KIEL
WORKING PAPER

Kind or contented? An investigation of the gift exchange hypothesis in a natural field experiment in Colombia

No. 2199 November 2021

Francesco Bogliacino, Gianluca Grimalda, David Pipke

Kiel Institute for the World Economy
ISSN 1862–1155

ABSTRACT

KIND OR CONTENTED? AN INVESTIGATION OF THE GIFT EXCHANGE HYPOTHESIS IN A NATURAL FIELD EXPERIMENT IN COLOMBIA

Francesco Bogliacino, Gianluca Grimalda, David Pipke[1]

November 2021

The gift exchange hypothesis postulates that workers reciprocate above market-clearing wages with above-minimum effort. This hypothesis has received mixed support in dyadic employer-worker relationships. We present a field-experimental test to assess this hypothesis in the context of a triadic relationship in which only one out of two workers receives a pay increase. We conjecture that inequality aversion motivations may thwart positive reciprocity motivations and analyze the interaction between such motivations theoretically. Across three treatments, the pay increase is justified to workers based on either relative merit or relative need or was arbitrary as no justification was offered. Two conditions in which either none or both workers receive a bonus serve as the reference. In contrast to the gift exchange hypothesis, we find that pay increases lead to a decrease in productivity. Such a decrease is most sizable in the condition where both workers receive the bonus. A post-diction of this result is that workers interpret the monetary bonus as a signal of the employer's contentment with their effort, which makes them feel entitled to reduce their effort. In other treatments, receiving the pay increase while the coworker does not has a positive effect on productivity, especially when the pay increase is based on merit. This result is consistent with statusseeking preferences rather than aversion against advantageous inequality. Conversely, not receiving the pay increase while the coworker does, leads to lower productivity, especially when the pay increase is assigned based on relative needs.

Keywords: Gift exchange, employer-worker relationship, pay inequality, field experiment, reciprocity, labor market, effort provision, fairness, wage inequality

Francesco Bogliacino
Universidad Nacional de Colombia
Cra 45, Bogotá, Colombia
Email: fbogliacino@unal.edu.co
www. https://unal.edu.co/en.html

Gianluca Grimalda
Kiel Institute for the World Economy
Kiellinie 66
D-24105 Kiel, Germany
Email: gianluca.grimalda@ifw-kiel.de
www.ifw-kiel.de

David Pipke
Kiel Institute for the World Economy
Kiellinie 66
D-24105 Kiel, Germany
Email: david.pipke@ifw-kiel.de
www.ifw-kiel.de

The responsibility for the contents of this publication rests with the author, not the Institute. Since working papers are of a preliminary nature, it may be useful to contact the author of a particular issue about results or caveats before referring to, or quoting, a paper. Any comments should be sent directly to the author.

[1]Francesco Bogliacino acknowledges funding from Fundación Universitaria Konrad Lorenz (Internal Grant 7INV1141: "Desigualdad, aspiraciones y esfuerzo: un análisis experimental"), Open Evidence (Research Grant 009_Inequalities "Desigualdad, Aspiraciones y esfuerzo: Un análisis experimental"), and Universidad Nacional (Internal Grant 26007 "Desigualdad, Aspiraciones y Compromiso Individual"). Suelen Castiblanco, Laura Jiménez and Daniel Reyes provide excellent research assistance. We appreciate comments and suggestions from all the participants at the BEBES seminar in Bogotá, the Brown Bag seminar at Kiel University, and the FAIR Workshop at NHH.

1 Introduction

According to standard competitive market theory, workers are employed at the market-clearing wage and provide the minimal effort required not to get fired by the firm (Lazear, 2000). Several economic mechanisms such as efficiency wages (Katz, 1986), implicit contracts (Azariadis, 1975), and insider-outsider relationships (Lindbeck and Snower, 1988) have been proposed to explain evidence at odds with this prediction. Another such mechanism is eminently psychological and rests on the notion of "reciprocity" (Rabin, 1993). That is the disposition to repay kind actions with kindness and spiteful actions with spite (Fehr and Gächter, 1998, 2000). The seminal work by Akerlof (1982) and Akerlof and Yellen (1990) reported anecdotal evidence consistent with what was called a "gift exchange" (Adams, 1963). Firms paying wages higher than the market-clearing level would be repaid by workers through the provision of effort above the minimal possible. The gift exchange hypothesis has received support in both laboratory experiments (Fehr, Kirchsteiger and Riedl, 1993; Fehr et al., 1998; Charness, 2004) and controlled experiments conducted in natural settings (Gneezy and List, 2006; Cohn, Fehr and Goette, 2015; Gilchrist, Luca and Malhotra, 2016; Englmaier and Leider, 2020). However, some natural experiments have suggested that such productivity gains may be transient (Gneezy and List, 2006; Bellemare and Shearer, 2009) or even negligible (Kube, Maréchal and Puppe, 2012, 2013; Esteves-Sorenson, 2018; DellaVigna et al., 2021). It is still an open question in which settings the gift exchange hypothesis actually drives individual behavior at the workplace.

Two topics have not been systematically investigated in this literature. Firstly, the firm may be willing to increase the wages only to some of its workers. This would be the case should the firm's resources be insufficient to increase all workers' wages. Or it could be a firm's goal to reward the most productive workers. In these cases, it is unclear how workers will react to selective rewards. To a large extent, experimental studies on gift exchange have only analyzed dyadic relationships where only one principal and one agent interact. Such studies are then silent on how horizontal inequality affects the gift exchange. Self-regarding individuals should be indifferent to what happens to others. However, both experimental studies and surveys suggest that many individuals are inequality-averse. That is, they experience envy if someone else is rewarded more than themselves and compassion otherwise (Fehr, Goette and Zehnder, 2009; Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000). If inequality aversion is strong enough, then both disadvantaged and advantaged workers may react negatively to the firm offering selective rewards, thus reducing their effort. Such a strategy may then backfire on the firm. Secondly, people are generally found to be more tolerant of inequality when this can be justified on the grounds of fairness (Konow, 2003). Typically, when inequality is caused by a worker being more productive than another because of their higher effort or ability, the resulting inequality is more accepted than when it is caused by luck (Krawczyk, 2010; Alesina and Giuliano, 2011; Durante, Putterman and van der Weele, 2014; Almås, Cappelen and Tungodden, 2020). However, even if the firm tried to follow fair criteria to select who is the most deserving worker to reward, a self-serving bias[1] may cause unrewarded workers not to recognize the fairness of the process. Workers' resentment may lead to shirking or sabotage (Bewley, 1999). Conversely, if workers find the firm justification of rewards as fair, they may be willing to respond with positive reciprocity even if the reward is directed to another worker. It is then important to ascertain how individuals weigh up reciprocity and inequality aversion with respect to different possible sources of inequality, which may be perceived as either fair or unfair. The existing literature has however only analyzed one possible source of wage inequality at a time, typically focusing on pay disparity that appeared arbitrary and not explicitly linked to individual merit (Hennig-Schmidt, Sadrieh and Rockenbach, 2010; Cohn et al., 2014).

The present study provides evidence on these topics from a controlled field experiment conducted in a natural work environment. We invited pairs of workers to a university department for a one-day data-entry job to be conducted individually. After the morning session, the job instructor announced a surprise bonus payment for one worker, worth one-third of the previously announced earnings. We manipulated the source of earnings inequality across three treatments. The bonus was assigned based on (a) relative productivity (with the most productive worker in the morning session receiving the bonus) in the Productivity treatment; (b) relative needs (evaluated through a measure of participants socio-economic status) in the Needs treatment; or (c) no justification (a method which we expected as being perceived as arbitrary by workers) in the Arbitrary treatment. These three treatment conditions are contrasted with two baseline conditions in which (d) no worker receives a bonus (Control condition), or (e) both workers receive the bonus (Double Bonus condition). In this way, we can weigh up reciprocity and inequality aversion motivations as a response to the bonus by both the bonus recipient and the nonrecipient. We do this for three different sources of inequality, which we expected to be viewed as fair (in the Productivity treatment), unfair (in the Arbitrary treatment), or morally acceptable (in the Needs treatment). Assigning bonuses based on relative needs has rarely been investigated.[2] We conjectured that this method may have stirred moral approval by both workers, thus spurring workers' dispositions to positive reciprocity.

This study contributes to the growing literature studying labor market interactions via field experiments (Harrison and List, 2004; List and Rasul, 2011; Al-Ubaydli and List, 2017). Identifying causal effects from firm-level data would be confounded by the presence of many unobservable factors that could drive an observed correlation between effort and wages. Firms' strategies are equilibrium responses to exogenous conditions that are jointly determined with workers' behavior.

[1]Miller and Ross (1975) define the self-serving bias as individual psychological dispositions whereby "[...] people indulge both in self-protective attributions under conditions of failure and in self-enhancing attributions under conditions of success". A self-serving bias, then, prompts individuals to attribute their failure to situational factors, and their success to their own abilities (Deffains, Espinosa and Thöni, 2016). See also Mezulis et al. (2004) and Babcock and Loewenstein (1997).

[2]In his discussion of distributive justice, Konow (2003) suggests that assigning resources on the basis of relative needs is one principle, among several others, that is normally seen as fair by individuals. Nicklisch and Paetzel (2020) provide experimental support to this claim.

In particular, one would not be able to rule out various unobservable drivers of work morale, such as reputational concerns, social norms, rules of behavior, managerial practices, and interpersonal relationships, especially at the shop-floor level and with regards to motivations and intentions. Even when we could observe discrete changes in remuneration policy, a selection problem in identifying a proper counterfactual would arise because multiple omitted variables may vary at the same time of the treatment across organizations. Our experimental approach controls for these confounds and allows for causal inference.

The gift exchange hypothesis has received extensive support in the context of laboratory experiments (Fehr, Kirchsteiger and Riedl, 1993; Fehr and Falk, 1999; Charness, 2004; Charness, Frechette and Kagel, 2004) and some support in field studies (Cohn, Fehr and Goette, 2015; Gilchrist, Luca and Malhotra, 2016; Englmaier and Leider, 2020). However, some other studies utilizing data-entry tasks with student workers found only statistically insignificant effects of monetary pay rises (Hennig-Schmidt, Sadrieh and Rockenbach, 2010; Kube, Maréchal and Puppe, 2012; Al-Ubaydli et al., 2015; Esteves-Sorenson, 2018; DellaVigna et al., 2021) on productivity, or effects turned out to be largely transient (Gneezy and List, 2006; Bellemare and Shearer, 2009). Although Gneezy and List (2006) report initial productivity boosts in response to surprisingly higher than advertised wages in two samples of student workers, either hired to perform a data entry task for a library or a fundraising task for a charity, the effect after the "gift" waned quickly over the six hours of the experiment. While the productivity boost lost statistical significance as early as after 90 minutes among students hired to perform the data entry task, it also diminished, but remained overall statistically significant, with the fundraising task. Bellemare and Shearer (2009) report a similar finding based on an experiment in a real tree-planting firm in British Columbia. A one-day pay rise significantly increases the number of trees planted on that day compared to previous and subsequent days, especially among experienced workers. However, the long-term relationship of workers casts doubt on whether effects may be confounded by reputational concerns.

Our study is also closely related to the literature on how horizontal inequality affects productivity at the workplace. Laboratory studies examining pay inequality between two workers generally found that workers are less sensitive to the treatment of other workers compared to themselves. All the same, workers' behavior is less responsive to advantageous inequality than to disadvantageous inequality (Charness, Rigotti and Rustichini, 2007; Gächter and Thöni, 2010; Bracha, Gneezy and Loewenstein, 2015). Only very few studies studied the effects of pay inequality within natural field experiments. Hennig-Schmidt, Sadrieh and Rockenbach (2010) did not find evidence for peer comparison effects, neither for advantageous nor disadvantageous wage inequality. Cohn et al. (2014) found that workers whose wage was cut decreased effort twice as much as when both team members' wage was cut. However, this was not the case when only the coworker's wage was cut, pointing to an asymmetric effect. Breza, Kaur and Shamdasani (2018) found that wage inequality led to lower attendance and output when the productivity of coworkers was hard to observe in a large-scale field experiment with Indian manufacturing workers. However, such a negative effect of inequality disappeared when workers learned that inequality reflected differences in their baseline productivity. In summary, studies investigating selective worker rewards have done so either in a laboratory context devoid of many features characterizing real-life work relationships or in a context when only one possible source of inequality - either luck or productivity - existed. Our framework enables us to assess the relevance of varying fairness justifications for different sources of inequality in workers' response to selective pay rises.

We also contribute to the literature on fairness, legitimacy, and the behavioral responses to procedures (Kahneman, Knetsch and Thaler, 1986; Konow, 2000, 2003; Bolton, Brandts and Ockenfels, 2005; Cappelen et al., 2007; Trautmann, 2009) and to the theoretical literature using social preferences to explain labor market outcomes. Our theoretical model merges two types of social preferences that have rarely been considered jointly thus far (Charness and Rabin, 2002; Falk and Fischbacher, 2006; Cox, Friedman and Gjerstad, 2007; Cohn et al., 2014), i.e., "reciprocity" (Rabin, 1993) and inequality aversion (Fehr and Schmidt, 1999). We hypothesized that workers would be willing to "punish" (Fehr and Fischbacher, 2004) the firm with lower effort when not receiving the bonus while the other worker received the bonus. As for the bonus recipient, our model accommodates both a decrease in effort, if the worker is inequality averse as in Fehr and Schmidt (1999), or an increase in effort, if the worker desires to repay the firm for having been singled out as the bonus recipient as in the status-seeking model of Frank (1985).

Our main result is that, contrary to the gift exchange hypothesis, not only do pay rises fail to increase workers' productivity, but actually, they decrease productivity. This was most clearly the case in the Double Bonus condition where both workers received the bonus, which recorded the largest drop in productivity across conditions. It was also the case in all three treatments where only one worker received the bonus. Our post-diction is that this result is due to what we define as a *contentment effect*. That is, workers may interpret the bonus as a signal that the manager is contented with their work and is unlikely to punish or even fire them, thus feeling entitled to reduce their effort. Furthermore, we do not find evidence that advantageous pay inequality leads to inequality aversion and thus negative reciprocity. Conversely, the productivity of bonus recipients in all three treatments is higher than in the Double Bonus condition, reaching statistical significance in the Productivity treatment. This result may be interpreted as weak evidence for status-seeking preferences as opposed to aversion against advantageous pay inequality (Frank, 1985; Abbink and Sadrieh, 2009; Abbink and Herrmann, 2011; **?**). Similarly, there are no significant effects of disadvantageous inequality among nonrecipients when the bonus is assigned arbitrarily or to the more productive worker. Only when the bonus is assigned to the worker who is needier, nonrecipients significantly reduce their effort relative to the Control condition. This suggests that workers view a justification of pay inequality based on needs as inappropriate in a labor market context.

The remainder of our paper is structured as follows. The next section describes the experimental design. In section 3 we develop a theoretical model of worker effort based on which we develop our hypotheses regarding the experiment. Empirical results are presented in section 4. Section 5 discusses the results with an emphasis on their generalizability to cultural contexts different from Colombia. Section 6 concludes.

2 Experimental design and procedures

Recruitment. — Our field experiment was conducted in Colombia's capital Bogotá. Participants were recruited through advertisements at the university as well as through social networks, ensuring that our sample includes both university students and people with lower levels of education (see Appendix Table A.1 for sample characteristics). The advertisement invited people to register for a one-time work opportunity ruling out reemployment. This aspect should rule out reputational concerns and repeated game strategies. The advertisement mentioned that some basic abilities in computer work were needed. Participants were requested to send a CV as a part of their job application. The advertised hourly wage rate of 15,000 Pesos per hour (about 6 USD) stood well above the standard payment for temporary work. A pilot study ascertained that such a wage rate was necessary to ensure a smooth recruitment process and to attract applicants with the required skill level, also considering the possibility of large commuting times in Bogota. Workers' payment was independent of the number of entries that workers completed, and no mention of a desired effort level was made throughout the recruitment process or the initial induction session. The final sample contains 236 participants[3] who were involved in 126 daily sessions between October 2014 and January 2015. In 15 cases in which a recruited worker did not show up, a confederate acted as a worker.

The sessions. — The sessions were held at the Fundación Universitaria Konrad Lorenz, a university institute located in the center of Bogotá. On each working day of fieldwork, two participants were invited to come to the university at 10 am. The same female research assistant acted as an instructor in all sessions. The fact that the working sessions were part of an experiment and that pay conditions would be manipulated across sessions was concealed to participants. Either a female or a male aid was also present on a rota basis. Workers were prevented from talking with each other throughout the session to prevent effects from peer pressure or group bonding, or workers communicating information on their own productivity to the other worker (Falk and Ichino, 2006). After the second worker arrived, the instructor started explaining the job to be performed throughout the day. After instructions, workers were asked to sign an informed consent form for the handling of their personal data. The work consisted of entering the street addresses of some randomly selected phone users into an Excel spreadsheet. The collected data was needed to conduct a survey in another research project by the university. Each worker carried out this task alone in an office, whose location was unknown to the other worker. Workers used a university

[3]We had to remove one observation for which technical problems with the USB connector made it impossible to recover the exact distribution of output between the morning and the afternoon session.

computer each, which was connected via the internet to our server computer. In this way, workers' hourly output could be monitored by researchers. The morning session lasted two hours. At its end, our assistant went to workers's rooms and communicated a fifteen-minute lunch break.[4] Afterward, workers were separately reconvened to the job instructior's room. The instructor announced treatment-specific instructions for the afternoon session (see Appendix section A.5 for instructions). Treatment conditions were randomized, administered according to a sequence that was randomly selected before the start of fieldwork. The afternoon session lasted three hours. At the end of the workday, workers filled out a questionnaire inquiring about demographic characteristics and their evaluation of the work session, and were given their payment.

The treatments. — In the Control condition, work continued without any announcement during the break. In the Double Bonus condition, a bonus payment of 25,000 Pesos (about 10.5 USD, worth one-third of advertised earnings) was paid to both workers, mentioning no specific reason. In single-bonus treatments, workers were told that a 25,000 Pesos bonus would be paid to one worker. Three different justifications for the bonus assignment were provided: In the Productivity treatment, the worker who had the higher number of entries in the morning session received the bonus. In the Needs treatment, the worker residing in the block classified as less affluent according to the official evaluation of Bogotá city council[5] received the bonus. [6]In the Arbitrary treatment, a worker was assigned the bonus without giving any information on the criteria being used. It was instead specified that neither relative productivity nor relative needs determined the bonus assignment. [7]. These three treatments were meant to evaluate workers' reactions to three different methods to introduce inequality in the wage structure, expecting that acceptance of inequality would be higher in treatments with higher perceived fairness. As argued by Bewley (1999), workers' perceived fairness of the internal pay structure depends on whether pay differences are based on reasonable and impartial criteria. The Productivity treatment arguably can be deemed as reasonable and impartial, as paying a productivity premium is common in the labor market and is justifiable by the objective to reward the worker bringing the higher profit to the firm. At the other

[4]Lunch was provided for free by the researchers. The preference for a relatively short lunch break emerged during the pilot study, as workers typically preferred a short break to finish earlier.

[5]Each of Bogotá dwelling is assigned a ranking (so-called "Estratificación socioeconómica"), ranging from 1 to 6, which classifies the value and quality of the housing for the purpose of differentiating the payment of utility bills. The ranking used in the research session was taken from a copy of a utility bill, which participants were asked to bring to the work session. This ensured that the Bonus assignment in the Needs treatment was based on truthful reporting. Each participant was asked to bring such bill and to hand it over to the instructor at the beginning of the session.

[6]Such evaluation is commonly used as a proxy for socioeconomic status in Colombia (Hagenlocher et al., 2013; Martinsson, Villegas-Palacio and Wollbrant, 2015; Bogliacino, Jiménez Lozano and Reyes, 2018). In case both blocks had the same evaluation, the bonus was assigned to the worker whose block was located in the poorer district, according to the city council classification.

[7]In fact, each worker had a 50% probability of being assigned the bonus. We did not release this information because an unbiased random procedure may have been perceived as a fair criterion by workers (Bolton, Brandts and Ockenfels, 2005; Krawczyk, 2011; Trautmann, 2009; Schurter and Wilson, 2009). Our intention was instead to maximize the perceived unfairness of the procedure. The Arbitrary treatment was indeed perceived as the last fair of the three (see section 4.3)

side of the spectrum, the Arbitrary treatment is unlikely to be seens as reasonable and impartial, precisely because no reason is offered for the bonus assignment. This treatment was meant to mirror situations in the workplace where workers perceive promotions or differential treatments of workers as unfair. The Needs treatment introduces a method that is rarely used in the labor market. Nonetheless, relative needs are psychologically salient for many people (Konow, 2003) and have wide-ranging policy relevance. Allocation of resources based on needs is often invoked as a principle of distributive justice (Nicklisch and Paetzel, 2020). Affirmative action or means-based intervention may be interpreted as preferential treatments based on needs. The purpose of the Needs treatment is to investigate workers' reaction to rewards that arguably redresses real-life inequalities, and may therefore be considered equitable for reasons different from economic productivity (Konow, 2003). We conjectured that the Needs treatment would be perceived as lying in the middle of the fairness spectrum, with the Productivity and Arbitrary treatments lying at its extremes.

3 Theoretical framework

We propose a simple theoretical model of worker effort which shall guide the analysis of our experiment. In the tradition of the gift exchange hypothesis (Akerlof, 1982; Akerlof and Yellen, 1990), the model captures the idea that workers' effort choice under fixed hourly pay depends on the generosity of the wage in comparison with the standard. We also assume that individuals are concerned with horizontal pay inequality, i.e. how the own wage compares to the coworkers' wage. We build on a series of more general models that incorporate either reciprocity motives (Rabin, 1993; Dufwenberg and Kirchsteiger, 2004) or inequality aversion (Fehr and Schmidt, 1999) into individuals' utility functions .

3.1 The setup

Akin to DellaVigna et al. (2021), we propose a utility function for a worker that depends on three components:

$$U_i = w_i - c(e_i) + K_{f_i} \cdot e_i \tag{1}$$

The first two components are purely self-regarding, as they only depend on variables attaining to the agent. w_i is the wage earned by the worker and $c(e_i)$ the disutility from exerting effort. As standard, we assume an invertible convex cost function $c(e_i) = re_i^2$ that is increasing in effort and fulfills the regularity conditions ensuring a unique solution. Those are $c'(e) > 0$, $c''(e) > 0$ and $\lim_{e\to\infty} c(e) = \infty$.. The third component of (1) is other-regarding as it includes variables pertaining to other agents' payoffs. We call it the *social preferences* component. It is the product of social preferences towards the firm K_{f_i} and worker's effort e_i.

$$\begin{aligned} K_{f_i} &= \kappa + \varphi \{ l \cdot (w_i - w_{norm}) + m \cdot (w_j - w_{norm}) \} \\ &\quad + \left\{ -(I_{w_i > w_j} \cdot \beta - I_{w_j > w_i} \cdot \alpha) \cdot (w_i - w_j) \right\} \end{aligned} \tag{2}$$

The social preferences towards the firm K_{f_i} consists of three parts. The first is a constant κ which may capture unconditional value attrinuted to the firm's payoff. This may originate from altruism (Becker, 1974), social norms demanding positive effort in employment relations (Bénabou and Tirole, 2006), or even utility from doing useful work (Ariely, Kamenica and Prelec, 2008).[8] κ may also be interpreted as capturing fears of getting punished when not providing an acceptable level of effort to the employer (Lazear, 2000). In addition, we assume that workers' concerns about the employer's payoffs are affected by two terms. The first term captures the "gift exchange" Akerlof (1982) component. It expresses the sensitivity of the worker to the firm paying wages above or below an exogenous norm. Such a social norm may coincide with the market equilibrium, or with the level that was set in previous bargaining or past interactions. Wages equal to the norm are perceived as neither kind nor unkind. In general, this motivation should apply to both the worker's own wage and the other worker's wage (Akerlof, 1982). Accordingly, the higher $(w_i - w_{norm})$ and $(w_j - w_{norm})$, the higher the perceived kindness, and the stronger are the social preferences towards the firm. We assume a self-serving bias by agent i , such that the agent's own wage matters no less than the other agent's wage, i.e. $l \geq m \geq 0$. $l = m$ corresponds to agent i holding the treatment of agent j exactly on a par with the treatment of her own wage. If $l > m \geq 0$, agent i will consider her own pay as more relevant than the coworker pay's as relevant to assess judge the firm's kindness. In the case of $m = 0$ and $l > 0$, agent i will only be concerned with her own wage. According to the gift exchange hypothesis, the weight that measures the relative relevance of the gift exchange component for the social preferences is positive: $\varphi_1 > 0$. Nonetheless, one may put forward an alternative hypothesis. A worker receiving a wage higher than w_{norm} may think that the firm wants to manifest its *contentment* for the worker's behavior. The worker may in this case feel entitled to provide *less* effort, rather than *more,* possibly because the worker now has less fear of being fired . Hence, a wage rise may be taken as a justification to reduce effort, rather than to increase it. We can model the "contentment hypothesis" with $\varphi < 0$.

The second term in K_{f_i} refers to the horizontal inequality between coworkers' wages. $I_{()}$ is an indicator function that takes the value of 1 if the subscripted condition is fulfilled. Individuals are concerned with both advantageous ($w_i > w_j$) and disadvantageous ($w_i < w_j$) inequality (Fehr and Schmidt, 1999). Advantageous and disadvantageous inequalities are weighted with $-\beta$ and $-\alpha$, respectively. Fehr and Schmidt (1999) posit that individuals dislike both advantageous and disadvantageous inequality, experiencing *compassion* when earning more than the other worker,

[8] See DellaVigna et al. (2021) for a similar theoretical approach.

and *envy* when earning less. Setting $\alpha \geq 0$ and $\beta \geq 0$ accordingly, individuals perceive the firm as less kind if the wage differential is positive. Individuals may be expected to be more sensitive to disadvantageous inequality than advantageous inequality (see evidence by Loewenstein, Thompson and Bazerman, 1989; Fehr and Schmidt, 1999), so that $\alpha \geq \beta \geq 0$. Another possibility is that individuals are not driven by compassion when earning more than another individual, but rather enjoy relative status (Frank, 1985) or experience spite. In this case, individuals attach a positive utility to earning more than the other, rather than disutility. Experiments demonstrate that a non-negligible share of individuals indeed displays spiteful preferences (Abbink and Sadrieh, 2009; Abbink and Herrmann, 2011; Fehr, Glätzle-Rützler and Sutter, 2013). We can capture status concerns or spite when earning more and envy when earning less than a coworker by setting $\alpha \geq 0 \geq \beta$.

Given that (2) and e_i enter (1) multiplicatively, the worker will maximize his or her utility by exerting positive effort as long as his or her social preferences or sentiments for the firm are positive, i.e., $K_{f_i} > 0$. The utility function for the second worker j is assumed to be identical to that of worker i with reversed indices i and j. In our experimental setup, wages can only take two values. Either the wage is equal to the fixed wage w or the worker additionally receives a lump-sum bonus B such that the wage of worker i can be expressed as $w_i = w + I_{Bonus_i} \cdot B$. Worker i's maximization problem can be written as:

$$\underset{e_i \in \mathbb{R}}{\operatorname{argmax}}\, w_i - c(e_i) + K_{f_i} \cdot e_i \tag{3}$$

Maximizing utility with respect to effort gives the first order condition defining the optimal level of effort e_i^* provided by worker i:

$$\frac{\delta U_i}{\delta e_i} = -c'(e_i) + K_f = 0 \Longleftrightarrow e_i^* = c'^{-1}(K_f) = \frac{K_f}{2r} \tag{4}$$

The second-order condition is automatically satisfied since $c''(e) > 0$. The optimal effort choice depends negatively on parameter r from the quadratic effort-cost function and positively on the social preferences towards the firm. For $K_f = 0$, the optimal level of effort is zero. In the case of $K_f < 0$, the optimal effort is negative which may be interpreted as intentionally harming (e.g., by acts of sabotage) the firm's objectives.

3.2 Hypotheses

Our primary hypotheses are grounded on the assumptions that are more directly in line with the existing literature. However, our model is general enough to accommodate for alternative hypotheses. In particular, our primary hypothesis is that workers' behavior is driven by reciprocity motives akin to the gift exchange hypothesis with respect to both the own wage and the other worker's wage, i.e., $\varphi > 0$; $l > m > 0$. Our primary assumption is also that workers feel both envy and compassion (as opposed to spite) when inequality exists, i.e. $-\alpha < -\beta < 0$. Analytical derivations of the underlying predictions generated by the model can be found in the Appendix, section A.7.

The first set of hypotheses concerns how workers react to the bonus payment, regardless of inequality concerns:

- *Hypothesis 1a* (*Gift exchange*): Workers are expected to increase their effort in response to a bonus payment. Alternatively:
- *Hypothesis 1b* (*Contentment*): Workers are expected to decrease their effort in response to a bonus payment.

The comparison between the Double Bonus condition and the Control condition provides us with a clean test of the first set of hypotheses, because inequality is by construction absent. This is arguaby the condition in which the perceived kindness of the firm should be at its highest.

The second and third hypotheses concern how inequality affects the effort of the worker not receiving the bonus, and receiving the bonus, respectively. In general, aversion against horizontal pay inequality should lead to lower perceived kindness of the firm in single-bonus treatments, ceteris paribus. A disadvantaged worker will only increase effort if the satisfaction for the other worker's payoffs exceeds envy in its impact on the perceived kindness of the firm. This is equivalent to $\varphi m > \alpha_T$, where T is the index for the treatment condition. On the other hand, an advantaged worker will increase effort as long as the gift exchange component outweighs the effect of compassion. From the perspective of our model parameters, this corresponds to $\varphi_1 l > \beta_T$. Conversely, if the advantaged worker is motivated by status seeking (or spite), then she will attach positive utility to earning more than the other worker, as $\beta_T < 0$. Effort increases more for the advantaged worker than for the disadvantaged worker under the assumptions that own payoffs have a stronger weight than other's payoffs ($l > m > 0$) and that envy is a stronger psychological motivation than compassion ($\alpha > \beta$).

- *Hypothesis 2* (*Envy*): We expect aversion against disadvantageous horizontal pay inequality to lead to negative effects on bonus nonrecipients' effort because of envious feelings ($\alpha > 0$). Accordingly, nonrecipients productivity should drop in single-bonus treatments compared to the Control condition.
- *Hypothesis 3a (Compassion)*: Compassion leads to bonus recipients to decrease their output in comparison with the Double Bonus condition ($\beta > 0$).
- *Hypothesis 3b (Status seeking/Spite)*: Status-seekingleads to bonus recipients to increase their output compared to the Double Bonus condition ($\beta < 0$).

The fourth hypothesis concerns the effect of the fairness of the source of inequality on effort. According to the discussion in section (4.3), we posit the following:

- *Hypothesis 4a (Fairness Perception):* We expect the Productivity treatment and the Needs treatment to be perceived as fairer than the Arbitrary treatment, because a reason is provided for the bonus allocation in the former but not in the latter.
- *Hypothesis 4b (Fairness Relevance)*: We expect that the higher the perceived fairness, the higher the effort. In particular, effort by both bonus recipients and nonrecipients should be higher in the Productivity treatment than in the the Arbitrary treatment. Likewise, effort should be higher in the Needs treatment than in the Arbitrary treatment.[9]

4 Empirical results

We report the general patterns of the experiment results in section 4.1. Section 4.2 provides the econometric analysis and the calibration of the utility function parameters. Section 4.3 analyzes perceptions of treatment fairness and their impact on productivity. Section 4.4 performs robustness checks of the main results.

4.1 Descriptive results

Table 1 consolidates productivity within the morning and afternoon sessions, thus providing an overview of average productivity during the pre- and post-treatment phases. We use the number of typed characters as our main outcome measure because it provides the most precise measurement of individual effort. In section 4.4, we analyze the robustness of our results to using alternative outcome measures. Overall, the average number of typed characters per hour increased from roughly 971 during the morning session to about 1057 entries per hour during the afternoon session, an increase which is statistically significant (Wilcoxon matched-pairs signed-rank test: $p < 0.001$, N= 236). Since reemployment concerns were ruled out by design (see sections 2 and 5), and since we observe a significant productivity increase in the Control condition, learning must have played a role in such a productivity increase.

Figure 1 plots hourly productivity broken down by treatment and recipient status (bonus recipient vs. nonrecipient). A visual inspection of both Figure 1 and Table 1 reveals little variability in productivity during the morning shift. Indeed, non-parametric tests fail to reject the null hypothesis of equal productivity (Kruskal-Wallis rank test: $p = 0.750$, $N = 236$) across the five treatment conditions at the level of worker pairs in the morning sessions. This result suggests that

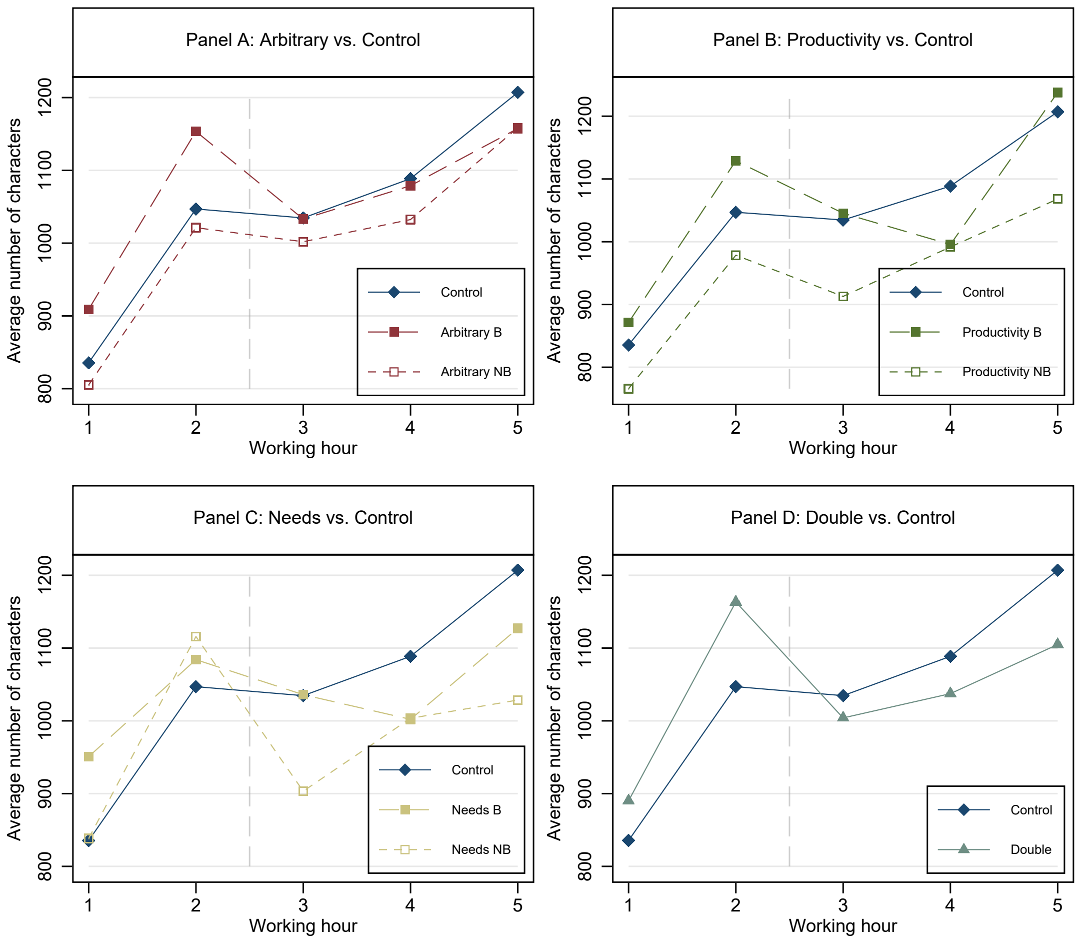
[9] Our research is exploratory as to whether productivity is higher or the same in the Productivity treatment compared to the Needs treatment.

Table 1: Average Number of Characters per Hour by Block

Condition	Morning	Afternoon	Diff.	p-value	Obs.
All workers	971.53	1057.40	85.87	< 0.01	236
	(274.44)	(285.14)	(201.70)		
Control	941.26	1110.10	168.85	< 0.01	39
	(235.46)	(262.94)	(217.12)		
Double	1026.71	1048.78	22.07	0.39	26
	(269.02)	(329.50)	(134.94)		
Arbitrary	974.37	1077.54	103.17	< 0.01	56
	(298.12)	(305.64)	(193.49)		
Arbitrary B	1031.29	1089.83	58.54	0.10	29
	(299.82)	(313.31)	(208.69)		
Arbitrary NB	913.22	1064.33	151.11	< 0.01	27
	(289.30)	(302.57)	(166.39)		
Productivity	939.41	1044.53	105.12	< 0.01	59
	(252.99)	(268.54)	(181.47)		
Productivity B	1000.11	1092.95	92.83	< 0.01	31
	(263.54)	(232.64)	(163.02)		
Productivity NB	872.20	990.92	118.72	< 0.01	28
	(226.72)	(298.51)	(202.11)		
Needs	998.02	1018.13	20.11	0.12	56
	(300.56)	(276.84)	(221.15)		
Needs B	1017.47	1055.05	37.58	0.07	29
	(369.39)	(320.66)	(260.78)		
Needs NB	977.13	978.48	1.35	0.79	27
	(208.02)	(219.65)	(171.80)		

Notes: Standard deviations in parentheses. p-values from Wilcoxon two-sided sign-rank tests.

Figure 1: Hourly Data: Typed Characters



Note: Average number of typed characters are plotted for the five hours of the experiment for each treatment condition. The Control condition is plotted in every graph for reasons of comparison. In case hourly data could not be tracked by our assistants for one hour due to technical reasons the output count was assigned to the subsequent hour resulting in a productivity peak. To account for this, the above figure shows "smoothed" data in which the average productivity of both hours was assigned to each of the hours. The figure looks qualitatively identical plotting only observations where output was recorded in every single hour.

our treatment manipulation, as intended, was exogenous with respect to workers' inherent ability.[10] Looking at post-treatment productivity, the null hypothesis of equal productivity between bonus recipients and nonrecipients in the afternoon session cannot be rejected (Wilcoxon rank-sum test: p = 0.409, N = 236). We can neither reject the null that productivity was the same in the afternoon session across the eight treatment conditions (Kruskall-Wallis rank test: p = 0.582, N = 236), nor can we reject the null between the five conditions at the level of worker pairs (Kruskall-Wallis rank test: p = 0.713, N = 236). Thus, on average, the payment of bonuses apparently failed to induce higher productivity in the afternoon session.

Most of the previous experimental literature found a productivity burst soon after receiving the bonus, which may fade away with time (see section 1). We find the contrary in our experiment. While productivity increases sizably within both the two-hour morning session (pre-treatment) and the three-hour afternoon session (post-treatment), productivity in the third hour is always lower than in the second hour (see Figure 1) in every treatment condition as well as in the Control condition.[11]

4.2 Econometric analysis and parameter calibration

Because all workers have experienced the same working conditions during the morning session, differences in the changes of effort from the morning to the afternoon session across treatments can be causally attributed to the treatment variation. Therefore, we base our analysis of treatment effects on the following difference-in-difference regression model:

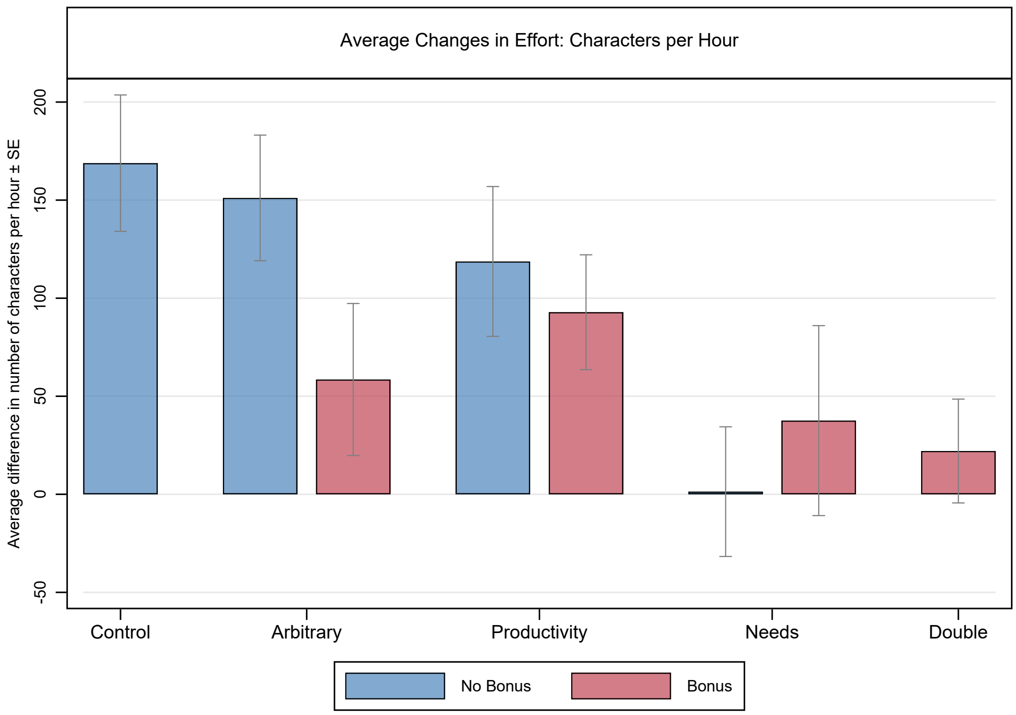
$$\begin{aligned} y_{it} = b_0 + \mu_i + b_1(Double \times a) + \sum_T b_{2_T} \cdot (Advantaged_T \times a) \\ + \sum_T b_{3_T}(Disadvantaged_T \times a) + \gamma \cdot a + \epsilon_{it} \end{aligned} \tag{5}$$

y_{it} denotes the outcome measure of worker i in session t. The constant b_0 captures the average outcome in the morning session ($t = 1$). Individual fixed effects μ_i control for individual heterogeneity in time-invariant characteristics, e.g. typing skills or the ability to concentrate which may affect productivity. The indicator variable a equals one if the observation is from the post-intervention ($t = 2$) period in the afternoon. We allow for the error term ϵ_{it} to be clustered at the worker pair level at which we administered the treatments (Abadie et al., 2017). The interactions $Double \times a$, $Advantaged_T \times a$, and $Disadvantaged_T \times a$ of treatment indicators (for each of the single-bonus treatments) with the post-intervention indicator deliver our estimated coefficients of interest, which capture the differences across treatments in the outcome change between the morn-

[10] In the Appendix Table A.2, we further show that treatment conditions are balanced with respect to several observable characteristics.

[11] Since a short lunch break took place between the second and third hour (see section 2), part of the productivity drop in the third hour may be due to a physiological drop in concentration due to food digestion, which we could call a "siesta effect".

Figure 2: Changes in Average Hourly Productivity across Treatment Conditions



Note: The figure displays the mean difference and its standard error of the average hourly number of typed characters between the morning and afternoon session in the respective treatment conditions.

Table 2: Treatment Effects

Dependent variable:	Characters (1)	Characters (2)	Quality (3)	Quality (4)
Double × Afternoon	-146.8****	-146.8****	0.06	0.06
	(43.00)	(42.81)	(0.53)	(0.53)
Arbitrary NB × Afternoon	-17.74		0.02	
	(49.47)		(0.50)	
Arbitrary B × Afternoon	-110.3**		0.18	
	(54.05)		(0.50)	
Productivity NB × Afternoon	-50.13		0.02	
	(53.64)		(0.87)	
Productivity B × Afternoon	-76.01		-0.81	
	(47.81)		(0.51)	
Needs NB × Afternoon	-167.5***		0.16	
	(50.14)		(0.51)	
Needs B × Afternoon	-131.3**		0.02	
	(61.29)		(0.56)	
Advantaged × Afternoon		-105.2**		0.07
		(43.95)		(0.44)
Disadvantaged × Afternoon		-78.11*		-0.21
		(43.19)		(0.38)
Afternoon	168.8****	168.8****	-0.70**	-0.70**
	(37.88)	(37.72)	(0.29)	(0.29)
Constant	971.5****	971.5****	98.01****	98.01****
	(6.06)	(6.25)	(0.08)	(0.08)
No. Individuals	236	236	236	236
No. Clusters	126	126	126	126
Hypothesis tests				
Double = Arbitary B	p = 0.405		p = 0.834	
Double = Productivity B	p = 0.049		p = 0.160	
Double = Needs B	p = 0.767		p = 0.958	
Double = Advantaged		p = 0.173		p = 0.596
Arbitrary B = Productivity B	p = 0.479		p = 0.094	
Arbitrary B = Needs B	p = 0.735		p = 0.799	
Needs B = Productivity B	p = 0.328		p = 0.194	
Arbitrary NB = Productivity NB	p = 0.514		p = 0.998	
Arbitary NB = Needs NB	p = 0.001		p = 0.819	
Needs NB = Productivity NB	p = 0.021		p = 0.885	

Notes: This table shows regression results from fixed effects model with the average of characters entered per hour in the morning and the afternoon session as dependent variable in the first two columns and the share of correct entries relative to all entries in the remaining two columns. Standard errors in parentheses clustered at session level. **** $p < 0.001$, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Hypothesis tests are Wald tests of the null hypotheses that the changes in the dependent variable are equal in the contrasted conditions.

Table 3: Calibrated Model Parameters

	Pooled	Arbitrary	Productivity	Needs
φ	-11.74	-11.74	-11.74	-11.74
	(3.43)****	(3.44)****	(3.44)****	(3.44)****
β_T	-3.33	-2.92	-5.66	-1.24
	(2.42)	(3.47)	(2.84)**	(4.18)
α_T	6.25	1.42	4.01	13.40
	(3.45)	(3.96)	(4.29)	(4.01)***

Hypothesis tests	β_T	α_T
Arbitrary = Productivity	p = 0.479	p = 0.514
Arbitrary = Needs	p = 0.735	p = 0.001
Needs = Productivity	p = 0.328	p = 0.021

Notes: The table contains calibrated structural parameters for each single-bonus treatment and pooled treatments. We assume m = 0, i.e. gift exchange or contentment effects are only dependent on own income. Parameter of the quadratic cost function fixed at r = 1. **** indicates p < 0.001, *** indicates p < 0.01, ** indicates p <0.05 and * indicates p < 0.1 of significant difference to zero.

ing and the afternoon session, relative to the Control condition. According to an F-test, individual fixed effects are highly significant ($p < 0.001$, F-test).

Table 2 presents the regression results. The first two columns display coefficients for the number of typed characters as the dependent variable whereas the last two columns use the accuracy rate (share of correct entries relative to all entries) as the dependent variable. While we look at treatment conditions separately in the first and the third columns, single-bonus treatment conditions are pooled in the second and fourth columns. Table 3 shows the calibrated parameters for the utility function developed in section 3.1. A detailed explanation of the calibration exercise can be found in section A.8 of the Appendix.

The cleanest way to test for the existence of behavior compatible with the gift exchange hypothesis (Hypothesis 1) is to compare the Control condition with the Double Bonus condition because this comparison is not confounded with the introduction of earnings inequality between the two workers. We find that workers in the Double Bonus condition significantly reduced effort in response to a bonus relative to the Control condition (-15 percent of the morning session average, -146.8 characters per hour, $p = 0.001$). The estimated parameter φ, capturing the willingness to reciprocate a wage increase in (1) and (2), equals -11.7 and is significantly lower than 0 ($p < 0.001$), pointing to a contentment effect rather than a gift exchange effect (see section 3.1). The prevalence of a contentment effect is also confirmed by noting that bonus recipients, on average, decreased effort from the morning to the afternoon session relative to the Control condition pooling the three single-bonus conditions (-10.8 percent[12], -105.2 characters per hour, $p = 0.018$). We conclude:

[12] We express the productivity change in percentage terms of the morning session average (971.5 characters per hour).

- *Result 1*: Effort in the Double Bonus condition decreased by roughly 15 percent in response to the wage bonus, reaching statistical significance at the 1 percent level. This result is consistent with a contentment effect (Hypothesis 1b) rather than a gift exchange effect (Hypothesis 1a).

To test Hypotheses 2 and Hypothesis 3 about the impact of earnings inequality on effort, we contrast productivity changes among bonus recipients (nonrecipients) in single-bonus treatments relative to the productivity change of workers in the Double Bonus (Control) condition. As the monetary payoff is the same in these treatment conditions, differences in productivity changes must be attributed to the payoff differential relative to the other worker. Consistently with Hypothesis 2, the calibration of the envy parameter α yields a positive sign pooling the single-bonus treatments, albeit not significantly different from zero. While nonrecipients in the Arbitrary and Productivity treatments do not reveal statistically significant differences in their output change relative to the Control condition (-1.8 percent and -5.2 percent, -17.7 and -50.1 characters per hour, $p = 0.72$ and $p = 0.35$, respectively), nonrecipients in the Needs treatment strongly reduced effort relative to the Control condition (-17.2 percent, -167.5 characters per hour, $p < 0.01$).[13] As for advantageous pay inequality, contrary to our primary Hypothesis 3a we do not find negative effects on productivity. Output increased among bonus recipients relative to the Double Bonus condition, the calibrated parameter β equaling -3.33 pooling the single-bonus treatments. However, this difference is not significantly different from zero (+4.3 percent, 41.6 characters per hour, $p = 0.173$). The sign is negative within each treatment, being significantly different from 0 only in the Productivity treatment (+7.3 percent, +70.8 characters per hour, $p = 0.049$, Wald test), but not in the Needs (+1.6 percent, +15.5 characters per hour, $p = 0.767$) nor the Arbitrary treatment (+3.8 percent, +36.5 characters per hour, $p = 0.405$). This result is consistent with Hypothesis 3b of status-seeking motivations, rather than compassion (see section 3.1). We conclude:

- *Result 2*: Disadvantageous pay inequality tended to yield a significantly negative effect on output, consistently with Hypothesis 2, although this was statistically significant (at the 1 percent level) only in the Needs treatment.

- *Result* 3: Advantageous pay inequality tended to yield a significantly positive effect on output, consistently with Hypothesis 3b of status-seeking behavior (rather than compassion), although this was statistically significant (at the 5 percent level) only in the Productivity treatment.

[13] As a further robustness check, we compared the output change among nonrecipients in the Productivity treatment to those workers in the Control condition who were less productive than their coworker in the morning, leaving conclusion unchanged (-62 characters per hour, $p = 0.36$, rank-sum test).

4.3 Impact of procedural fairness

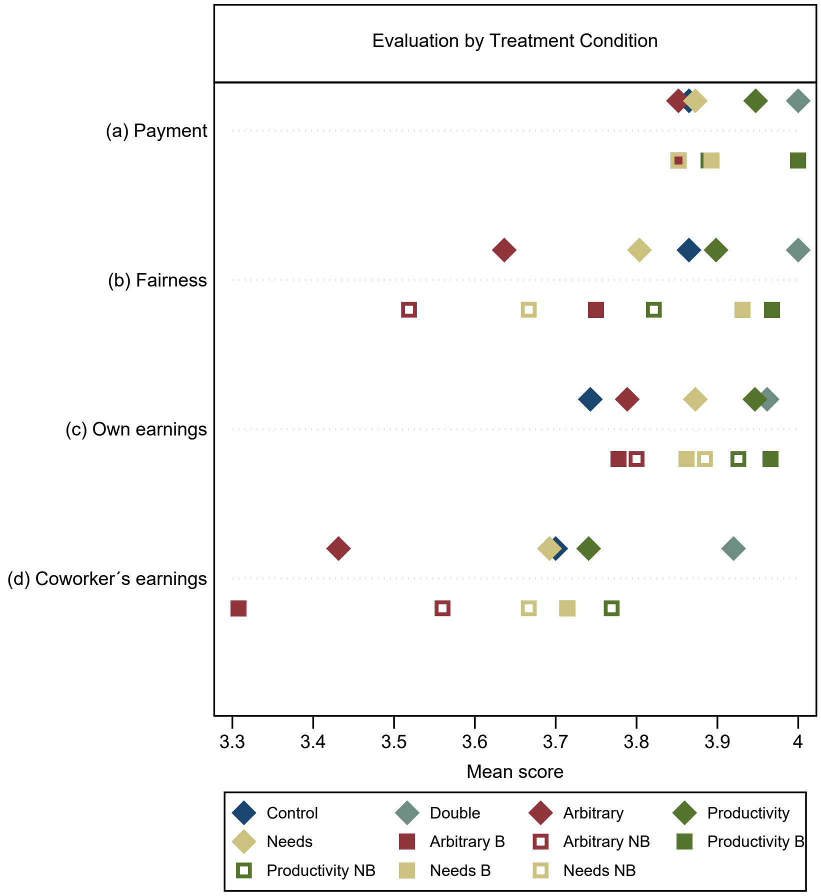
We first analyze the perception of pay fairness across treatments. We use questions from the post-experiment questionnaire asking workers to rate (a) the adequacy of payment, (b) the fairness of how workers had been treated, and whether (c) their own earnings, or (d) the coworker's earnings in the afternoon session were deserved. Answers could range from 1 (absolute disagreement) to 4 (absolute agreement). Figure 3 shows mean answer scores by treatment condition. Section A.4 in the Appendix reports an analysis of differences between treatments from an OLS regression.

Most workers perceived the payment as adequate, with little difference between treatments (see Appendix Table A.4).[14] Nonrecipients in the Arbitrary treatment stand out for reporting the lowest average score for the overall fairness in the way the workers were treated. Their evaluation score is weakly significantly lower than in the Control condition (p = 0.074, Wald test) and the Double Bonus condition (p = 0.006, Wald test), and is at the margins of statistical significance with respect to nonrecipients under the Productivity treatment (p = 0.12, Wald test; Table A.4, column 2). Considering recipients and nonrecipients together, reported fairness in the Arbitrary treatment is significantly lower than in the Productivity treatment (p = 0.025, Wald test) and the Double Bonus condition (p = 0.001, Wald-test; Table A.3, column 2). It is remarkable that all workers in the Double Bonus condition reported having been treated fairly.

Perception of deservedness of their own earnings in the afternoon session was highest for bonus recipients in the Productivity and the Double Bonus condition, reaching (marginal) statistical significance relative to the Control condition (p = 0.074 and p = 0.081, respectively Wald tests; Table A.4, column 4). The most sizable differences across treatments are observed when participants rated coworkers' earnings deservedness. Again, workers in the Arbitrary treatments expressed the lowest scores. Nonrecipients in the Arbitrary treatment report a lower score than workers in the Double bonus condition (p = 0.050, Wald test). Bonus recipients' score was even lower than nonrecipients' score within the Arbitrary treatment and was lower than in the Control condition (p = 0.110, Wald test), the Double Bonus condition (p = 0.003, Wald test), and compared to bonus recipients in the Productivity treatment (p = 0.071, Wald test) and the Needs treatment (p = 0.063, Wald test). Bonus recipients and nonrecipients in the Needs treatment rated the deservedness of their coworker's earnings lower than in the Double bonus condition (p = 0.046 and p = 0.048, respectively, Wald tests).

[14] Nonrecipients in the Arbitrary and the Needs treatment report the lowest scores, in case of the latter significantly lower than in the Double bonus condition (p = 0.036, Wald test).

Figure 3: Post-experimental Evaluation



Note: This figure shows the average answer score from the post-experimental survey by treatment condition. Arbitrary, Productivity, and Needs are averages over the respective conditions of recipients and nonrecipients. Answers could range between 1 "absolutely disagree" and 4 "absolutely agree" with respect to the following statements: (a) The payment I received was adequate. (b) The treatment of the two persons hired was fair. (c) The earnings I received in the second part of the day were deserved. (d) The earnings my coworker received in the second part of the day were deserved.

These mild differences in fairness perceptions in both Arbitrary treatment conditions failed however to lead to significant behavioral responses, i.e. productivity differences, contrary to Hypothesis 4b. We detect no difference between Productivity and Arbitrary treatments in the calibrated parameters relative to inequality aversion neither for bonus recipients ($p = 0.479$, Wald test) nor for nonrecipients ($p = 0.514$, Wald test) (see Table 3). Bonus recipients' output increased more in the Productivity treatment than the Needs treatment, although the difference is insignificant at conventional levels (+5.7 percent, +55.3 characters per hour, $p = 0.328$, Wald test). Contrary to our expectations, nonrecipients' output dropped in the Needs treatment significantly relative to both the Arbitrary (-15.4 percent, -149.8 characters per hour, $p = 0.001$, Wald test) and the Productivity (-12.1 percent, -117.4 characters per hour, $p = 0.021$, Wald test) treatment. We conclude:

- Result 4: Hypothesis 4a receives some support, albeit at levels that are often statistically insignificant. The Arbitrary treatment is generally perceived as less fair than the Productivity treatment, although differences reach statistical significance at 5% level only considering bonus recipients and nonrecipients jointly. Such differences in perception do not lead to any significant change in productivity in the Arbitrary treatment compared to the other single-bonus conditions, thus contradicting Hypothesis 4b. Surprisingly, we observe a strongly significant drop in productivity by nonrecipients in the Needs treatment, which is not reflected in fairness or deservedness ratings.

4.4 Robustness checks

In principle, workers' behavioral reactions to the treatments are not restricted to the quantity margin of their output, but possibly extend to their quality. In other words, workers may, consciously or unconsciously, adjust the precision of their work to either repay a gift from the employer or, adversely, punish them. In particular, workers dissatisfied with their treatment may try to "sabotage" the firm, keeping the number of characters relatively unchanged, but willingly introducing mistakes in their output. This would result in mistakes of various kinds, such as spelling mistakes, omissions, entry swaps, in the output returned by the worker. We investigated treatment effects in the share of correct entries (relative to all entries) in the afternoon session compared to the morning session in the third and fourth columns of Table 2. Our analysis fails to find any appreciable change in output quality between the afternoon and morning session, neither in the aggregate nor across treatments. The quality measured as the number of correct entries relative to all entries is roughly 98 percent[15] across all conditions, only slightly varying across them. Consequently, a non-parametric test of the null of equal changes in output quality between the morning and afternoon session across treatment groups cannot be rejected ($p = 0.570$, $N = 236$, Kruskal-Wallis test). Similarly, results are qualitatively identical using the number of correct entries as the outcome variable, which provides us with a composite measure of the output quantity and quality. Both findings indicate that the quality dimension of work is not significantly impacted by the treatments. Thus, we can also rule out that the contentment effect was driven by working with less speed but with higher precision.

Furthermore, the results documented in the previous sections are unaffected by several robustness checks. Pairwise treatment comparisons with Wilcoxon rank-sum tests, as shown in Table A.5 in the Appendix, yield identical conclusions to the regression analysis. We do not find any sign of a trend[16] in the output changes from the morning to the afternoon session over the 126

[15] Interestingly, Hennig-Schmidt, Sadrieh and Rockenbach (2010) report an almost identical share of correct entries in a similar data-entry task.

[16] The p-value of the time trend is $p = 0.6$ in an OLS regression with the output change as dependent variable.

days of data collection, which speaks against the possibility of contagion effects among workers. Omitting outliers by the method of Tukey's fences does not have a significant impact on our results. Results are also robust to excluding fifteen observations where a confederate acted as the second worker in case a recruited worker did not show up (see section 2 and Table A.6). We plot empirical cumulative distribution functions (ECDFs) in Figure A.1 in the Appendix, showing that the contentment effect is not only apparent in the differences in means, but also when we consider the full distributions, thus revealing a broad behavioral pattern. The ECDF corresponding to the productivity change in the Double Bonus condition is clearly shifted to the left compared to the ECDF of the Control condition. Workers who received a bonus in the single-bonus treatments also showed productivity changes more similar to those in the Double Bonus condition, consistent with the presence of a contentment effect among them. On the contrary, the productivity changes of nonrecipients in single-bonus treatments are more similar to those of the Control condition, except for the nonrecipients in the Needs treatment whose productivity relatively decreased.

5 Discussion

Experimental results are normally assessed in terms of internal and external validity. Esteves-Sorenson (2018) considers eight possible confounds affecting internal validity in gift exchange studies. First, peer effects may create idiosyncratic effects on outcomes. Such effects were minimized in our setting because workers were prevented from interacting and communicating with each other at any stage of the work session (see section 2). Second, concerns that within-subject pay manipulation may lead agents to inflate their effort after receiving the bonus because they want to avoid experiencing disutility from being perceived as selfish seem to be irrelevant in our naturalistic setting. We observe treated participants displaying, if anything, selfish behavior in comparison to non-treated participants, as the former reduce their productivity after receiving the bonus in relation to the Control condition. Moreover, participants ignore being part of an experiment and are therefore not motivated by the desire to comply with the researcher's perceived goals. Third, we also do not believe that insufficient wage raises played a part in our experiment. The bonus equaled 33% of the advertised wage rate, and this is at least on par with other experiments that found a gift exchange effect (Bellemare and Shearer, 2009; Cohn, Fehr and Goette, 2015; Gilchrist, Luca and Malhotra, 2016). Fourth, effort ceiling effects were likely absent in our experiment, because the workers demonstrated that higher productivity was possible. In fact, productivity peaked in the last hour of the experiment and the three-hour afternoon session gave ample time to workers to reciprocate the bonus. Fifth, fatigue effects most likely also played a minor role, as demonstrated by the fact that productivity was particularly low after the lunch break offering time to recover but, on average, increased in all treatment conditions as well as in the Control condition over time (see Figure 1). Sixth, we cannot rule out that the relatively high wage set in the Control condition, in relation to the market-clearing wage rate, led to the selection of workers abler than normal (Esteves-Sorenson, 2018). Nevertheless, the base wage was arguably appropriate, taking into account the commuting time and travel costs for Bogotá residents. Its value was decided after indications from the pilot experiment that a lower wage may have led to significant attrition. Moreover, the random assignment of workers into treatment minimizes the chance that results based on differences between treatments are due to imbalances in workers' abilities (see section 4.1 for the discussion of treatment exogeneity w.r.t. pre-intervention productivity). Seventh, as for reemployment concerns, we followed the best practice from other experiments and repeatedly ruled out, starting from the recruitment stage, that there would be any possibility for the worker to be re-employed in the future. We cannot of course exclude that participants did not fully believe, or take into account, this announcement. If this were the case, it would be however unclear why reemployment concerns should affect differently the Control and treatment conditions. Relatedly, nonrecipients in single-bonus treatments may have felt motivated to increase their effort levels, if they had thought that not receiving the bonus was a signal that their re-employment chances were lowered. Likewise, non-bonus recipients may have imagined that another bonus would have been paid at the end of the working day. If these motivations were at play, they should have presumably led nonrecipients in the Productivity treatment to raise their effort, because the bonus depended directly on workers' output. Productivity by nonrecipients was however not higher in the Productivity treatment than in the Arbitrary treatment in the afternoon session relative to the morning session (see Figure 2), suggesting that this factor did not play a significant role. More generally, productivity by nonrecipients in single-bonus treatments was in every case lower than in the Control condition in the afternoon session relative to the morning session. This suggests that effects due to reemployment concerns, or to the desire to receive a hypothetical second bonus, were largely irrelevant. The eighth and final point of concerning Esteves-Sorenson (2018) is that of small samples. Our study had an overall larger sample size than other experiments detecting gift exchange in natural experimental labor markets. Hence, we had adequate power to detect the existence of reciprocity effects. One may argue that we were still under-powered to capture differences between the various treatment effects, in particular between the Productivity and the Arbitrary treatments. The fact that we observe very small differences between Arbitrary and Productivity treatments (the observed effect sizes amount to a Coehn's d of roughly 0.18 comparing both treatments for bonus recipients and nonrecipients, see Table 1) make us doubtful that we are incurring into a Type-II error. Clearly, only a larger sample size could fully ascertain this aspect.[17] An additional concern is that the strength of a reciprocal response crucially depends on whether

[17]Laboratory studies comparing analogous treatments in Dictator Games find effect sizes in the range of d = 0.4, which would have required a sample size of roughly 100 observations in each contrasted condition to ensure an a-priori power of 80 percent given alpha = 0.05. Our initial assumption was that the effect size in our sample would have been considerably larger than d = 0.4, because we believed that the Arbitrary treatment would have caused a larger negative effect than treatments where assignment to roles is determined by an unbiased "fair" lottery (Bolton, Brandts and Ockenfels, 2005; Grimalda, Kar and Proto, 2016). We also believed that effects would have been larger in a natural setting than in a laboratory one. From an ex-post perspective, our assumptions did not turn out to be the case on either account.

the value of effort for the employer is emphasized to the workers, as discussed in parts of the gift exchange literature (DellaVigna et al., 2021; Englmaier and Leider, 2020). We cannot exclude the relevance of this concern for our study. But even if the value of effort for the employer during the relatively tedious data-entry task may not have been clear to the workers, this fact alone cannot explain why we observe a significant decrease of effort in response to the bonus although it might render the emergence of a contentment effect more likely.

As for external validity, Shadish (2002) and Cronbach (1980) argue that inference from local experimental evidence to general ones should be analyzed in terms of participants, treatments, settings, and observations (outcomes). Starting from participants, the first limitation of our study is participants' young age. Although a substantial portion of participants was not the usual convenience sample of university students, we are aware that significant differences across different age groups exist. Research comparing experimental student samples and nationally representative samples (Bellemare and Kröger, 2007; Cappelen et al., 2015) generally find that the former tend to behave more selfishly than other groups in the population. If that was the case, then we may indeed expect different patterns of reciprocity from workers of older age.

A second relevant aspect concerns the relevance of cultural effects in the determination of our results. Several studies have unveiled the relevance of culture for economic outcomes (Guiso, Sapienza and Zingales, 2006; Bandiera and Fischer, 2013). It is then unclear whether our findings would replicate to a different context, in particular to high-income Western societies where previous field tests of the gift exchange hypothesis were performed (Gneezy and List, 2006; Kube, Maréchal and Puppe, 2012; Cohn, Fehr and Goette, 2015). We believe that two points are in order on this aspect. First of all, Henrich, Heine and Norenzayan (2010) and Henrich (2020) have forcefully argued that a disproportionate amount of research has been carried out in Western societies and that people from these societies are rather peculiar in comparison to other societies. If economics, as a science, strives for developing universally valid notions, then it should confront itself with cultural diversity and see the extent to which principles of behavior are truly universal or culture-specific. Moreover, Western societies are probably the exception, rather than the norm, when it comes to cultural or psychological traits (Henrich, Heine and Norenzayan, 2010). Secondly, the growing body of literature on cross-cultural analysis enables us to at least gauge the impact of the cultural traits that are specific to Colombia. The mapping of social preferences carried out by Falk et al. (2018) shows that people from Colombia have an average score on positive reciprocity (the variable particularly relevant in a gift exchange situation) comparable to Northern Americans and Europeans. On the grounds of this study, then, one may expect that the results found in Colombia would be generalizable to Western countries, too. If one digs deeper into cultural traits, though, some differences emerge. According to the six-dimension model of national culture by Hofstede, Hofstede and Minkov (2010), Colombia is ranked as one of the most collectivistic countries in the world within the collectivist-individualism dimension (see also Hofstede Insights, 2020). It also tops the rankings for power distance, that is, the degree to which "the less powerful members of institutions and organizations within a country expect and accept that power is distributed unequally". On the other hand, North-Western countries typically rank high in individualism and low in power distance. One may conjecture that both collectivism – when workers perceive themselves as belonging to a different group than the entrepreneur – and power distance contribute to workers perceiving themselves as being socially distant from the entrepreneur. If that is the case, then workers may not perceive the payment of a bonus as a "kind" act that warrants gratitude. Or they may expect that the entrepreneur will not construe their extra effort as an act of kindness. As a result, they may not feel the need, or the moral obligation, to reciprocate a change in the wage rate, even if such a change goes to their advantage. In societies where power distance and collectivism are not dominant cultural traits, employees may perceive to engage with entrepreneurs on a more equal standing, and may then have more psychological incentives to perform positive reciprocity. In another natural experiment spanning different cultural areas, Bandiera, Dahlstrand-Rudin and Fischer (2020) find that employees from most collectivistic countries respond less to wage incentives in contracts offered by companies. Our results mirror such results from Bandiera, Dahlstrand-Rudin and Fischer (2020), although the incentivization scheme in Bandiera, Dahlstrand-Rudin and Fischer (2020) was based on standard economic incentives rather than on positive reciprocity. At a basic level, such results suggest that monetary incentives may be construed differently across different cultures.

As for the nature of our treatments, we believe that most of the elements of our design are standard components of real-life labor contracts, except for the Needs treatment. Bonus payments are widespread in labor markets around the world. Linking bonuses to productivity also seems common. Admittedly, it would be extravagant for a company to offer a selective bonus without giving any justification for it, as we do in the Arbitrary treatment. Nevertheless, it is also arguably the case that most workers, when being prone to a self-serving bias, may fail to recognize the legitimacy of the justification given by a manager for the bonus being assigned to another worker, and may thus consider the selective bonus as tantamount to being arbitrary (Bewley, 1999). Selective payments based on needs are rarely used as part of labor market contracts. As argued in section 2, the purpose of this treatmet was precisely to investigate reactions to an allocation principle having psychological salience and being policy relevant. A possible explanation for the strong productivity decrease among nonrecipients in the Needs treatment could precisely be that bonus assignment due to social status is not considered by workers to be a fair or appropriate reason at the workplace (Kahneman, Knetsch and Thaler, 1986).

As for settings, we believe that a strength of our study is that we designed a natural field experiment, where participants are not aware to be participating in an experiment and the work environment appears natural for the type of task (Harrison and List, 2004). In such a setting, people are generally less restricted in their behavior favoring the emergence of unforeseen results (List and Rasul, 2011). Regarding the outcomes of our job interaction, data entry output is common to many tasks and offers the obvious advantage of being a standardized measure that can be compared across different settings. In many jobs, outcomes are the result of a collective effort by several coworkers. Since such efforts cannot be singled out, adding this dimension into our experiment would have introduced undesirable confounds. Another limitation of our study is the focus on short-term employment (see also point seven in the discussion above), thus neglecting the long-term effects of the treatment. Nevertheless, casual employment is particularly relevant in countries with higher informality rates and is also increasing in developed countries (International Labour Organization, 2016).

6 Concluding remarks

The main result of this paper has been the finding of significant productivity decreases in response to a bonus payment in the context of a one-day labor contract involving two workers. Such a decrease among bonus recipients occurs in all treatments considered, but the effect is larger when both workers are bonus recipients. This evidence contradicts the gift exchange hypothesis and is consistent with what we termed a contentment effect. That is, the worker interprets the bonus payment as a signal that the entrepreneur is satisfied with the worker's performance, with no moral obligation to reciprocate for this action.

We also found that social comparison effects only played a minor role: workers who received a bonus as unique beneficiaries tended to increase productivity, albeit statistically insignificantly, compared to the Double Bonus condition, pointing to status-seeking preferences (Frank, 1985). Consistent with the observation that relative wages matter when evaluating whether the treatment was fair (Bewley, 1999), a bonus may create a perception of higher recognition when the wage of the non-rewarded serves as a reference point (Kahneman and Tversky, 1979). Hence, this reaction partially offsets the contentment effect. Nonrecipients significantly reduced productivity only in the Needs treatment. In general, workers' behavior responded only marginally to different justifications for pay inequality. More research is needed to ascertain the extent to which these results would hold across different age and cultural groups. Given that we only observed one data point for the base wage and the bonus payment, further research should also investigate the possible existence of a "ceiling effect" in reciprocity and a diminishing sensitivity to kindness.

Overall, our findings echo Bewley's (1999) insight that introducing pay inequality in the wage structure of a company may backfire. We have however uncovered a new channel whereby this effect manifests itself, which has not to do with work morale but rather to workers' lack of responsiveness to the "gift" offered by the firm.

A Appendix

A.1 Descriptives

Table A.1: Sample Characteristics

Variable	Obs	Mean	Std. Dev.	Min	Max
Demographics					
Age	235	21.16	1.92	18	27
Female	234	0.54	0.50	0	1
Married	236	0.08	0.27	0	1
Occupation					
Student	236	0.83	0.38	0	1
Unemployed	236	0.10	0.30	0	1
Other	236	0.07	0.25	0	1
Estratificación socioeconómica					
1	234	0.05	0.21	0	1
2	234	0.38	0.49	0	1
3	234	0.47	0.50	0	1
4	234	0.09	0.28	0	1
5	234	0.02	0.13	0	1
6	234	0.00	0.07	0	1
Education participants					
Bachelor	236	0.10	0.30	0	1
Some college semester	236	0.55	0.50	0	1
Technical univ.	236	0.15	0.36	0	1
University	236	0.19	0.39	0	1
Education attainment father					
None	236	0.01	0.11	0	1
Primary	236	0.17	0.37	0	1
Bachelor	236	0.35	0.48	0	1
Some college semester	236	0.08	0.27	0	1
Technical univ.	236	0.14	0.35	0	1
University	236	0.23	0.42	0	1
Education attainment mother					
None	236	0.00	0.07	0	1
Primary	236	0.17	0.38	0	1
Bachelor	236	0.33	0.47	0	1
Some college semester	236	0.10	0.30	0	1
Technical univ.	236	0.19	0.39	0	1
University	236	0.19	0.39	0	1

A.2 Treatments Balance

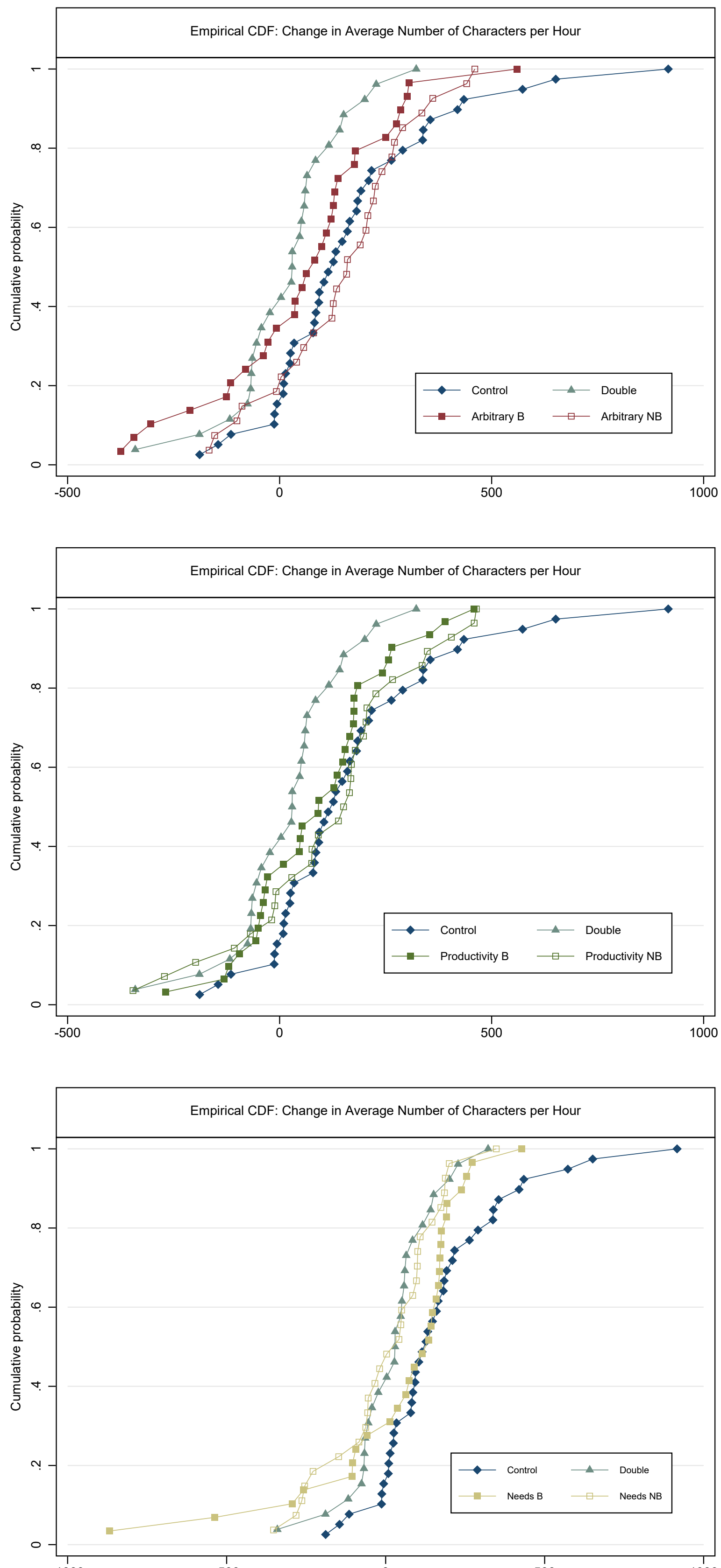
Table A.2: Balance Table

	Control	Arbitrary	Productivity	Needs	Double	Total	F-test
Age	21.711	20.446	21.288	21.321	21.269	21.162	0.008
	(2.205)	(1.548)	(1.894)	(1.879)	(2.070)	(1.921)	
Female	0.590	0.518	0.627	0.491	0.423	0.540	0.368
	(0.498)	(0.504)	(0.488)	(0.505)	(0.504)	(0.499)	
Married	0.128	0.054	0.085	0.071	0.077	0.081	0.818
	(0.339)	(0.227)	(0.281)	(0.260)	(0.272)	(0.273)	
Estrato	2.526	2.582	2.678	2.673	2.962	2.661	0.208
	(0.687)	(0.896)	(0.753)	(0.840)	(0.774)	(0.805)	
Degree	0.538	0.464	0.390	0.446	0.308	0.436	0.371
	(0.505)	(0.503)	(0.492)	(0.502)	(0.471)	(0.497)	

Notes: Table displays background characteristics for each treatment condition and the total sample. On average, workers are slightly younger in the Arbitrary treatment condition, significant according to an F-test, although the mean difference to the other conditions is 1.25 years at the maximum. "Degree" is a dummy for having an higher-education degree (Bachlor, Technical University or University degree). The p-value we report in the last column is from a F-test of joint significance in a regression of background characteristic on treatment indicators.

A.3 Empirical distribution functions of output changes

Figure A.1: Empirical distribution functions: Productivity changes



A.4 Post-experimental questionnaire: Satisfaction analysis

Table A.3: Regressions: Satisfaction Questions (Pooled Treatment Conditions)

	(1)	(2)	(3)	(4)
	Payment Adequate	Treatment Fair	Own profits deserved	Other's profits deserved
Double	0.135	0.135	0.219*	0.22
	-0.0846	-0.0848	-0.124	-0.15
Arbitrary	-0.013	-0.229*	0.0456	-0.269
	-0.115	-0.136	-0.141	-0.195
Productivity	0.0825	0.0334	0.204*	0.0407
	-0.0895	-0.0958	-0.121	-0.154
Needs	0.00786	-0.0613	0.13	-0.00769
	-0.098	-0.106	-0.128	-0.152
Constant	3.865****	3.865****	3.743****	3.700****
	-0.0846	-0.0848	-0.118	-0.14
Obs.	228	233	224	212
R2	0.0145	0.0447	0.0337	0.0527
Hypothesis tests				
Double = Control	0.113	0.114	0.079	0.144
Arbitrary = Control	0.910	0.095	0.748	0.171
Arbitrary = Double	0.062	0.001	0.048	0.001
Arbitrary = Productivity	0.256	0.025	0.061	0.041
Arbitrary = Needs	0.822	0.181	0.366	0.079
Productivity = Control	0.358	0.728	0.096	0.792
Productivity = Double	0.072	0.024	0.751	0.031
Productivity = Needs	0.195	0.228	0.208	0.575
Needs = Control	0.936	0.565	0.312	0.960
Needs = Double	0.011	0.003	0.158	0.004

Notes: Table shows OLS regression results from post-experimental questionnaire. p-values for pairwise tests between treatment conditions reported. Standard errors (clustered at session level) in parentheses. Column (1)-(4) have the answer score (1 absolutely disagree - 4 absolutely agree) to the following statements as dependent variable. (1) The payment I received was adequate. (2) The treatment of the two persons hired was fair. (3) The earnings I received in the second part of the day were deserved. (4) The earnings received by my coworker in the second part of the day were deserved.

Table A.4: Regressions: Satisfaction Questions

	(1)	(2)	(3)	(4)
Double	0.135	0.135	0.219*	0.22
	(0.09)	(0.09)	(0.12)	(0.15)
Arbitrary × NB	-0.013	-0.346*	0.0571	-0.14
	(0.14)	(0.19)	(0.16)	(0.22)
Arbitrary × B	-0.013	-0.115	0.0349	-0.392
	(0.14)	(0.16)	(0.17)	(0.24)
Productivity × NB	0.024	-0.0434	0.183	0.0692
	(0.11)	(0.12)	(0.13)	(0.17)
Productivity × B	0.135	0.103	0.223*	0.0143
	(0.09)	(0.09)	(0.12)	(0.17)
Needs × NB	-0.013	-0.198	0.142	-0.0333
	(0.11)	(0.16)	(0.15)	(0.18)
Needs × B	0.028	0.0662	0.119	0.0143
	(0.12)	(0.10)	(0.14)	(0.17)
Constant	3.865****	3.865****	3.743****	3.700****
	(0.09)	(0.09)	(0.12)	(0.14)
Obs.	228	233	224	212
R2	0.019	0.074	0.035	0.062
Double = Control	0.115	0.116	0.081	0.147
Arbitrary NB = Control	0.928	0.074	0.713	0.534
Productivity NB = Control	0.820	0.727	0.159	0.675
Needs NB = Control	0.906	0.207	0.333	0.855
Arbitrary B = Control	0.928	0.468	0.839	0.110
Productivity B = Control	0.115	0.262	0.074	0.935
Needs B = Control	0.810	0.500	0.380	0.932
Arbitrary NB = Double	0.203	0.006	0.133	0.050
Productivity NB = Double	0.074	0.050	0.577	0.132
Needs NB = Double	0.036	0.012	0.408	0.048
Double = Arbitary B	0.203	0.062	0.157	0.003
Double = Productivity B	1.000	0.321	0.938	0.074
Double = Needs B	0.176	0.153	0.189	0.046
Arbitrary B = Productivity B	0.203	0.113	0.146	0.071
Arbitrary B = Needs B	0.770	0.201	0.547	0.063
Needs B = Productivity B	0.176	0.527	0.164	1.000
Arbitrary NB = Productivity NB	0.778	0.121	0.265	0.282
Arbitary NB = Needs NB	1.000	0.494	0.520	0.610
Needs NB = Productivity NB	0.691	0.332	0.677	0.474

Notes: Table shows OLS regression results from post-experimental questionnaire. p-values for pairwise tests between treatment conditions reported. Standard errors (clustered at session level) in parentheses. Column (1)-(4) have the answer score (1 absolutely disagree - 4 absolutely agree) to the following statements as dependent variable. (1) The payment I received was adequate. (2) The treatment of the two persons hired was fair. (3) The earnings I received in the second part of the day were deserved. (4) The earnings received by my coworker in the second part of the day were deserved.

A.5 Translated treatment instructions

The original instructions were in Spanish language and are available on request.

- Control: "We are ready to continue the work. The work is the same as the first session, but the duration of the second session will be 3 hours. Today's salary will be 15,000 per hour as in the morning. At the end of the day we will pay you 75,000 pesos. I am going to accompany you to your office now. "
- Double: "We are ready to continue the work. The work is the same as the first session, but the duration of the second session will be 3 hours. Today's salary will be 15,000 per hour as in the morning. However we want to pay a premium of 25,000 pesos to both workers of the couple. At the end of the day we will pay 100,000 pesos. I am going to accompany you to your office now."
- Arbitrary: "We are ready to continue the work. The work is the same as the first session, but the duration of the second session will be 3 hours. Today's salary will be 15,000 per hour as in the morning. However, we want to pay a premium of 25,000 pesos to one of the couple's workers. Unfortunately we could not find out who was more qualified to receive it by looking at their resumes or their activity this morning, or who had fewer resources, so we already chose one among you. You will receive the premium and at the end of the day we will pay you 100,000 pesos. You will not receive the premium and at the end of the day you will receive 75,000 pesos. I am going to accompany you to your office now. "
- Productivity: "We are ready to continue the work. The work is the same as the first session, but the duration of the second session will be 3 hours. Today's salary will be 15,000 per hour as in the morning. However, we want to pay a premium of 25,000 pesos to one of the couple's workers. We have decided that it would be appropriate to assign the premium to the person who, according to the available information, would be more productive. We have found out your performance in this morning's session, and it turns out that you have completed more addresses than you. Then we have decided that it would be more appropriate for you to receive the premium, because based on these criteria it appears to be more productive. At the end of the day we will pay 100,000 pesos. You will not receive the premium and at the end of the day you will receive 75,000 pesos. I am going to accompany you to your office now. "
- Needs: "We are ready to continue the work. The work is the same as the first session, but the duration of the second session will be 3 hours. Today's salary will be 15,000 per hour as in the morning. However, we want to pay a premium of 25,000 pesos to one of the couple's workers. We have decided that it would be appropriate to assign the premium to the person who, based on the available information, had fewer resources. We have found out the stratum

of their homes, the level of poverty of the town where they live and the age. Then we have decided that it would be more appropriate for you to receive the premium, because based on these criteria it appears to have lower resources. At the end of the day we will pay 100,000 pesos. You will not receive the premium and at the end of the day you will receive 75,000 pesos. I am going to accompany you to your office now."

A.6 Rank-sum Tests: Differences-in-Differences across Treatment Conditions

Table A.5 shows treatment effects calculated as pairwise differences in differences between treatment conditions. We compare the change in the treatment condition from the first column to the change in the treatment condition of the first row (differences in changes from morning to afternoon productivity across treatment conditions).

Table A.5: Treatment Effects by Treatment Condition

	Control	Arbitrary B	Arbitrary NB	Productivity B	Productivity NB	Needs B	Needs NB
Arbitrary B	-110.31						
	(-65.33%)						
p-value	0.09						
Arbitrary NB	-17.74	92.58					
	(-10.50%)	(54.83%)					
p-value	0.72	0.07					
Productivity B	-76.01	34.30	-58.28				
	(-45.02%)	(20.31%)	(-34.51%)				
p-value	0.19	0.63	0.16				
Productivity NB	-50.13	60.19	-32.39	25.89			
	(-29.69%)	(35.64%)	(-19.18%)	(15.33%)			
p-value	0.71	0.21	0.54	0.39			
Needs B	-131.27	-20.95	-113.53	-55.25	-81.14		
	(-77.74%)	(-12.41%)	(-67.24%)	(-32.72%)	(-48.05%)		
p-value	0.24	0.69	0.10	0.84	0.31		
Needs NB	-167.49	-57.18	-149.76	-91.48	-117.37	-36.23	
	(-99.19%)	(-33.86%)	(-88.69%)	(-54.18%)	(-69.51%)	(-21.46%)	
p-value	0.00	0.27	0.00	0.08	0.02	0.15	
Double	-146.78	-36.46	-129.04	-70.76	96.65	-15.51	-20.72
	(-86.93%)	(-21.60%)	(-76.42%)	(-41.91%)	(57.24%)	(-9.19%)	(-12.27%)
p-value	0.00	0.29	0.00	0.09	0.02	0.12	0.89

Notes: ATE as difference in differences (afternoon-morning) between compared treatments in absolute and percentage terms. We compare the difference in the treatment condition from the first column relative to the difference in the treatment condition from the first row. p-values from non-parametric two-sided Wilcoxon rank-sum tests against the null of equal distributions in the pairwise compared treatment groups.

A.7 Calculation of effort predictions

We examine optimal effort choices in reaction to possible payoff combinations which can arise in our setup. In the first period, no bonus is paid by the firm such that $w_i = w_j = w$ holds, i.e. there is no wage inequality. The optimal level of effort is equal to:

$$\hat{e_1} = \frac{1}{2r}[\kappa + \varphi(l+m)(w - w_{norm})] \tag{6}$$

The effort level in the first period must be equal across treatments. The second period differs across treatment conditions depending on whom received the bonus. In the Control condition, no worker receives the bonus in the second period. Nevertheless, learning or fatigue effects are possible. We therefore set the effort level in the second period to be equal to the sum of the first period effort and a random variable $\tilde{L}$. We assume $\tilde{L}$ to be normally distributed across individuals with mean μ_L and variance σ_L^2.

$$e_{2\hat{Control}} = \hat{e_1} + \tilde{L}^i \tag{7}$$

In the Double Bonus condition, both workers receive the bonus, hence the second term capturing pay inequality of 2 equals zero.

- If both workers receive a bonus ($w_i = w_j = w + B$), the difference in the optimal effort between the second and first period must be equal to:

$$\Delta e_{DOUBL\hat{E}-BONUS} = \frac{v}{2r}\varphi(l+m)B + \tilde{L} \tag{8}$$

- The difference in optimal effort between the first and the second period for the worker receiving the bonus (whom we label as ADV)is:

$$\Delta \hat{e}_T^{ADV} = \frac{1}{2r}\left[\varphi\left\{l \cdot B\right\} + \left\{-\alpha_T \cdot B\right\}\right] + \tilde{L} \tag{9}$$

where $T = \{PRODUCTIVITY; NEEDS; ARBITRARY\}$ are the three single-bonus treatment conditions.

- The difference in optimal effort between the first and the second period for the worker not receiving the bonus (whom we label as DIS):

$$\Delta \hat{e}_T^{DIS} = \frac{1}{2r}\left[\varphi\left\{m \cdot B\right\} + \left\{-\beta_T \cdot B\right\}\right] + \tilde{L} \tag{10}$$

We can now examine how optimal effort responds to bonus payments in our model's framework, assuming $\tilde{L} > 0$. First, in case of a situation in which both workers receive a bonus $B > 0$, effort unambigously increases relative to the first period if $\varphi > 0$ since

$$\varphi_1 \cdot (l+m) \cdot B + \tilde{L} > 0 \tag{11}$$

In this situation of a "Double Bonus", the firm does not induce any wage inequality between the workers. The predicted behaviour by the model corresponds to the assumed reciprocating behaviour of paying back a kind action of the firm with kindness. If instead only one worker receives the bonus, the firm is no longer seen as unambigously kind in our proposed model. The effort of the benefitted worker increases relative to the first period in which payoffs were equal if

$$\left\{\varphi l + (-\alpha_T)\right\} B + \tilde{L} > 0 \tag{12}$$

holds. This corresponds to a situation where the appreciation of the gift received by worker i outweighs inequality aversion in the effect on the perceived kindness of the firm.

In case of $\alpha_T < 0$, which corresponds to a preference for advantageous inequality under the respective justification T, effort may increase even more than under the Double Bonus if this preference is stronger than the weighted delight for the other worker:

$$\varphi \{l \cdot B\} + (-\alpha_T \cdot B) > \varphi(l+m)B \Leftrightarrow (-\alpha_T) > \varphi_1 \cdot m \tag{13}$$

If instead the other worker receives the bonus, the disadvantaged worker's effort increases relative to the first period without payoff inequality if

$$\begin{aligned} &\varphi \{m \cdot B\} + (-\beta_T \cdot B) > 0 \\ &\Longleftrightarrow \varphi \cdot m - \beta_T > 0 \Leftrightarrow \varphi_1 \cdot m > \beta_T \end{aligned} \tag{14}$$

holds. In words, the disadvantaged worker must be delighted strongly enough for the other worker who receives the bonus such that the total effect on the perceived kindness of the firm is still positive despite the inequality in payoffs. Comparing both cases of inequality, the increase in effort from period 1 to period 2 is smaller in the case when the worker is disadvantaged than when the worker is benefitted by the bonus assignment if

$$\begin{aligned} &\varphi \cdot l \cdot B + (-\alpha_T \cdot B) > \varphi \cdot m \cdot B + (-\beta_T \cdot B) \\ &\qquad \Leftrightarrow \varphi \cdot l + (-\alpha_T) > \varphi_1 \cdot m + (-\beta_T) \Leftrightarrow \varphi(l-m) > \alpha_T - \beta_T \end{aligned} \tag{15}$$

, which holds for a parameterization in which $l > m$ (higher weight for own than coworker's earnings in gift exchange part) and $\beta > \alpha$ (envy is stronger than compassion) hold, assuming $\varphi > 0$. Predicted effort changes vary with the parameters l and m defining interdependent preferences. For reasons of clarity, we assume $\varphi > 0$ in the following discussion. First, we take a look at the case of a purely selfish worker i, i.e. for which $m = 0$, i.e. a purely selfish worker does not care about the other worker's payoffs in the gift exchange part. In this case, worker i's effort changes by $\frac{1}{2r}\varphi B$ from the first to the second period if worker i receives the bonus. In turn, effort of a purely selfish worker i decreases by $\frac{1}{2r}(-\beta_T \cdot B)$ if worker j receives the bonus since there is no compensating effect on the perceived kindness of the firm by feeling delighted for the other worker. Spiteful preferences which put a negative weight of the other worker's material payoffs, can be captured by the model assuming an $m < 0$. If the other worker "j" receives the bonus, the effort of the spiteful worker "i" changes by $\frac{1}{2r}(\varphi \cdot m \cdot B + (-\beta_T \cdot B))$. Hence, the negative effect of "envy" due to disadvantageous wage inequality on the perceived kindness of the firm gets further amplified by the spitefulness. The total effect on worker i's effort in case of spiteful ($m < 0$) preferences will be negative if $\varphi \cdot m < \beta_T$ which automatically holds for $\varphi > 0$.

A.8 Calibration on experimental data

The model shall serve as a structural micro-foundation for the reduced-form estimation with observed effort as the outcome variable. The identification strategy leads to under-identification of our key theoretical parameters. We would like to estimate the following set of parameters: $\{\varphi l; \varphi m; \alpha_T; \beta_T\}$, whereof $\beta_T; \alpha_T$ are vectors containing a unique parameter related to advantageous and disadvantageous inequality aversion, respectively, for each treatment. We only have seven treatment conditions in addition to the Control treatment: The Double Bonus condition and the three other treatment conditions where we observe effort levels under both, advantageous and disadvantageous wage inequality. The under-identification is due to the fact that, in our setup, we can only observe β_T together with l (weight of the own wage in gift exchange part), and α_T together with m (weight of the other worker's wage in the gift exchange part), but we cannot observe, say, β_T together with m, because we do not have the possibility to compare treatments with the same size of inequality but at the same time different size of the bonus. Due to those limitations, we assume $m = 0$. This means that the worker experiences a gift exchange or contenment effect only with respect to oneself, but not with respect to the other worker. This assumption seems relatively mild in the context of our theoretical framework.

We also assume that the parameter r from the quadratic effort cost function is a normally distributed random variable, which we rescale to have mean 1. In order to recover the parameters of our theoretical model from the observed behavior in the experiment, we estimate a fixed-effects regression model with the average hourly effort (either measured in characters, entries or correct entries) y_{it} in the two periods, given by the morning and the afternoon session, as dependent variable. To be precise, we estimate the following differences-in-differences regression:

$$y_{it} = b_0 + \mu_i + b_1(Double \times a) + \sum_T b_{2_T} \cdot (Advantaged_T \times a) + \sum_T b_{3_T}(Disadvantaged_T \times a) + \gamma \cdot a + \epsilon_{it} \tag{16}$$

to calibrate our model. This differences-in-differences approach eliminates the estimation of some unobservable parameters, such as w_{norm}. While in the simplified world of our theoretical model, the intercept would be zero, the constant term b_0 which corresponds to the change in effort in the Control group captures effects not accounted for in the theoretical model (e.g. learning and/or fatigue effects, model parameter $\tilde{L}$). Hence, the Control group serves as the reference category[18] and enables us to get rid of the $\tilde{L}$ function. The indicator functions for each single-

[18]Without adding controls this is equivalent to compare the conditional means of the difference in provided effort

bonus treatment are also included as vectors. The coefficient vectors b_{2_T} and b_{3_T} contain the regression coefficients from each of the single-bonus treatments in the afternoon. Due to the unobservability of effort costs, we assume $r = 1$ in our quadratic cost function such that $c(e) = e^2$. The remaining coefficients allow to identify the parameters φ, α_T and β_T. The parameters α_T and β_T defining inequality aversion can be inferred for each single-bonus treatment separately to identify differences in the source of horizontal inequality.

A.9 Relating regression coefficents to model parameters

Simplification: $m = 0; l = 1$ Our estimated regression coefficients can be related to the theoretical parameters from the model equations as follows. We take the difference in optimal effort as given by the model solution from the model subsection between the first and second period and impose the simplifying restrictions. From the observed effort change in the Double Bonus condition where no payoff inequality arises we arrive at an expression to determine the product φ. This defines the reaction of worker effort to a symmetric pay rise without wage inequality.

$$\Delta e_{Double} = b_1 = \frac{1}{2b}\varphi(w + B - w_{norm}) - \frac{1}{2b}\varphi(w - w_{norm}) = \frac{1}{2b}\varphi \cdot B \iff \varphi = b_1 \frac{2b}{B}. \quad (17)$$

Applying the same procedure for the difference in effort when either receiving or not receiving the bonus, we can express the remaining theoretical model parameters in terms of the fitted coefficients b_1,b_{2_T} and b_{3_T} (matching the $b's$ from the regression model) after rearranging equations.

$$\Delta e_{Advantaged_T} = b_2 = \frac{1}{2b}(\varphi + (-\beta)) \cdot B = b_1 - \beta_T \cdot \frac{B}{2b} \iff \beta_T = \frac{2b \cdot (b_1 - b_2)}{B} \quad (18)$$

$$\Delta e_{Disadvantaged_T} = b_3 = \frac{1}{2b}(\varphi + (-\alpha_T)) \cdot B \iff \alpha_T = -b_3 \cdot \frac{2b}{B} \quad (19)$$

levels between the treatment conditions.

A.10 Additional Regressions

Table A.6: Panel Data Fixed-Effects Model Regressions

	Whole Sample		Outliers removed		
	(1)	(2)	(3)	(4)	(5)
	Entries	Correct entries	Characters	Entries	Correct entries
Double × Afternoon	-5.885**	-5.645**	-127.1****	-4.676**	-4.434**
	(2.56)	(2.55)	(37.05)	(2.26)	(2.23)
Arbitrary NB × Afternoon	-0.324	-0.151	1.94	0.885	1.06
	(2.64)	(2.56)	(44.33)	(2.34)	(2.24)
Arbitrary B × Afternoon	-5.807*	-5.506*	-90.64*	-4.598*	-4.295
	(2.98)	(2.94)	(49.34)	(2.72)	(2.67)
Productivity NB × Afternoon	-0.797	-0.918	-30.45	0.411	0.293
	(3.12)	(3.11)	(48.89)	(2.88)	(2.85)
Productivity B × Afternoon	-3.658	-3.993	-56.34	-2.449	-2.781
	(2.53)	(2.53)	(42.48)	(2.22)	(2.21)
Needs NB × Afternoon	-9.151***	-8.713***	-147.8***	-7.942***	-7.501***
	(2.95)	(2.86)	(45.07)	(2.68)	(2.58)
Needs B × Afternoon	-4.031	-3.805	-79.25	-2.868	-2.665
	(2.87)	(2.81)	(48.26)	(2.66)	(2.57)
Afternoon	9.410****	8.799****	149.2****	8.202****	7.588****
	(2.04)	(2.00)	(31.03)	(1.63)	(1.57)
Constant	55.18****	54.14****	968.7****	55.28****	54.23****
	(0.42)	(0.42)	(6.84)	(0.41)	(0.40)
No. Individuals	236	236	234	234	234
No. Clusters	126	126	126	126	126
Wald-tests					

Table A.6: Panel Data Fixed-Effects Model Regressions

	Whole Sample		Outliers removed		
	(1)	(2)	(3)	(4)	(5)
Double = Arbitary B	0.977	0.959	0.402	0.977	0.959
Double = Productivity B	0.306	0.46	0.047	0.306	0.46
Double = Needs B	0.469	0.469	0.258	0.490	0.496
Arbitrary B = Productivity B	0.419	0.571	0.477	0.419	0.571
Arbitrary B = Needs B	0.551	0.562	0.831	0.568	0.585
Needs B = Productivity B	0.883	0.941	0.627	0.871	0.964
Arbitrary NB = Productivity NB	0.871	0.789	0.512	0.871	0.789
Arbitary NB = Needs NB	0.001	0.001	0.001	0.001	0.001
Needs NB = Productivity NB	0.009	0.014	0.020	0.009	0.014

Notes: Table shows Fixed-effects panel data regression results. The first three columns show the results for entries, and correct entries as dependent variable using the whole sample as in the main text. The remaining columns show the results for the same dependent variables plus characters when excluding observations whose change in effort lies outside of Tukey's outer fences (below or above the first or third quartile by three times the interquartile range, respectively), thus excluding strong outliers also visible in the ECDFs above. Clustered standard errors at the session (couple) level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$.

References

Abadie, Alberto, Susan Athey, Guido Imbens, and Jeffrey Wooldridge. 2017. "When Should You Adjust Standard Errors for Clustering?"

Abbink, Klaus, and Abdolkarim Sadrieh. 2009. "The pleasure of being nasty." *Economics Letters*, 105(3): 306–308.

Abbink, Klaus, and Benedikt Herrmann. 2011. "The Moral Costs of Nastiness." *Economic Inquiry*, 49(2): 631–633.

Adams, J. Stacy. 1963. "Toward an understanding of inequity." *Journal of abnormal psychology*, 67: 422–436.

Akerlof, George A. 1982. "Labor Contracts as Partial Gift Exchange." *The Quarterly Journal of Economics*, 97(4): 543–569.

Akerlof, George A., and Janet L. Yellen. 1990. "The Fair Wage-Effort Hypothesis and Unemployment." *The Quarterly Journal of Economics*, 105(2): 255–283.

Alesina, Alberto, and Paola Giuliano. 2011. "Preferences for Redistribution." In *Handbook of social economics*. Vol. 1 of *Handbooks in economics*, , ed. Alberto Bisin, Jess Benhabib and Matthew O. Jackson, 93–131. Amsterdam and London and Boston:Elsevier/North-Holland.

Almås, Ingvild, Alexander W. Cappelen, and Bertil Tungodden. 2020. "Cutthroat Capitalism versus Cuddly Socialism: Are Americans More Meritocratic and Efficiency-Seeking than Scandinavians?" *Journal of Political Economy*, 128(5): 1753–1788.

Al-Ubaydli, Omar, and John A. List. 2017. "Field Experiments in Markets." In *Handbook of economic field experiments*. Vol. 1 of *Handbooks in economics*, , ed. Abhijit V. Banerjee and Esther Duflo, 271–307. Amsterdam and Oxford:North-Holland an imprint of Elsevier.

Al-Ubaydli, Omar, Steffen Andersen, Uri Gneezy, and John A. List. 2015. "Carrots that look like sticks: Toward an understanding of multitasking incentive schemes." *Southern Economic Journal*, 140212121118009.

Ariely, Dan, Emir Kamenica, and Dražen Prelec. 2008. "Man's search for meaning: The case of Legos." *Journal of Economic Behavior & Organization*, 67(3-4): 671–677.

Azariadis, Costas. 1975. "Implicit Contracts and Underemployment Equilibria." *Journal of Political Economy*, 83(6): 1183–1202.

Babcock, Linda, and George Loewenstein. 1997. "Explaining Bargaining Impasse: The Role of Self-Serving Biases." *Journal of Economic Perspectives*, 11(1): 109–126.

Bandiera, Oriana, Amanda Dahlstrand-Rudin, and Greg Fischer. 2020. "Incentives and Culture: Evidence from a Multi-Country Field Experiment." *Working Paper.*

Bandiera, Oriana, and Greg Fischer. 2013. "Can "good" HR practices be exported? Evidence from a field experiment in Ghana." *Working Paper.*

Becker, Gary S. 1974. "A Theory of Social Interactions." *Journal of Political Economy*, 82(6): 1063–1093.

Bellemare, Charles, and Bruce Shearer. 2009. "Gift giving and worker productivity: Evidence from a firm-level experiment." *Games and Economic Behavior*, 67(1): 233–244.

Bellemare, Charles, and Sabine Kröger. 2007. "On representative social capital." *European Economic Review*, 51(1): 183–202.

Bénabou, Roland, and Jean Tirole. 2006. "Incentives and Prosocial Behavior." *American Economic Review*, 96(5): 1652–1678.

Bewley, Truman Fassett. 1999. *Why wages don't fall during a recession.* Cambridge, Mass:Harvard University Press.

Bogliacino, Francesco, Laura Jiménez Lozano, and Daniel Reyes. 2018. "Socioeconomic stratification and stereotyping: Lab-in-the-field evidence from Colombia." *International Review of Economics*, 65(1): 77–118.

Bolton, Gary E., and Axel Ockenfels. 2000. "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review*, 90(1): 166–193.

Bolton, Gary E., Jordi Brandts, and Axel Ockenfels. 2005. "Fair Procedures: Evidence from Games Involving Lotteries." *The Economic Journal*, 115(506): 1054–1076.

Bracha, Anat, Uri Gneezy, and George Loewenstein. 2015. "Relative Pay and Labor Supply." *Journal of Labor Economics*, 33(2): 297–315.

Breza, Emily, Supreet Kaur, and Yogita Shamdasani. 2018. "The Morale Effects of Pay Inequality*." *The Quarterly Journal of Economics*, 133(2): 611–663.

Cappelen, Alexander W., Astri Drange Hole, Erik Ø. Sørensen, and Bertil Tungodden. 2007. "The Pluralism of Fairness Ideals: An Experimental Approach." *American Economic Review*, 97(3): 818–827.

Cappelen, Alexander W., Knut Nygaard, Erik Ø. Sørensen, and Bertil Tungodden. 2015. "Social Preferences in the Lab: A Comparison of Students and a Representative Population." *The Scandinavian Journal of Economics*, 117(4): 1306–1326.

Charness, G., and M. Rabin. 2002. "Understanding Social Preferences with Simple Tests." *The Quarterly Journal of Economics*, 117(3): 817–869.

Charness, Gary. 2004. "Attribution and Reciprocity in an Experimental Labor Market." *Journal of Labor Economics*, 22(3): 665–688.

Charness, Gary, Guillaume R. Frechette, and John H. Kagel. 2004. "How Robust is Laboratory Gift Exchange?" *Experimental Economics*, 7(2): 189–205.

Charness, Gary, Luca Rigotti, and Aldo Rustichini. 2007. "Individual Behavior and Group Membership." *American Economic Review*, 97(4): 1340–1352.

Cohn, Alain, Ernst Fehr, and Lorenz Goette. 2015. "Fair Wages and Effort Provision: Combining Evidence from a Choice Experiment and a Field Experiment." *Management Science*, 61(8): 1777–1794.

Cohn, Alain, Ernst Fehr, Benedikt Herrmann, and Frédéric Schneider. 2014. "Social comparison and effort provision: Evidence from a field experiment." *Journal of the European Economic Association*, 12(4): 877–898.

Cox, James C., Daniel Friedman, and Steven Gjerstad. 2007. "A tractable model of reciprocity and fairness." *Games and Economic Behavior*, 59(1): 17–45.

Cronbach, Lee Joseph. 1980. *Toward reform of program evaluation.* San Francisco/Calif.:Jossey-Bass.

Deffains, Bruno, Romain Espinosa, and Christian Thöni. 2016. "Political self-serving bias and redistribution." *Journal of Public Economics*, 134(3): 67–74.

DellaVigna, Stefano, John List, Ulrike Malmendier, and Gautam Rao. 2021. "Estimating Social Preferences and Gift Exchange at Work." *Working Paper.*

Dufwenberg, Martin, and Georg Kirchsteiger. 2004. "A theory of sequential reciprocity." *Games and Economic Behavior*, 47(2): 268–298.

Durante, Ruben, Louis Putterman, and Joël van der Weele. 2014. "Preferences for Redistribution and Perception of Fairness: An Experimental Study." *Journal of the European Economic Association*, 12(4): 1059–1086.

Englmaier, Florian, and Stephen Leider. 2020. "Managerial Payoff and Gift-Exchange in the Field." *Review of Industrial Organization*, 56(2): 259–280.

Esteves-Sorenson, Constança. 2018. "Gift Exchange in the Workplace: Addressing the Conflicting Evidence with a Careful Test." *Management Science*, 64(9): 4365–4388.

Falk, Armin, and Andrea Ichino. 2006. "Clean Evidence on Peer Effects." *Journal of Labor Economics*, 24(1): 39–57.

Falk, Armin, and Urs Fischbacher. 2006. "A theory of reciprocity." *Games and Economic Behavior*, 54(2): 293–315.

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde. 2018. "Global Evidence on Economic Preferences*." *The Quarterly Journal of Economics*, 133(4): 1645–1692.

Fehr, E., and K. M. Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *The Quarterly Journal of Economics*, 114(3): 817–868.

Fehr, E., G. Kirchsteiger, and A. Riedl. 1993. "Does Fairness Prevent Market Clearing? An Experimental Investigation." *The Quarterly Journal of Economics*, 108(2): 437–459.

Fehr, Ernst, and Armin Falk. 1999. "Wage Rigidity in a Competitive Incomplete Contract Market." *Journal of Political Economy*, 107(1): 106–134.

Fehr, Ernst, and Simon Gächter. 1998. "Reciprocity and economics: The economic implications of Homo Reciprocans." *European Economic Review*, 42(3-5): 845–859.

Fehr, Ernst, and Simon Gächter. 2000. "Fairness and Retaliation: The Economics of Reciprocity." *Journal of Economic Perspectives*, 14(3): 159–182.

Fehr, Ernst, and Urs Fischbacher. 2004. "Third-party punishment and social norms." *Evolution and Human Behavior*, 25(2): 63–87.

Fehr, Ernst, Daniela Glätzle-Rützler, and Matthias Sutter. 2013. "The development of egalitarianism, altruism, spite and parochialism in childhood and adolescence." *European Economic Review*, 64(1): 369–383.

Fehr, Ernst, Erich Kirchler, Andreas Weichbold, and Simon Gächter. 1998. "When Social Norms Overpower Competition: Gift Exchange in Experimental Labor Markets." *Journal of Labor Economics*, 16(2): 324–351.

Fehr, Ernst, Lorenz Goette, and Christian Zehnder. 2009. "A Behavioral Account of the Labor Market: The Role of Fairness Concerns." *Annual Review of Economics*, 1(1): 355–384.

Frank, Robert H. 1985. *Choosing the right pond: Human behavior and the quest for status.* New York, NY:Oxford Univ. Press.

Gächter, Simon, and Christian Thöni. 2010. "Social comparison and performance: Experimental evidence on the fair wage–effort hypothesis." *Journal of Economic Behavior & Organization*, 76(3): 531–543.

Gilchrist, Duncan S., Michael Luca, and Deepak Malhotra. 2016. "When 3 + 1 > 4: Gift Structure and Reciprocity in the Field." *Management Science*, 62(9): 2639–2650.

Gneezy, Uri, and John A. List. 2006. "Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments." *Econometrica*, 74(5): 1365–1384.

Grimalda, Gianluca, Anirban Kar, and Eugenio Proto. 2016. "Procedural fairness in lotteries assigning initial roles in a dynamic setting." *Experimental Economics*, 19(4): 819–841.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2006. "Does Culture Affect Economic Outcomes?" *Journal of Economic Perspectives*, 20(2): 23–48.

Hagenlocher, Michael, Eric Delmelle, Irene Casas, and Stefan Kienberger. 2013. "Assessing socioeconomic vulnerability to dengue fever in Cali, Colombia: Statistical vs expert-based modeling." *International journal of health geographics*, 12: 36.

Harrison, Glenn W., and John A. List. 2004. "Field Experiments." *Journal of Economic Literature*, 42(4): 1009–1055.

Hennig-Schmidt, Heike, Abdolkarim Sadrieh, and Bettina Rockenbach. 2010. "In Search of Workers' Real Effort Reciprocity-a Field and a Laboratory Experiment." *Journal of the European Economic Association*, 8(4): 817–837.

Henrich, Joseph. 2020. *The WEIRDest People in the World: How the West Became Psychologically Peculiar and Particularly Prosperous.* [S.I.]:Farrar, Straus and Giroux.

Henrich, Joseph, Steven J. Heine, and Ara Norenzayan. 2010. "The weirdest people in the world?" *The Behavioral and brain sciences*, 33(2-3): 61–83; discussion 83–135.

Hofstede, Geert, Gert Jan Hofstede, and Michael Minkov. 2010. *Cultures and organizations: Software of the mind ; intercultural cooperation and its importance for survival.* . Rev. and expanded 3. ed. ed., New York:McGraw-Hill.

Hofstede Insights. 2020. "Country comparisons."

International Labour Organization. 2016. *Non-standard employment around the world: Understanding challenges, shaping prospects.* Geneva:International Labour Office.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica*, 47(2): 263.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler. 1986. "Fairness as a Constraint on Profit Seeking: Entitlements in the Market." *The American Economic Review*, 76(4): 728–741.

Katz, Lawrence F. 1986. "Efficiency Wage Theories: A Partial Evaluation." *NBER Macroeconomics Annual*, 1: 235.

Konow, James. 2000. "Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions." *American Economic Review*, 90(4): 1072–1092.

Konow, James. 2003. "Which Is the Fairest One of All? A Positive Analysis of Justice Theories." *Journal of Economic Literature*, 41(4): 1188–1239.

Krawczyk, Michał. 2010. "A glimpse through the veil of ignorance: Equality of opportunity and support for redistribution." *Journal of Public Economics*, 94(1-2): 131–141.

Krawczyk, Michal Wiktor. 2011. "A model of procedural and distributive fairness." *Theory and Decision*, 70(1): 111–128.

Kube, Sebastian, Michel André Maréchal, and Clemens Puppe. 2012. "The Currency of Reciprocity: Gift Exchange in the Workplace." *American Economic Review*, 102(4): 1644–1662.

Kube, Sebastian, Michel André Maréchal, and Clemens Puppe. 2013. "Do wage cuts damage work morale? Evidence from a natural field experiment." *Journal of the European Economic Association*, 11(4): 853–870.

Lazear, Edward P. 2000. "Performance Pay and Productivity." *American Economic Review*, 90(5): 1346–1361.

Lindbeck, A., and Dennis Snower. 1988. "Cooperation, Harassment, and Involuntary Unemployment: An Insider-Outsider Approach." *American Economic Review*, 78(1): 167–188.

List, John A., and Imran Rasul. 2011. "Field Experiments in Labor Economics." In *Handbook of labor economics*. Vol. 4 of *Handbooks in economics*, , ed. David E. Card and Orley Ashenfelter, 103–228. Amsterdam and New York and New York, N.Y., U.S.A:North-Holland.

Loewenstein, George F., Leigh Thompson, and Max H. Bazerman. 1989. "Social utility and decision making in interpersonal contexts." *Journal of Personality and Social Psychology*, 57(3): 426–441.

Martinsson, Peter, Clara Villegas-Palacio, and Conny Wollbrant. 2015. "Cooperation and social classes: Evidence from Colombia." *Social Choice and Welfare*, 45(4): 829–848.

Mezulis, Amy H., Lyn Y. Abramson, Janet S. Hyde, and Benjamin L. Hankin. 2004. "Is there a universal positivity bias in attributions? A meta-analytic review of individual, developmental, and cultural differences in the self-serving attributional bias." *Psychological bulletin*, 130(5): 711–747.

Miller, Dale T., and Michael Ross. 1975. "Self-serving biases in the attribution of causality: Fact or fiction?" *Psychological bulletin*, 82(2): 213–225.

Nicklisch, Andreas, and Fabian Paetzel. 2020. "Need-Based Justice and Distribution Procedures: The Perspective of Economics." In *Need-Based Distributive Justice.* , ed. Stefan Traub and Bernhard Kittel, 161–190. Cham:Springer International Publishing.

Rabin, Matthew. 1993. "Incorporating Fairness into Game Theory and Economics." *American Economic Review*, 83(5): 1281–1302.

Schurter, Karl, and Bart J. Wilson. 2009. "Justice and Fairness in the Dictator Game." *Southern Economic Journal*, 76(1): 130–145.

Shadish, William R. 2002. "Revisiting field experiments: Field notes for the future." *Psychological Methods*, 7(1): 3–18.

Trautmann, Stefan T. 2009. "A tractable model of process fairness under risk." *Journal of Economic Psychology*, 30(5): 803–813.