown charities whose cause is unlikely to be particularly appealing to German students in Cologne.[26] In the second stage of the experiment, subjects were again randomly divided into agents and principals. For each principal-agent pair, the recipient who received the least number of points in the dictator games by the agent was then selected to be the neutral third-party for that pair. In case of ties, the recipient was randomly selected among those who received the least numbers of points in the dictator game.[27]

We conducted the treatment in May 2022 at the experimental lab of the University of Cologne. We recruited our subjects from the same subject pool as before, but excluded those that had participated in an earlier session. In total, 90 subjects took part in the "neutral treatment". A limitation of this treatment is that it is just a weaker version of the prosocial treatment: the externality towards the third-party is still present. However, the main difference between this treatment and the prosocial treatment, as also described in previous studies with a similar design (Cassar, 2019), is that the subjects should care less about the charities involved (i.e., they should be less prosocially motivated) in the former than in the latter treatment. This is indeed the case: in the dictator game, the number of points that agents allocate to the charity used in the neutral treatment is significantly lower than the respective number of points allocated to the charity used in the prosocial treatment ($p < 0.02$). Hence, based on subjects' behavior in the dictator games preceding the main part of the experiment, subjects are less prosocially motivated (i.e., in terms of the model, m is lower) in the neutral than in the prosocial treatment.

We start by comparing the agents' behavior in the neutral treatment to their respective behavior in the prosocial and GE treatments. As can be seen in Table A.4, consistent with our Result 1, agents are less (more) likely to accept a wage below their outside option in the neutral treatment than in the prosocial (GE) treatment. However, the differences remain insignificant. As far as effort is concerned, Fig. A.2 gives a graphical representation of the agents' effort function in each of the three treatments. As can be seen, starting from wage 25, the agents' effort function in the neutral treatment lies very close to the respective function in the GE treatment and below the respective function in the prosocial treatment.[28] Consistent with this graphical overview and with our Result 2, Table A.6 shows that for any given wage, agents exert lower (higher) effort in the neutral treatment than in the prosocial (GE) treatment – although these differences remain again insignificant. Taken together, these results suggest that the agents' behavior in the neutral treatment lies somehow "in the middle" between the behavior in the prosocial treatment and the behavior in a standard gift-exchange. The direction of the results is consistent with our previous findings, but we lose the (marginal) significance. This is not surprising given that the neutral treatment is just a weaker version of the prosocial treatment and the differences between the prosocial and GE treatments were only marginally significant since the beginning. Furthermore, previous laboratory findings by Cassar (2019) also show that the choice of the specific charity plays no additional role in motivating effort: exerted effort is the same independent of whether the charity is chosen by the agent, the principal, or randomly determined.

As far as the principals' behavior is concerned, we find that principals in the neutral treatment behave more closely to the prosocial treatment than to the GE treatment. This is not surprising given that, for the principals, the difference between the prosocial and the neutral treatment is even weaker than for the agents.[29] As can be seen in Table A.8 they offer a wage that is only slightly below the offered wage in the prosocial treatment but substantially higher than the wage offered in the GE treatment. Furthermore, a comparison between the effort chosen by the agents and the expected effort guessed by the principals (see Columns 1 and 2 in Table A.3 for the regressions' results) suggest that, in line with the findings in the prosocial treatment, principals do not underestimate agents' reciprocity towards an increase in wage. The slope of the linear approximation of the real effort function is identical to the linear approximation of the slope of the principals' expected effort function. Finally, similar to the prosocial treatment, the offered wage in the neutral treatment is not different from the guessed profit-maximizing wage (Wilcoxon signed-rank test, $p > 0.91$), namely, from the wage that the principals believe maximize their profits, and the offered wage is only marginally higher than the real profit-maximizing wage (Wilcoxon signed-rank test, $p = 0.06$). Thus, overall, it is clear that principals in the neutral treatment – similarly to their counterpart in the prosocial treatment – do not have biased beliefs about the agents' effort function. This suggests that the behavior of principals is not driven by the choice of the charities. These new findings, however, cannot really tell us whether it is efficiency or the presence of the externality itself that drives the differences observed between the prosocial and GE treatment. To investigate this issue further, we ran a second new treatment, which we describe below.

[25] We thank two anonymous referees for suggesting this additional treatment.

[26] The causes of these charities were the fight against addiction (Deutsche Hauptstelle fuer Suchtfragen e.V.), the preservation and restoration of historic organs in Protestant churches (Stiftung Orgelklang), the species-appropriate husbandry of rats (Rattenhilfe Nordwest e.V.), and the improvement of the lives of people with adiposity (Adipositas Stiftung Deutschland).

[27] Note that, in Stage 1 of the experiment, subjects still didn't know what would happen in Stage 2. In Stage 1, they were simply informed that after Stage 1 is over, they will receive the instructions for Stage 2, and that either Stage 1 or Stage 2 will count for payment. Principals were informed how the neutral third-party was selected.

[28] See Fig. A.7 for the individual effort choice function of each agent in neutral treatment. As can be seen, there are about 8 non-reciprocal agents.

[29] For the principals, the only difference is that in the prosocial (neutral) treatment principals are told that the recipient of the donation is the one who was allocated the highest (lowest) number of point form the dictator game.

5.2. Efficiency treatment with no externality

In the second new treatment (henceforth, "efficiency treatment"), we keep the same overall efficiency as in the prosocial treatment but remove the externality completely as in the GE treatment. This is achieved by modifying the principals' payoff from (10*effort - wage) in the prosocial treatment to (35*effort - wage) in the efficiency treatment. Given that in the prosocial treatment the agent's effort also generates a donation to a third-party equal to (25*effort), total efficiency is the same in both treatments. However, as in the GE treatment, in the efficiency treatment there is no social impact/externality/third party because the "extra generated surplus" (i.e., 25*effort) is given to the principal. This mimics well the difference between "equally efficient" organizations, one in which the entire surplus goes to the employer (standard profit-maximizing organization), and the other one in which part of the surplus is devoted to the social good (social enterprises).[30] We conducted the "efficiency treatment" in October 2022 at the experimental lab of the University of Cologne. We recruited our subjects from the same subject pool as before, but excluded those that had participated in an earlier session. In total, 120 subjects took part in the "efficiency treatment".

The regression results on the agents' acceptance wages are reported in Table A.5. Consistent with our Result 1, we find that agents are willing to accept a significantly lower wage in the prosocial treatment than in the efficiency treatment ($p < 0.01$). This holds true more generally, and not only for the wages below the outside options. As far as effort is concerned, Fig. A.2 gives a graphical representation of the agents' effort function in each of the treatments.[31] As can be seen, the agents' effort function in the efficiency treatment lies well below the respective function in the prosocial treatment in the whole wage domain. Consistent with this graphical overview and with our Result 2, Table A.7 further shows that for any given wage, agents exert significantly lower effort in the efficiency treatment than in the prosocial treatment ($p \leq 0.016$ for all comparisons). We also find that the agents' effort function in the efficiency treatment is slightly flatter than in the prosocial treatment but the effect is very small, with the coefficient of the interaction between the treatment dummy and wage being very close to zero (-0.018) and only significant in the random effects linear regressions but not in the Tobit regressions. So overall, all these new findings are very much consistent with our original findings concerning the agents' behavior. We now turn to the principals' behavior.

Based on the theoretical predictions, given that the principals' returns from the agents' effort are higher in the efficiency treatment than in the prosocial treatment, offered wages should be higher in the former than in the latter. Fig. A.3, which depicts the four types of wages – offered wage, profit-maximizing wage, expected profit-maximizing wage and guessed profit-maximizing wages – within each of our four treatments, shows that this is indeed case. Similar results are found in the regressions reported in Columns 1 and 2 of Table A.9. However, Columns 3 and 4 of the same table also show that once we control for the principals' charity motivation (elicited through the dictator game with charity at the beginning of the experiment), the difference in wages across the two treatments is no longer significant.[32] Relatedly, Fig. A.3 also shows that the four types of wages in the efficiency treatment share a similar pattern than the four types of wages in the GE treatment, which is different from the pattern within the prosocial and the neutral treatments. As in the GE treatment, principals in the efficiency treatment offer a wage which is neither different from their guessed profit-maximizing wage (Wilcoxon signed-rank test $p = 0.67$) nor from their expected profit-maximizing wage (Wilcoxon signed-rank test $p = 0.14$) but which is, in fact, 17 points (significantly) lower than the actual profit-maximizing wage (Wilcoxon signed-rank test $p < 0.01$). This explains why, even though according to the theory, wages should be higher in the efficiency treatment than in prosocial treatment after controlling for principals' characteristics, this is not the case. Principals in the efficiency treatment are leaving some profits on the table. As in the GE treatment, this seems to be due to the principals in the efficiency treatment underestimating the slope of the agents' effort function. Columns 3 and 4 in Table A.3 present the respective chosen effort and expected effort regressions on wage levels in the efficiency treatment. The wage coefficient in Column 3 is 17 percent larger than the wage coefficient in Column 4, suggesting that, similarly to the GE treatment, the linear approximation of the agents' real effort function is steeper than the linear approximation of the principals' average expected effort function. In other words, principals in the efficiency treatment underestimate agents' reciprocity – albeit to a lower extent than in the GE treatment – which lead them to offer a wage below the profit-maximizing level.

Taken together, the results from these two robustness-checks suggest that our findings are unlikely to be driven by a simple efficiency effect but rather by the presence of a positive externality. These results, however, are independent from the specific third-party who benefits from this externality.

6. Discussion

In this section we discuss (i) the link between the model and the experiment; (ii) limitations on causality; (iii) the potential extension to a multiple round setting with learning opportunities and (iv) the potential extensions to a setting with endogenous and costly social impact of the job and selection.

[30] The downside of this treatment, however, is that it may trigger some inequity aversion that is specific to the principal but not to the third party benefiting from the externality. Note that also in the prosocial and neutral treatments agents could be inequity averse towards the third-party benefitting from the externality. The comparison with the efficiency treatment becomes more difficult if we assume that there is some inequity aversion which is specific to the principal and thus not comparable with inequity aversion towards a third party.

[31] See Fig. A.8 for the individual effort choice function of each agent in efficiency treatment. As can be seen, there are about 10 non-reciprocal agents.

[32] Principals in the efficiency treatment seem indeed to be more charity-motivated than principals in the prosocial treatment based on their allocations in the dictator game with charity (two-sample Kolmogorov-Smirnov test $p = 0.09$).

6.1. Link between the theoretical model and the experiment

In section 1 we present a theoretical model whose predictions we tested in a follow-up experiment. As the purpose of the model was to provide a theoretical benchmark, we took the point of view of profit-maximizing managers; that is, we assumed that the principals could correctly predict agents' effort responses and that they did not care about the job's social impact. Our experimental findings, however, unexpectedly revealed that in the absence of the positive externality, individuals tend to underestimate other people's reciprocity. This discrepancy between the theoretical predictions and the experimental findings should not be interpreted as the experiment not being sufficiently related to the theoretical model. The purpose of the model was not to explain the experimental findings. Rather, the experimental findings help us to understand how human behavior in organizations can deviate from a previously defined theoretical benchmark.

6.2. Limitations on causality

As in many other studies in the field of prosocial motivation (for a recent example see Jones et al. (2023)), our initial study compares treatments with and without a positive externality and, thus, cannot cleanly disentangle the effects of the positive externality from those of the efficiency differences that are generated by the externality. Therefore, we go one step further and, as a robustness check, also report the results of an "efficiency" treatment in which we remove the positive externality but keep efficiency constant (by increasing the payoff of the principal). We find that effort levels, offered wages, and the principals' beliefs are very different across the prosocial and the efficiency treatments, which suggests that our initial results are not purely driven by efficiency differences. That being said, agents might have different inequity aversion preferences towards the principal compared to the charity, which might contribute to explaining these treatment differences. Hence, even with this additional robustness check, our study cannot fully address the endogeneity issue and shares this inherent limitation with the rest of the literature. It is worth pointing out, however, that, from a policy perspective, distinguishing the effects of externality vs efficiency is less relevant given that, outside of the laboratory, a positive externality by definition entails efficiency gains. Additionally, if those efficiency gains are not enjoyed by society or other third parties (as in our prosocial treatment), they are typically distributed to shareholders (as in our efficiency treatment), which inevitably leads to inequity concerns.

6.3. Multiple rounds and learning opportunities

This paper focuses only on a one-shot interaction between the agent and the principal and, therefore, fails to capture long-term relationships between an employer and employee. This design choice was necessary in order to rule out reputation concerns. However, a potential extension of our study would be to allow multiple rounds and to rematch principals and agents at each round. While this design feature would not add external validity in terms of capturing long-term relationships, it would allow some learning opportunities to take place. Note, however, that even if agents and principals interact over a longer time-frame, biased beliefs are likely to persist. In particular, an overly pessimistic principal will not offer a higher wage and will thus never receive the counterfactual information that would lead her to offer higher wages. To that end, Caria and Falco (2022) show, with a lab experiment in urban Ghana between entrepreneurs and workers, that (i) principals significantly underestimate workers' trustworthiness in a real-effort task, and (ii) the low expectations persists.

6.4. Endogenous purpose, CSR and selection

Consistent with the seminal theoretical paper by Besley and Ghatak (2005) and its subsequent implementations in the laboratory (Cassar, 2019), in our setting the social purpose of the job is exogenous and set to be equal to the agents' preferred charities.[33] This is meant to capture the situation in which workers derive an intrinsic benefit from doing their job; from the social value they create in providing a collective good. Examples, as stated by Besley and Ghatak (2005) and Cassar (2019), include: doctors who are committed to saving lives, researchers to advancing knowledge, judges to promoting justice, teachers to transferring knowledge and values to students, journalists to reporting information to the world, engineers to promoting green technology and so on. These workers are employed in organizations that may share their social propose to a larger or to a smaller extent: hospitals, as well as schools and companies, can be more or less driven by profit-maximization than by social value creation. We derived theoretical predictions taking as benchmark the situation in which the employer is profit-maximizer. However, in our experiment, the principals' preferences are obviously endogenous (i.e., the subjects in the role of the principals may or may not care to generate a donation to the agents' chosen charity) but we elicit them and can control for them in the empirical analysis thanks to the dictator game played in the first stage of the experiment. We decided to let the agents, rather the principals, choose the charity because the aim of the study was to look at the effect of efficiency wages when agents are prosocially motivated and, therefore, it was important for us that the agents in our experiment cared about the charities. This being said, the earlier experimental study by Cassar (2019) – which has a similar setting to the one of this paper but where the principals choose piece-rates instead of efficiency wages – finds that the choice of the charity (whether it is chosen by agent, randomly determined, or chosen by the principal) has no effect on the effort provided by the agents

[33] Thus, differently from Cassar and Armouti-Hansen (2020) and Besley and Ghatak (2017), this paper leaves aside questions related to the optimal choice of the mission or of the organizational form.

and on the piece-rates offered by the principals. In other words, what matters in determining behavior (at least in the laboratory) is the presence of the externality in the form of the charitable donation rather than the specific charity that receives the donation. Based on these previous results we do not think that our findings would fundamentally change if it were the employers choosing the charity. In fact, as our first robustness-check shows, even if the recipient of the donation is a neutral third-party, the results still hold.

Another characteristic of our theoretical and experimental setting is that the social impact does not entail any cost for the employer. This speaks, again, to many organizations in which employees may be more or less intrinsically motivated by the nature of their task: e.g., to schools who hire teachers who can be more or less intrinsically motivated by the purpose of transferring knowledge and values to their students. This additional intrinsic drive that some teachers might have compared to others come with no additional cost for the employer.

This being said, there exist, of course, also many settings in which the social impact is costly for the employer, e.g., in the case of an investment CSR made by a company. An extension of this study could thus be to test how the agents' reaction function – and in turn the optimal wage – vary if the principals would bear the costs of the externality generated by the agents' effort. We believe that this richer setting could lead to potentially different outcomes. A costly CSR investment might induce some CSR-reciprocity (in the form of indirect reciprocity) in addition to a wage-reciprocity if combined with a high wage offer. One might expect agents to be more reciprocal towards higher wages if the latter are also combined with some costly social initiatives. On the other hand, some agents might be more willing to punish a low wage if combined with a costly CSR investment because they might wonder why the company is willing to sacrifice money to help a social cause but not to increase their wage.[34] Then the question arises if employers can correctly predict these workers' responses or whether new biases emerge. New theoretical and experimental analyses should be conducted to answer these questions.

Finally, an additional natural extension of this study is to allow agents to self-select into organizations with or without a social impact. In fact, a job's social purpose plays a major role in attracting a prosocial workforce (Kosfeld and von Siemens, 2009; Fehrler and Kosfeld, 2013; Friebel et al., 2019). If the principals can anticipate this, one should expect our findings to become even stronger once we allow self-selection into organizations.

7. Conclusion

In this paper, we provide a first attempt at analyzing, both theoretically and experimentally, the interaction between two non-financial motives that are very common in the workplace and that have received high attention by the economic literature: workers' motivation to reciprocate efficiency wages and workers' prosocial motivation to contribute to the social good. Using a laboratory experiment instead of field data has several advantages. First, it addresses most endogeneity issues head-on: we are able to disentangle the effects of efficiency wages from those of the job's social impact on the agents' effort and on the principals' wage offers. We find that the job's social impact and efficiency wages are independent in motivating agents' effort and that, contrary to the resulting theoretical predictions, the job's social impact increases principals' wage offers. Second, our design allows us to elicit the effort response function of each agent and the belief profile of each principal, which, in turn, allows us to gain useful insights on the principals' behavior and the presence of potential biases. We show that the treatment difference in wage offers is the result of principals underestimating agents' reciprocity in a standard gift-exchange game and, therefore, of principals offering a wage that is below the profit-maximizing wage. On the contrary, in the presence of a positive externality, principals are behaving optimally – in the sense that they are maximizing profits based on their correct beliefs. We conjecture that this might be due to principals trusting agents more to do the "right thing" (i.e., to be more reciprocal and thus to exert more effort in response to a higher wage) if their reciprocal actions also benefit a third-party.

Studying workers' motivation and wage contracting in the laboratory inevitably also has its downside. The use of the strategy method and of monetary effort levels certainly makes the environment more abstract and artificial. These design features, however, should not fundamentally affect the results (Brandts and Charness, 2011). Furthermore, the laboratory grants exogenous control over key variables, but the same control implies that many features and assumptions that are made in the theory are directly imposed. For example, our analysis does not take into account the organization's choice of the social impact or workers' self-selection into organizations with a social purpose. These and other extensions to our analysis are left to future research.

Declaration of competing interest

None.

Data availability

Data will be made available on request.

[34] A finding along these lines (albeit in a quite different context) is documented by the field experiment in Cassar and Meier (2020) which shows that if CSR is perceived to be done strategically to increase profits, it may actually backfire and reduces workers effort. The potentially negative effects of CSR are also documented by List and Momeni (2021) which shows that CSR can backfire due to moral licensing: the "doing good" nature of CSR induces workers to misbehave on another dimension that hurts the firm. Differently from this latter study, our setting does not allow multiple or multidimensional tasks and, therefore, it cannot speak to the potential effect of CSR on moral licensing.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.geb.2024.03.001.

References

Akerlof, G.A., Kranton, R.E., 2008. Identity, supervision, and work groups. Am. Econ. Rev. 98 (2), 212–217.
Andreoni, J., 1989. Giving with impure altruism: applications to charity and Ricardian equivalence. J. Polit. Econ. 97 (6), 1447–1458.
Andreoni, J., 1990. Impure altruism and donations to public goods: a theory of warm-glow giving. Econ. J. 100 (401), 464–477.
Ariely, D., Bracha, A., Meier, S., 2009. Doing good or doing well? Image motivation and monetary incentives in behaving prosocially. Am. Econ. Rev. 99 (1), 544–555.
Ashraf, N., Bandiera, O., Jack, B.K., 2014. No margin, no mission? A field experiment on incentives for public service delivery. J. Public Econ. 120, 1–17.
Bartling, B., Fehr, E., Schmidt, K.M., 2012. Screening, competition, and job design: economic origins of good jobs. Am. Econ. Rev. 102 (2), 834–864.
Benabou, R., Tirole, J., 2003. Intrinsic and extrinsic motivation. Rev. Econ. Stud. 70 (3), 489–520.
Bénabou, R., Tirole, J., 2006. Incentives and prosocial behavior. Am. Econ. Rev. 96 (5), 1652–1678.
Besley, T., Ghatak, M., 2005. Competition and incentives with motivated agents. Am. Econ. Rev. 95 (3), 616–636.
Besley, T., Ghatak, M., 2017. Profit with purpose? A theory of social enterprise. Am. Econ. J.: Econ. Policy 9 (3), 19–58.
Besley, T., Ghatak, M., 2018. Prosocial motivation and incentives. Annu. Rev. Econ. 10, 411–438.
Bolton, G.E., Ockenfels, A., 2000. ERC: A theory of equity, reciprocity, and competition. Am. Econ. Rev. 91 (1), 166–193.
Brandts, J., Charness, G., 2011. The strategy versus the direct-response method: a first survey of experimental comparisons. Exp. Econ. 14 (3), 375–398.
Brown, M., Falk, A., Fehr, E., 2004. Relational contracts and the nature of market interactions. Econometrica 72 (3), 747–780.
Børsting, C., Thomsen, S., 2017. Foundation ownership, reputation, and labour. Oxf. Rev. Econ. Policy 33 (2), 317–338.
Caria, S.A., Falco, P., 2022. Skeptical employers: Experimental evidence on biased beliefs constraining firm growth. Rev. Econ. Stat., 1–45.
Cassar, L., 2019. Job mission as a substitute for monetary incentives: benefits and limits. Manag. Sci. 65 (2), 896–912.
Cassar, L., Armouti-Hansen, J., 2020. Optimal contracting with endogenous project mission. J. Eur. Econ. Assoc. 18 (5), 2647–2676.
Cassar, L., Meier, S., 2018. Nonmonetary incentives and the implications of work as a source of meaning. J. Econ. Perspect. 32 (3), 215–238.
Cassar, L., Meier, S., 2020. Intentions for doing good matter for doing well: the negative effects of prosocial incentives. Econ. J. 131 (637), 1988–2017.
Charness, G., Kuhn, P., 2011. Lab labor: What can labor economists learn from the lab?. In: Handbook of Labor Economics, vol. 4. Elsevier, pp. 229–330.
Charness, G., Rabin, M., 2002. Understanding social preferences with simple tests. Q. J. Econ. 117 (3), 817–869.
Delfgaauw, J., Dur, R., 2007. Signaling and screening of workers' motivation. J. Econ. Behav. Organ. 62 (4), 605–624.
Delfgaauw, J., Dur, R., 2008. Incentives and workers' motivation in the public sector. Econ. J. 118 (525), 171–191.
DellaVigna, S., List, J.A., Malmendier, U., Rao, G., 2022. Estimating social preferences and gift exchange at work. Am. Econ. Rev. 112 (3), 1038–1074.
Dufwenberg, M., Kirchsteiger, G., 2004. A theory of sequential reciprocity. Games Econ. Behav. 47 (2), 268–298.
Esteves-Sorenson, C., 2017. Gift exchange in the workplace: addressing the conflicting evidence with a careful test. Manag. Sci. 64 (9), 4365–4388.
Falk, A., Fischbacher, U., 2005. Modeling strong reciprocity. In: Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life, vol. 6, p. 193.
Fehr, E., 2009. On the economics and biology of trust. J. Eur. Econ. Assoc. 7 (2–3), 235–266.
Fehr, E., Falk, A., 1999. Wage rigidity in a competitive incomplete contract market. J. Polit. Econ. 107 (1), 106–134.
Fehr, E., Falk, A., 2002. Psychological foundations of incentives. Eur. Econ. Rev. 46 (4–5), 687–724.
Fehr, E., Fischbacher, U., Kosfeld, M., 2005. Neuroeconomic foundations of trust and social preferences: initial evidence. Am. Econ. Rev. 95 (2), 346–351.
Fehr, E., Gächter, S., 1998. Reciprocity and economics: the economic implications of homo reciprocans. Eur. Econ. Rev. 42 (3–5), 845–859.
Fehr, E., Gächter, S., 2000. Fairness and retaliation: the economics of reciprocity. J. Econ. Perspect. 14 (3), 159–181.
Fehr, E., Gächter, S., Kirchsteiger, G., 1997. Reciprocity as a contract enforcement device: experimental evidence. Econometrica, 833–860.
Fehr, E., Goette, L., Zehnder, C., 2009. A behavioral account of the labor market: the role of fairness concerns. Annu. Rev. Econ. 1 (1), 355–384.
Fehr, E., Kirchler, E., Weichbold, A., Gächter, S., 1998a. When social norms overpower competition: gift exchange in experimental labor markets. J. Labor Econ. 16 (2), 324–351.
Fehr, E., Kirchsteiger, G., Riedl, A., 1993. Does fairness prevent market clearing? An experimental investigation. Q. J. Econ. 108 (2), 437–459.
Fehr, E., Kirchsteiger, G., Riedl, A., 1996. Involuntary unemployment and non-compensating wage differentials in an experimental labour market. Econ. J. 106 (434), 106–121.
Fehr, E., Kirchsteiger, G., Riedl, A., 1998b. Gift exchange and reciprocity in competitive experimental markets. Eur. Econ. Rev. 42 (1), 1–34.
Fehr, E., Schmidt, K.M., 1999. A theory of fairness, competition, and cooperation. Q. J. Econ., 817–868.
Fehrler, S., Kosfeld, M., 2013. Can you trust the good guys? Trust within and between groups with different missions. Econ. Lett. 121 (3), 400–404.
Fehrler, S., Kosfeld, M., 2014. Pro-social missions and worker motivation: an experimental study. J. Econ. Behav. Organ. 100, 99–110.
Fischbacher, U., 1999. z-tree: Experimental Software. University of Zurich.
Friebel, G., Kosfeld, M., Thielmann, G., 2019. Trust the police? Self-selection of motivated agents into the German police force. Am. Econ. J. Microecon. 11 (4), 59–78.
Gallus, J., Frey, B.S., 2016. Awards: a strategic management perspective. Strateg. Manag. J. 37 (8), 1699–1714.
Gneezy, U., List, J.A., 2006. Putting behavioral economics to work: testing for gift exchange in labor markets using field experiments. Econometrica 74 (5), 1365–1384.
Guido, A., Martinez-Marquina, A., Rholes, R., 2020. Information asymmetry and beliefs reveal self interest not fairness. GREDEG Working Paper Series.
Handy, F., Katz, E., 1998. The wage differential between nonprofit institutions and corporations: getting more by paying less? J. Comp. Econ. 26 (2), 246–261.
Hennig-Schmidt, H., Sadrieh, A., Rockenbach, B., 2010. In search of workers' real effort reciprocity—a field and a laboratory experiment. J. Eur. Econ. Assoc. 8 (4), 817–837.
Hurley, T.M., Shogren, J.F., 2005. An experimental comparison of induced and elicited beliefs. J. Risk Uncertain. 30 (2), 169–188.
Ichniowski, C., Shaw, K., Prennushi, G., 1997. The effects of human resource practices on manufacturing performance: a study of steel finishing lines. Am. Econ. Rev. 87 (3), 291–313.
Jones, D.B., 2015. The supply and demand of motivated labor: when should we expect to see nonprofit wage gaps? Labour Econ. 32, 1–14.
Jones, D.B., Tonin, M., Vlassopoulos, M., Winichakul, K.P., 2023. Paying for what kind of performance? Performance pay, multitasking, and sorting in mission-oriented jobs. Games Econ. Behav. 142, 480–507.
Kosfeld, M., Neckermann, S., 2011. Getting more work for nothing? Symbolic awards and worker performance. Am. Econ. J. Microecon. 3 (3), 86–99.
Kosfeld, M., Neckermann, S., Yang, X., 2017. The effects of financial and recognition incentives across work contexts: the role of meaning. Econ. Inq. 55 (1), 237–247.
Kosfeld, M., von Siemens, F.A., 2009. Worker self-selection and the profits from cooperation. J. Eur. Econ. Assoc. 7 (2–3), 573–582.
Kvaløy, O., Nieken, P., Schöttner, A., 2015. Hidden benefits of reward: a field experiment on motivation and monetary incentives. Eur. Econ. Rev. 76, 188–199.
Leete, L., 2001. Whither the nonprofit wage differential? Estimates from the 1990 census. J. Labor Econ. 19 (1), 136–170.
Levine, D.K., 1998. Modeling altruism and spitefulness in experiments. Rev. Econ. Dyn. 1 (3), 593–622.
List, J.A., 2009. Social preferences: some thoughts from the field. Annu. Rev. Econ. 1 (1), 563–579.

List, J.A., Momeni, F., 2021. When corporate social responsibility backfires: theory and evidence from a natural field experiment. Manag. Sci. 67 (1), 8–21.
Mas, A., Pallais, A., 2017. Valuing alternative work arrangements. Am. Econ. Rev. 107 (12), 3722–3759.
Mocan, H.N., Tekin, E., 2003. Nonprofit sector and part-time work: an analysis of employer-employee matched data on child care workers. Rev. Econ. Stat. 85 (1), 38–50.
Murdock, K., 2002. Intrinsic motivation and optimal incentive contracts. Rand J. Econ. 33 (4), 650–671.
Prendergast, C., 2008. Intrinsic motivation and incentives. Am. Econ. Rev. 98 (2), 201–205.
Preston, A.E., 1988. The effects of property rights on labor costs of nonprofit firms: an application to the day care industry. J. Ind. Econ. 36 (3), 337–350.
Rabin, M., 1993. Incorporating fairness into game theory and economics. Am. Econ. Rev., 1281–1302.
Ruhm, C.J., Borkoski, C., 2003. Compensation in the nonprofit sector. J. Hum. Resour. 38 (4), 992–1021.
Schwerter, F., Zimmermann, F., 2020. Determinants of trust: the role of personal experiences. Games Econ. Behav. 122, 413–425.
Shapiro, C., Stiglitz, J.E., 1984. Equilibrium unemployment as a worker discipline device. Am. Econ. Rev. 74 (3), 433–444.