Journal of Behavioral and Experimental Economics 102 (2023) 101966

Contents lists available at ScienceDirect

Journal of Behavioral and Experimental Economics

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

Disentangling the influences of positive reciprocity and mood on gift exchange at work☆

Santiago I. Sautua

Universidad del Rosario, Department of Economics, Bogotá, Colombia

ARTICLE INFO

Dataset link: https://osf.io/y3f2z

JEL classification:
Codes
C93
D91
J33

Keywords:
Field experiment
Gift exchange
Reciprocity
Affect

ABSTRACT

While previous studies documented the existence of gift exchange at work, none of them identified the distinct influences of positive reciprocity and affective state. In this study, a field experiment was conducted in an effort to disentangle these two mechanisms. A subset of workers were promised an unexpected pay rise. Some of them started working immediately after the announcement, whereas others started working a week later. Because the pay rise announcement produced an immediate boost in elicited mood that disappeared after a week, this study design enabled separation of the reciprocity and mood motives. The results showed that the announcement increased productivity both among the workers who commenced work immediately after the announcement and among those who commenced work a week later. Moreover, while the increase in productivity was mainly driven by positive reciprocity, the boost in mood played a significant role immediately after the announcement.

1. Introduction

A key question for organizations is how to motivate workers to become more productive. Akerlof (1982) argued that employees who are paid above-market wages will work harder to return their employer's gift, thereby engaging in gift exchange. Over the last three decades, gift exchange in labor markets has been extensively studied, both in the laboratory and in the field (see Esteves-Sorenson, 2018 and DellaVigna et al., 2022 for detailed reviews). In a typical field experiment, workers are hired at a fixed wage for a one-off job, and a random sub-group receives an unexpected pay rise immediately before the work commences. Higher productivity among those who have received the pay rise is regarded as evidence of gift exchange.

Gift exchange, both in the labor market and elsewhere, is usually attributed to reciprocity. Several studies in anthropology, sociology, social psychology, and economics have argued that reciprocity is a universal social norm resulting from the moral duty to repay those who have helped us (Blau, 1964; Cialdini, 1993; Gouldner, 1960; Malmendier & Schmidt, 2017; Mauss, 1924; Pruitt, 1968; Wilke & Lanzetta, 1970).[1] Sometimes, such moral indebtedness is backed by a wish to reward the gift giver's good intentions (Dufwenberg & Kirchsteiger, 2004; Rabin, 1993).[2]

Productivity increases following a performance-independent pay rise, however, need not be entirely driven by positive reciprocity. Such increases could also be the result of an improved affective state because an unexpected pay rise is likely to boost a worker's mood (DellaVigna et al., 2022), and people who feel good are more likely to help others (Carlson et al., 1988; Isen & Levin, 1972; Levin & Isen, 1975; Weyant, 1978). Identifying to what extent gift exchange following a pay rise results from positive reciprocity versus positive mood has

☆ I thank the associate editor, two anonymous referees, and Sean Fahle for valuable comments and suggestions, and Diego Aycinena, Francesco Bogliacino, and Darwin Cortés for sharing the work tasks that I used in the experiment. I also thank the audiences at the Bogotá Experimental and Behavioral Economics Seminar and the Centre for Experimental Research on Fairness, Inequality and Rationality at the Norwegian School of Economics for their feedback. Steffanny Romero provided excellent research assistance, and María Camila Kairuz and Andrés F. Zambrano provided crucial support with the implementation of the experiment. I gratefully acknowledge financial support from Universidad del Rosario, Colombia. I thank Geoff Whyte, MBA, from Edanz Group (https://en-author-services.edanz.com/ac) for editing a draft of this manuscript. This study was approved by the Social Science Ethics Committee of Universidad del Rosario on 23 August 2019 (study 236-CS156).

E-mail address: santiago.sautua@urosario.edu.co.

[1] Some researchers have stressed that cognitive awareness of this social norm is often accompanied by a feeling of gratitude (Bartlett & DeSteno, 2006). Adam Smith labeled gratitude as the moral sentiment that "most immediately and directly prompts us to reward" (Smith, 1759–1809).

[2] While several studies, including this one, focus on *positive* reciprocity, other studies have emphasized the role of *negative* reciprocity, which leads individuals to punish others who have treated them unkindly (Akerlof & Yellen, 1990; Cohn et al., 2014; Kube et al., 2013; Lee & Rupp, 2007).

https://doi.org/10.1016/j.socec.2022.101966
Received 14 January 2022; Received in revised form 17 September 2022; Accepted 28 November 2022
Available online 30 November 2022

significant implications for organizations. On the basis of previous psychological research, an improved affective state is expected to produce only ephemeral increases in effort because the boost in mood fades quite rapidly (Isen et al., 1976), whereas reciprocity could have persistent positive effects on productivity. If pay rises induce mainly a temporary boost in mood, organizations are likely to reap only short-term benefits from wage increases. If, on the contrary, pay rises elicit substantial reciprocity, organizations may be able to lift worker productivity long after the rise is announced. Previous field experiments on gift exchange at work, pioneered by Gneezy and List (2006), have failed to identify the distinct contributions of reciprocity and mood.

This study attempts to fill this gap. Drawing on previous field studies, I conducted a field experiment with student workers in Bogotá, who earned a fixed wage for a one-off job involving data entry tasks. Crucially, unlike most previous studies, this study manipulated the period between the pay rise announcement and the work shift: some workers began working immediately after the announcement, whereas others began working a week later. In the latter group, the delay prior to the commencement of the work shift allowed sufficient time for the boost in mood resulting from the announcement to dissipate. This is the key feature of the study design that allowed clear separation of the reciprocity and mood motives. I complemented the variation in the period between the pay rise announcement and the commencement of the work shift with an elicitation of mood using the Self-Assessment Manikin (Bradley & Lang, 1994).

In the experiment, the subjects participated in two work sessions, which occurred 1 week apart. At the recruitment stage, they were informed that they would be paid at the end of the second session if they completed both sessions. The subjects were randomly assigned to one of three treatment groups. Those in the HighWage treatment group were offered 42,000 Colombian pesos (COP) (equivalent to approximately 12.5 USD). By contrast, subjects in the Rise-1 and Rise-2 treatment groups were initially offered a base wage of 30,000 COP (approximately 9 USD) and later received an unexpected pay rise of 12,000 COP. These two treatments differed in terms of the *timing* of the announcement of the rise. Subjects in the Rise-1 treatment group were promised the rise *at the end of session 1*, whereas those in the Rise-2 treatment group were promised the rise *at the beginning of session 2*.

This experimental design drew on (Gilchrist et al., 2016) by holding the final wage constant across treatment groups. This feature allowed me to isolate the effect on gift exchange of a pay rise announcement itself (a salient gift) from that of a higher wage. Because this required hiring subjects at different base wages, differential self-selection across treatment groups was a potential concern. I used the data on productivity in session 1 tasks, which preceded the pay rise announcement, to address this concern. Importantly, the analysis indicated that differential self-selection across treatment groups was not an empirical problem.

To assess the effects of the pay rise on productivity, it is necessary to focus on the work done by subjects in session 2. The gift-exchange hypothesis implies higher average productivity under the Rise-1 and Rise-2 treatments than under the HighWage treatment. To identify the contribution of positive reciprocity to gift exchange, I compared productivity in the Rise-1 and HighWage treatment groups, whereas to identify the contribution of positive mood, I compared productivity in the Rise-2 and Rise-1 treatment groups.

This identification strategy relies on three premises: (1) mood improved significantly immediately after the pay rise announcement, (2) the boost in mood vanished after a week, and (3) the level of reciprocity remained the same after a week. Elicited mood patterns confirmed the first two premises. The third premise is implied by the conventional view of reciprocity as a social norm. This view states that the unexpected pay rise triggered a sense of obligation to reciprocate among workers (Malmendier & Schmidt, 2017), which by its very nature was still present a week later. In Section 4.4.2, I show that the empirical findings of this study are consistent with this premise rather than with the alternative premise that reciprocity waned after a week as a result of adaptation to the higher wage.

Turning to the results, the pay rise announcement led to higher worker productivity, both in the work shift that began soon after the announcement (the Rise-2 treatment group) and in the shift that began a week later (the Rise-1 treatment group). Furthermore, the increase in productivity was larger in the Rise-2 treatment group. These findings suggest that both positive reciprocity and an improved mood played a role in gift exchange. While most of the increase in productivity was driven by reciprocity, the boost in mood had a significant contribution in relation to the work shift that began shortly after the announcement, accounting for about 25% of the effect.

This study complements that of Gagnon and Noussair (2020), who also investigated how gift exchange is affected by the delay between the moment when workers learn what their wage will be and the moment when they choose their level of effort, albeit in a different setting. Using a series of laboratory experiments, that study compared the effort choices of “short-term” workers, who made their choices within 3 h of learning what their wage was to be, with those of “long-term” workers, who made their choices 4 weeks after receiving their wage offer. That study found a strong positive wage–effort relationship in the short run and a significantly weaker relationship in the long run.

Gagnon and Noussair’s (2020) study differs from the present study in four key ways. First, following the laboratory gift exchange paradigm pioneered by Fehr et al. (1993), workers chose an artefactual effort level based on a known effort–cost schedule, instead of performing real-effort tasks in a work environment. Second, there were no pay rises because employers were required to make a single wage offer. Third, wages differed among workers, whereas I compared workers who received the same final wage. Fourth, while the researchers measured workers’ emotions during the laboratory session in which the workers learned what their wage was to be, emotional states were not measured again when the workers made their effort choices outside the laboratory. Therefore, the finding that gift exchange was weaker in the long term than in the short term is only suggestive of the role of emotions. The present study builds on this insight by providing direct evidence of the distinct influences of reciprocity and mood on gift exchange.

The rest of the paper is organized as follows. Section 2 presents the conceptual framework, which lays the groundwork for the field experiment. Section 3 describes the experiment, the identification strategy, and the data obtained. Section 4 presents the main results, assesses their robustness, and then rules out potential confounding factors. Section 5 concludes and compares the findings with results from previous studies.

2. Conceptual framework

In this section, I present a simple model of worker effort based on that of DellaVigna et al. (2022). I build on the previous model by incorporating mood into the worker’s utility function and distinguishing its effects on gift exchange from those of positive reciprocity. In the field experiment, effort is interpreted as productivity.

At $t = 0$, an employer hires two workers for a one-off job at a fixed wage. Worker L is offered w_L and Worker H is offered $w_H > w_L$. Wages are private information. At $t = 1$, before any work is done, the employer increases Worker L’s wage to w_H. The work shift begins at $t = T$, where $T \geq 1$. The workers perform identical individual tasks.

The workers have social preferences toward the employer, and thus care about the employer’s payoff, which increases with a worker’s effort. The satisfaction a worker derives from exerting effort e is given by $A(T) \cdot e$, where $A(T)$ represents the strength of the social preference at $t = T$.[3] Specifically,

$$A(T) = a + m + a_{rise} \cdot \mathbb{1}_{rise} + m_{rise}(T) \cdot \mathbb{1}_{rise}. \quad (1)$$

[3] An alternative specification is $A(T) \cdot \Pi(e, w_H)$, where $\Pi(e, w_H)$ denotes the employer’s return from the worker’s effort e given w_H. The simpler

The first two terms on the right-hand side of expression (1) capture factors that affect social preferences regardless of the announcement of a pay rise. The parameter $a > 0$ denotes baseline social preferences, which capture altruism, utility from exerting effort for the employer, or satisfaction in doing meaningful work (Ariely et al., 2008). The parameter $m \in \mathbb{R}$ denotes baseline mood. The assumption that mood positively affects social preferences is grounded on previous psychological research showing that individuals who feel good are more likely to help others (Carlson et al., 1988; Isen & Levin, 1972; Levin & Isen, 1975; Weyant, 1978).

The last two terms on the right-hand side of expression (1) highlight how the announcement of a pay rise affects social preferences. The indicator $\mathbb{1}_{rise}$ equals 1 if the worker received a pay rise (Worker L) and 0 if he or she received the same final wage directly (Worker H). The pay rise is perceived as a gift by Worker L because it was unexpected given the original work conditions (Macera & te Velde, 2018; Malmendier & Schmidt, 2017). Such a gift strengthens Worker L's social preferences toward the employer through two different channels. First, it triggers positive reciprocity, $a_{rise} > 0$. Second, it boosts mood, $m_{rise}(T) \geq 0$.[4]

The model captures an important difference in the endurance of reciprocity and mood. Because reciprocity stems from a sense of obligation to repay the gift (Malmendier & Schmidt, 2017), its intensity does not depend on when the work begins. (In Section 4.4.2, I further discuss this premise in light of the empirical results.) By contrast, the boost in mood depends on T (Isen et al., 1976). Specifically, mood improves immediately after the pay rise announcement (at $t = 1$) and then wanes. That is, $m_{rise}(1) > 0$, $m_{rise}(T)$ decreases with T, and there exists some $\hat{T} > 1$ such that $m_{rise}(T) = 0$ for all $T \geq \hat{T}$.

Each worker chooses effort e to maximize his or her utility, balancing his or her pay with effort on the employer's behalf. The workers' problem is given by

$$Max_{e \geq 0} \quad U(e) = w_H + A(T) \cdot e - C(e), \tag{2}$$

where $C(.)$ represents the cost of effort. I assume the regularity conditions $C'(.) > 0$, $C''(.) > 0$, and $\lim_{e \to \infty} C'(e) = \infty$, which guarantee the existence of a unique solution to (2). The first-order condition is

$$a + m + (a_{rise} + m_{rise}(T)) \cdot \mathbb{1}_{rise} - C'(e) = 0, \tag{3}$$

which implies that the optimal effort of each worker is given by

$$e_H^* = C'^{-1}(a + m); \tag{4}$$

$$e_L^* = C'^{-1}(a + m + a_{rise} + m_{rise}(T)). \tag{5}$$

Given the assumptions regarding $C(.)$, $C'^{-1}(.)$ exists and is strictly increasing. Furthermore, the second-order condition is satisfied because $-C''(e^*) < 0$. Assuming an interior solution, we can analyze the reduced-form effect of the pay rise on effort by comparing expressions (4) and (5).

When $T < \hat{T}$, both a_{rise} and $m_{rise}(T)$ are positive, and thus the effect of the pay rise through the reciprocity channel is confounded by the effect through the boost in mood. To isolate the effect through the reciprocity channel, we must compare Worker L's effort with Worker H's effort when the work starts at $T \geq \hat{T}$, so that $m_{rise}(T) = 0$:

$$e_L^*(T \geq \hat{T}) - e_H^*(T \geq \hat{T}) = C'^{-1}(a + m + a_{rise}) - C'^{-1}(a + m). \tag{6}$$

To identify the (full) effect of the pay rise through the boost in mood, we must compare Worker L's effort when the work starts at $T = 1$ with the effort he or she exerts when the work starts at $T \geq \hat{T}$:

$$e_L^*(T = 1) - e_L^*(T \geq \hat{T}) = C'^{-1}(a + m + a_{rise} + m_{rise}(1)) - C'^{-1}(a + m + a_{rise}). \tag{7}$$

3. Experiment

3.1. General procedures

A field experiment was carried out between October 2019 and March 2020 at Universidad del Rosario, Bogotá, using subjects drawn from the student pool managed by the Rosario Experimental and Behavioral Economics Laboratory (REBEL). In addition to participating in the study on gift exchange reported in this paper, the subjects participated in a separate study on charitable giving, which is reported in Sautua (2022). Here, I summarize the general procedures of the experiment. The Online Appendix contains all experimental instructions and procedures translated into English.

Subjects from various majors were invited by email to participate in two 1-hour sessions, 1 week apart, at the REBEL. Invitations were sent at various times, and only a subset of randomly chosen individuals from the large subject pool received the invitation each time. This recruitment strategy reduced the chances that friends or acquaintances ended up participating in different treatment groups and talking about the study between sessions 1 and 2, which could have generated contamination across treatments.[5] The invitation—which was sent about 4 days before the first of the two sessions—offered subjects a fixed wage for their participation and informed them that they would be paid at the end of the second session if they completed both sessions. The message did not contain any specific information about the study other than the wage.

On arrival at the laboratory, each subject was seated at a separate carrel in front of a computer terminal with an Internet browser. Although several subjects worked at the same time, they were not allowed to talk to each other and could not observe each other's work (because carrels were separated by panels), which eliminated productivity spillovers across peers within a session (Mas & Moretti, 2009). Subjects were informed that they would work on four tasks—two in each session—to assist researchers from the Economics Department at Universidad del Rosario in several research projects. To record his or her work on each task, each subject created a unique personal code that was known only to him or her, which guaranteed the anonymity of the information collected. Such anonymity removed reputation or repeated game effects.

All sessions were conducted by the same group of instructors. Task instructions followed a predetermined protocol, and each task lasted for 15 min. In session 1, the subjects commenced work on task 1, which consisted of measuring the lengths of the second and fourth fingers on the hands of several people using a set of recorded images and the AutoMetric software package. Once both fingers on a hand had been measured, the software automatically calculated the ratio of the lengths, that is, the 2D:4D ratio. The subjects then worked on task 2, which consisted of typing the responses from a few paper-and-pencil surveys into a Google Form. A week later, in session 2, the subjects worked on task 3, which was similar to task 2, and on task 4, in which they provided subjective answers to various multiple-choice questions about social norms.

version was adopted because the value to the employer of the worker's effort was unobserved by the subjects and held constant in the field experiment. See DellaVigna et al. (2022) for an experiment that manipulated the return to the employer.

[4] Malmendier and Schmidt (2017) model reciprocity in a similar way, building on economic models of action-based reciprocity. The utility of individual i depends on the material payoffs of other individuals. The weight that individual i (here, the worker) attaches to the welfare of individual j (here, the employer) depends on the actions of j that affect i, relative to the expected behavior of j. A favorable act that is unexpected (such as granting a pay rise) strengthens the bond between j and i, i.e., the weight that i attaches to j's payoff, inducing i to reciprocate.

[5] In Section 4.4.1, I argue that cross-contamination through social interaction was not an empirical problem in light of the main findings of this study.

Table 1
Summary of experimental design.

Treatment	Base wage (COP)	Final wage (COP)	Time of pay rise announcement
HighWage	42,000	42,000	–
Rise-1	30,000	42,000	End of session 1 (after task 2)
Rise-2	30,000	42,000	Start of session 2 (before task 3)

Tasks 1, 2, and 3 required objective data entry, and hence enabled measurement of both the quantity and quality of output. By contrast, there was no room for mistakes in task 4 given the subjective nature of the answers, and almost all subjects completed the task. For these reasons, hereafter I focus exclusively on tasks 1–3.

3.2. Measures of work performance

In task 1, the subjects could measure the lengths of up to 60 fingers; in task 2, they could type in the responses from up to three surveys, with 25 responses for each survey; and in task 3, they could type in the responses of up to two surveys (which were different from those of task 2), with 61 responses for each survey. Each response consisted of a few words or numbers. For each task, I used the proportion of measurements (task 1) or responses (tasks 2 and 3) recorded out of the maximum feasible number of measurements/responses as the main performance metric.

One limitation of tasks 2 and 3 is that, by design of the Google Forms, partial work was not accurately recorded by the computer. Specifically, the subjects' work on a given survey was *not* recorded *unless* he or she finished typing in the responses from the entire survey and clicked on a button on the Google Form to submit his or her work. As a result, the records of work performance in tasks 2 and 3 were coarser than those in task 1, where partial work was accurately recorded. For this reason, in tasks 2 and 3 I complemented the main performance metric—the proportion of responses recorded—with an additional measure, namely whether a subject typed in the responses from all of the available surveys. The results remain the same if I use the number of surveys completed instead of the probability of finishing the task as the complementary performance metric.

To conduct robustness checks of the results, I explored the proportion of *correct* measurements/responses out of the maximum feasible number of measurements/responses. This metric accounted for both the quantity and quality of the subjects' work. In tasks 2 and 3, a recorded response was classified as correct if all characters were correctly typed.[6]

3.3. Treatments

The subjects were allocated to one of three treatment groups, which varied in terms of three dimensions: (1) the base wage, (2) whether subjects were promised a pay rise, and (3) the timing of the pay rise announcement. Table 1 summarizes the experimental design.[7]

The recruitment message was randomized across subjects in the pool managed by the REBEL. Subjects in the HighWage treatment group were offered 42,000 COP (equivalent to approximately 12.5 USD). By contrast, those in the Rise-1 and Rise-2 treatment groups were initially offered 30,000 COP (approximately 9 USD). Once subjects in the Rise-1 treatment group had finished session 1, they were told that because of the availability of additional funds their initial wage of 30,000 COP would be increased by 12,000 COP if they finished session 2. Subjects in the Rise-2 treatment group were promised the same pay rise at the beginning of session 2, before they had undertaken any additional work. The experimental design draws on Gilchrist et al. (2016), who also compared subjects who were promised a pay rise with others who received the same final wage directly. The present design augments that of Gilchrist et al. (2016) by manipulating the period between the rise announcement and the work shift.

Several features of the experimental design are noteworthy.

First, the pay rise was explicitly linked to the availability of additional funds to make clear that it was a gift rather than an adjustment due to another reason, for example, that the employer realized the work was harder than expected, or that he or she was instructed by a higher authority to increase the wage. The wording used in the announcement of the pay rise followed Gilchrist et al. (2016).

Second, the oral announcement of the pay rise was reinforced with a written message on a board situated at the front of the room. This message remained on the board when the subjects returned to the room one week later, thus allowing subjects in the Rise-1 treatment group to remember that their final wage included a gift when they resumed their work in session 2. By removing memory issues, this feature of the design kept the intensity of reciprocity constant between the Rise-1 and Rise-2 treatment groups.

Third, I chose a 40% wage increase because it appeared to be large enough to affect reciprocity and mood, but not too large to raise suspicion among subjects about the employer's motives.[8]

Fourth, all subjects who attended a given work session received the same information about tasks and wages, and were not given any information about other groups. This feature of the design removed social comparison effects that could result from wage differences across workers (Abeler et al., 2010; Charness et al., 2016; Charness & Kuhn, 2007; Cohn et al., 2014; Gächter & Thöni, 2010; Hennig-Schmidt et al., 2010).

3.4. Elicitation of mood

The subjects' mood was elicited using the Self-Assessment Manikin (SAM) (Bradley & Lang, 1994), a well-known and validated non-verbal instrument. The SAM requires subjects to select one of nine manikins that best expresses their current feelings, ranging from very negative (encoded as –4) to very positive (encoded as 4). Subjects completed the SAM twice, once in each session. In session 1, completion of the SAM followed the pay rise announcement in the Rise-1 treatment group and task 2 under all treatments. In session 2, completion of the SAM followed the pay rise announcement in the Rise-2 treatment group and preceded task 3 under all treatments.

3.5. Identification strategy

This study focused on the effects of the announcement of the pay rise on gift exchange, which required an assessment of work performance in session 2 (i.e., task 3). In this section, I discuss how the experimental design allowed me to evaluate the specific roles of

[6] To check whether all characters of a response were correctly typed, a research assistant compared the recorded response with the original response from the paper form.

[7] Each treatment comprised several sessions. To minimize calendar effects, sessions were randomly distributed over several days of the week.

[8] See Macera and te Velde (2018) for further discussion on the importance of the credibility of the gift. The size of pay rises in data entry jobs using student subjects varies widely in the gift exchange literature. The 40% rise used in this experiment was intermediate in size, as most rises have ranged from 19% (e.g., in Kube et al., 2012) to 67% (e.g., in Gneezy & List, 2006). In one of her treatments, Esteves-Sorenson (2018) implemented a 100% pay rise.

positive reciprocity and mood after the pay rise. As explained below in Section 3.7, I used the data on work performance in tasks 1 and 2, which preceded the pay rise announcement, to rule out self-selection effects.

3.5.1. The role of positive reciprocity after the pay rise

To assess the role of positive reciprocity after the pay rise, I compared average productivity in task 3 between the Rise-1 and HighWage treatment groups. When the subjects started working on task 3, their final wage was the same under both treatments, but subjects in the Rise-1 treatment group had learned about the pay rise a week earlier. Crucially, as we shall see in Section 4.1, the average mood at the beginning of session 2 was the same under both treatments, indicating that any boost in mood caused by the pay rise announcement disappeared after a week. Additionally, the week-long gap between the pay rise announcement and the work shift that characterized the Rise-1 treatment eliminated fatigue and provided subjects with sufficient time to reflect on how to reciprocate in response to the pay rise (Esteves-Sorenson, 2018). Therefore, drawing on expression (6) from Section 2, the difference in average productivity between the Rise-1 and HighWage treatment groups identified the effect of the pay rise on productivity through the reciprocity channel.

3.5.2. The role of mood after the pay rise

To assess the role of mood after the pay rise, I compared average productivity in task 3 between the Rise-2 and Rise-1 treatment groups. When the subjects started working on task 3, both treatment groups had already been promised a pay rise. However, subjects in the Rise-2 treatment group had only heard the announcement a few minutes earlier, whereas those in the Rise-1 treatment group had heard it 1 week earlier. As we shall see in Section 4.1, the average mood at the beginning of session 2 was significantly higher in the Rise-2 treatment group than in the Rise-1 treatment group. Drawing on expression (7) from Section 2, this implies that the difference in average productivity between these two treatment groups identified the effect of the pay rise through the boost in mood.

3.6. Sample description

As mentioned earlier in Section 3.1, the full sample was used not only for the present study on gift exchange but also for a separate study on charitable giving. To accommodate the experimental design of the latter study, the sample was first randomly divided into five groups. For present purposes, the Rise-1 treatment comprised three of such groups, whereas the HighWage and Rise-2 treatments each comprised one group. For this reason, the Rise-1 treatment group was roughly three times as large as the others.

The full sample consisted of 397 subjects who completed at least session 1. Of these, 359 (90.4%) completed both sessions and there were no statistically significant differences in the attrition rate across treatments ($p = 0.884$, Fisher's exact test; see Table A1 in the Online Appendix for details).[9] To ensure consistency, hereafter the sample is restricted to the 359 subjects who completed both sessions. Therefore, the study sample consisted of 72 subjects in the HighWage treatment group, 223 in the Rise-1 treatment group, and 64 in the Rise-2 treatment group. Table A2 in the Online Appendix presents the summary statistics regarding the subjects' demographics (gender, age, number of semesters of study, and major) and self-reported risk, time, and prosocial preferences, which I elicited using a subset of questions from the streamlined preference survey module of Falk et al. (2016).[10] Demographics and self-reported preferences were used as control variables in the main empirical analysis to boost statistical power.

3.7. Ruling out self-selection effects

In principle, self-selection into the treatment groups might be a concern because the base wage was higher in the HighWage treatment group than it was in the Rise-1 and Rise-2 treatment groups. In particular, although subjects did not know anything about the tasks at the recruitment stage, different base wages could have attracted individuals with different ability or inclination to work. This could have confounded the effect of the pay rise on productivity in session 2, which was the focus of this study. Therefore, before proceeding to the main findings, I assess whether differential self-selection was an issue.

To this end, I first tested for treatment differences in average productivity in tasks 1 and 2, which preceded the pay rise announcement. If there was differential self-selection of workers across treatment groups, we should observe a significant difference in average productivity in session 1 tasks between the HighWage group and the other groups.

Table A3 in the Online Appendix shows the descriptive statistics for work performance by treatment group and task. To test for selection effects, I pooled observations from the Rise-1 and Rise-2 treatment groups because both groups featured the same base wage and hence were identical before the announcement of the pay rise.

In task 1, the subjects recorded, on average, 92.92% of the available measurements in the HighWage group and 90.75% of the measurements in the Rise-1 and Rise-2 groups; additionally, the mean percentage of correct measurements was 55.16 in the HighWage group and 55.81 in the Rise-1 and Rise-2 groups. Thus, there were no statistically significant differences between groups in terms of either the quantity of output or the percentage of correct measurements ($p = 0.534$ and $p = 0.806$, respectively, using Wilcoxon rank-sum tests).[11]

In task 2, the subjects recorded, on average, 98.15% of the available responses in the HighWage group and 98.14% of the responses in the Rise-1 and Rise-2 groups; additionally, the mean percentage of correct responses was 94.87 in the HighWage group and 94.33 in the Rise-1 and Rise-2 groups. Hence, there were no statistically significant differences between treatment groups in terms of either the quantity of output or the percentage of correct responses ($p = 0.900$ and $p = 0.723$, respectively, using Wilcoxon rank-sum tests).

As an additional test for differential self-selection, I examined whether different base wages attracted individuals with different self-reported preferences. I conducted an omnibus test based on a logit model wherein I regressed a dummy variable for the HighWage treatment group on self-reported preferences. The null hypothesis that subjects did not differentially self-select into the HighWage treatment group implies that all coefficients in the regression equal zero. The omnibus test did not reject the null hypothesis at conventional significance levels ($\chi^2 = 4.53$, $p = 0.476$).[12]

In summary, there were neither differences in productivity in session 1 nor differences in self-reported preferences between treatment groups. This suggests that potential self-selection into the HighWage

[9] A likely reason for the low attrition rate in all treatment groups is that subjects knew in advance that they had to complete both sessions to be paid (as mentioned earlier in Section 3.1).

[10] I included the second question from each of the risk-taking, time-discounting, altruism, and positive reciprocity sections, as well as the question on trust. See the experimental instructions in the Online Appendix for the exact wording of each question.

[11] Twenty-nine subjects in the Rise-1 treatment group and 4 subjects in the Rise-2 group had to be excluded from this comparison because their task 1 records were lost as a result of a technical glitch.

[12] The p-values of z-tests for each coefficient are as follows: reciprocity, $p = 0.854$; altruism, $p = 0.997$; trust, $p = 0.373$; risk, $p = 0.155$; patience, $p = 0.980$.

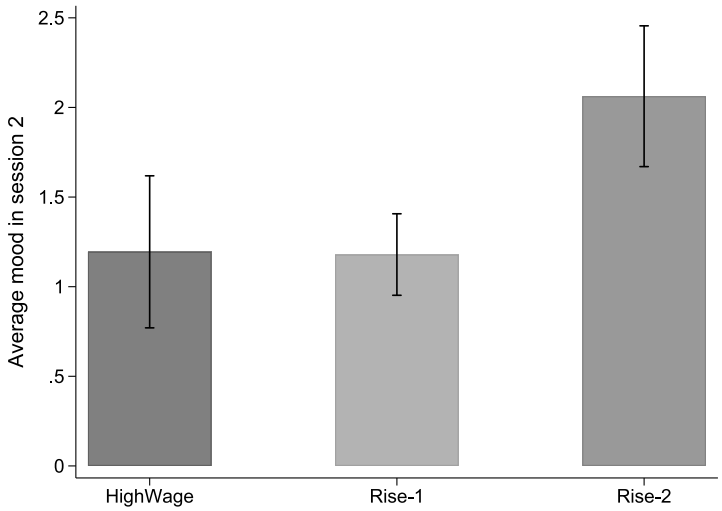

Fig. 1. Average mood in session 2 by treatment group. *Notes:* Mood ranges from –4 (very negative) to 4 (very positive). Bars indicate 95% confidence intervals for the means.

treatment was not an empirical problem and hence did not affect the main results regarding session 2.[13]

4. Results

4.1. Effect of the pay rise on mood

Fig. 1 displays the average mood at the beginning of session 2 by treatment group and reveals two key facts. First, as expected, the average mood was significantly higher in the Rise-2 treatment group than in the Rise-1 treatment group ($p < 0.001$, one-tailed t-test), suggesting that mood improved immediately after the pay rise announcement and then waned.[14] Second, the average mood was virtually identical in the Rise-1 and HighWage treatment groups ($p = 0.950$, two-tailed t-test), implying that any difference in productivity between these two groups must be attributed to positive reciprocity. (Table A4 in the Online Appendix shows the descriptive statistics for mood by treatment group and session.)

To further analyze these patterns, controlling for individual differences in demographics and self-reported preferences, I estimated the following regression model using ordinary least squares (OLS):

$$Mood_i = \gamma_0 + \gamma_1 R1_i + \gamma_2 R2_i + X_i' \gamma_3 + \epsilon_i. \quad (8)$$

The dependent variable, $Mood_i$, was subject i's self-reported mood at the beginning of session 2. The dummy variable $R1_i$ ($R2_i$) equals 1 if subject i was promised a pay rise in session 1 (2) and 0 otherwise. The HighWage treatment group constitutes the reference category. X_i is a vector that includes subject i's demographics and self-reported preferences. Finally, ϵ_i is an error term.

The results confirmed the patterns shown in Fig. 1. The estimate of γ_1 was –0.130 (s.e. 0.243), and was not statistically different from zero ($p = 0.593$, two-tailed t-test). The estimate of γ_2 was 0.745 (s.e. 0.293), which was statistically different from zero ($p = 0.011$, two-tailed t-test) and significantly larger than the estimate of γ_1 ($p < 0.001$, one-tailed t-test).

Overall, the patterns observed in relation to mood strongly supported the strategy that was proposed earlier to disentangle the reciprocity- and mood-driven effects of the pay rise on productivity.

4.2. Effect of the pay rise on productivity in session 2

Turning to productivity in session 2, Panel A of Fig. 2 shows the average percentage of responses recorded in task 3 by treatment group. Output was highest in the Rise-2 treatment group (100%), followed by the Rise-1 group (98.21%) and the HighWage group (93.75%).[15] As shown in Panel B of Fig. 2, these differences in the quantity of output resulted from differences in the probability of completing all of the surveys. The percentage of subjects who finished task 3 was highest in the Rise-2 treatment group (100%), followed by the Rise-1 group (96.41%) and the HighWage group (87.50%). (All the subjects managed to complete at least one of the two available surveys.)[16]

To further analyze the effects of the pay rise announcement on productivity in session 2, I estimated two regression models using OLS:

$$y_i = \beta_0 + \beta_1 R1_i + \beta_2 R2_i + X_i' \beta_3 + \epsilon_i. \quad (9)$$

The dependent variable y_i varies across models as follows: (1) the proportion of responses recorded by subject i in task 3 (main performance metric); and (2) a dummy variable that equals 1 if subject i finished task 3 and 0 otherwise (complementary performance metric). All independent variables are the same as for Eq. (8). The hypothesis that subjects in the Rise-1 treatment group responded to the pay rise with better performance than those in the HighWage treatment group implies that $\beta_1 > 0$. The hypothesis that subjects in the Rise-2 treatment group were even more productive than those in the Rise-1 treatment group because the more recent announcement had significantly improved their mood implies that $\beta_2 > \beta_1$.

Columns 1 and 2 of Table 2 display the results. Consider first the estimates of the treatment effects on the proportion of responses recorded in task 3 (column 1). The estimate of β_1 is positive and statistically significant ($p = 0.014$, one-tailed t-test). Compared with the HighWage treatment group, the proportion of responses recorded in the Rise-1 treatment group was, on average, 4.4 percentage points higher.

[13] The results regarding session 1 are consistent with those of Gilchrist et al. (2016), who found no productivity differences between two groups of workers whose wages differed by 33%.

[14] Further evidence that mood improved immediately after the pay rise announcement is provided by mood data from session 1: at the end of that session, the average mood was significantly higher in the Rise-1 group than in the Rise-2 group ($p < 0.001$, one-tailed t-test). Moreover, subjects in the Rise-1 group were happier at the end of session 1—having just learned about the pay rise—than they were at the beginning of session 2 ($p = 0.001$, one-tailed paired t-test), which provides further evidence that the boost in mood was temporary.

[15] The p-values of one-tailed Wilcoxon rank-sum tests are as follows: Rise-2 vs Rise-1, $p = 0.062$; Rise-1 vs HighWage, $p = 0.002$.

[16] The p-values of one-tailed χ^2-tests are as follows: Rise-2 vs Rise-1, $p = 0.062$; Rise-1 vs HighWage, $p = 0.002$.

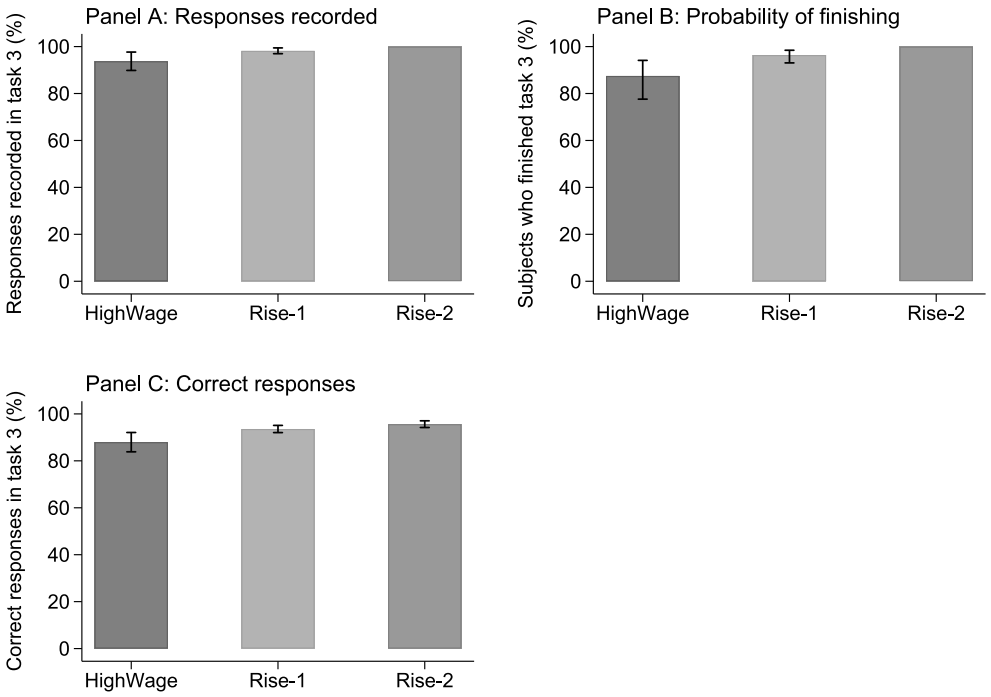

Fig. 2. Productivity in task 3 by treatment group. *Notes:* Panel A shows the average percentage of responses recorded in task 3 by treatment group. Panel B shows the percentage of subjects who finished task 3 by treatment group. Panel C shows the average percentage of correct responses in task 3 by treatment group. In Panels A and C, bars indicate 95% confidence intervals for the means. In Panel B, bars indicate 95% Clopper–Pearson confidence intervals for percentages.

Table 2
Treatment effects on productivity in session 2.

	(1) Responses	(2) Finish	(3) Responses	(4) Finish	(5) Correct
R1	0.044**	0.087**	0.044**	0.088**	0.051**
	(0.020)	(0.040)	(0.020)	(0.040)	(0.021)
R2	0.058***	0.115***	0.057***	0.115***	0.069***
	(0.019)	(0.037)	(0.019)	(0.037)	(0.021)
Donation			0.018	0.036	
			(0.033)	(0.066)	
Constant	0.933***	0.865***	0.929***	0.858***	0.904***
	(0.072)	(0.144)	(0.073)	(0.145)	(0.094)
Control variables:					
Demographics	Yes	Yes	Yes	Yes	Yes
Self-reported preferences	Yes	Yes	Yes	Yes	Yes
Mean of dep var in HighWage	0.94	0.88	0.94	0.88	0.88
Number of subjects	358	358	358	358	358
Hypothesis tests (p-values):					
H_o: $\beta_1 \leq 0$	0.014	0.014	0.014	0.014	0.007
H_o: $\beta_2 \leq 0$	0.001	0.001	0.001	0.001	0.001
H_o: $\beta_2 \leq \beta_1$	0.017	0.017	0.020	0.020	0.048

Notes: Columns 1 and 2 show the OLS estimates of treatment effects on the proportion of responses recorded in task 3 and the probability of finishing task 3, respectively. The econometric specifications are given by Eq. (9). Demographic controls include gender, age, number of semesters of study, and a dummy variable that equals 1 if a subject is an economics/finance major and 0 otherwise. Self-reported risk, time, and pro-social preferences were elicited using a subset of questions from the streamlined preference survey module of Falk et al. (2016). Columns 3 and 4 extend the specifications in columns 1 and 2 by including a subject's donation (as a proportion of the final wage) in a separate study as an additional control variable. Column 5 displays the OLS estimates of treatment effects on the proportion of correct responses in task 3. In all columns, robust standard errors are shown in parentheses. One subject in the Rise-1 treatment group was excluded because of missing demographics data.

**$p < 0.05$, using a two-tailed t-test.

***$p < 0.01$, using a two-tailed t-test.

This finding indicates that subjects who received a pay rise at the end of session 1 displayed positive reciprocity a week later when they returned to work. As expected, the estimate of β_2 is also positive and statistically significant ($p = 0.001$, one-tailed t-test). The proportion of responses recorded in task 3 was, on average, 5.8 percentage points higher in the Rise-2 treatment group than it was in the HighWage treatment group. Moreover, the estimate of β_2 was significantly larger than that of β_1 ($p = 0.017$, one-tailed t-test). That is, subjects increased their effort even further when the pay rise announcement immediately preceded the work shift. This additional increase in the quantity of output provides

evidence that the boost in mood caused by the pay rise announcement had a direct effect on performance.[17]

Column 2 shows the estimates of the treatment effects on the probability of completing task 3. Consistent with the results reported in column 1, the probability of finishing task 3 was significantly higher in the Rise-1 and Rise-2 treatment groups than it was in the HighWage treatment group ($p = 0.014$ for the Rise-1 group and $p = 0.001$ for the Rise-2 group, one-tailed t-tests). Additionally, the increase in the Rise-2 treatment group (11.5 percentage points) was larger than that in the Rise-1 treatment group (8.7 percentage points) ($p = 0.017$, one-tailed t-test).

Overall, the results showed that the pay rise announcement increased productivity in session 2 both by inducing positive reciprocity and by lifting mood. While reciprocity had the largest influence, the boost in mood made a significant contribution, accounting for approximately 25% of the effect in the Rise-2 treatment group.[18] Note that, given the high productivity of subjects in the Rise-1 treatment, the additional effect of mood in the Rise-2 treatment might have been constrained by a ceiling effect. Indeed, subjects in the Rise-2 treatment achieved maximum output. The contribution of mood to performance could be larger in harder tasks in which there is more room for productivity increases.

4.3. Robustness checks

4.3.1. Controlling for giving behavior in the separate study

As mentioned earlier, the subjects also participated in a simultaneous study on charitable giving. They were asked to make a donation out of their final wage to a charitable organization before working on task 3. In principle, such donation decision could have affected gift exchange in session 2, possibly contaminating treatment differences in relation to productivity.

To account for the potential influence of giving behavior on subsequent work performance, I reestimated Eq. (9) including subject i's donation (as a proportion of the final wage) in the separate study as an additional control variable. Columns 3 and 4 in Table 2 show that the results from columns 1 and 2 are strongly robust to the inclusion of this additional regressor. Moreover, the estimate of the coefficient for donations is not statistically different from zero ($p = 0.590$ both in column 3 and in column 4). In summary, the observed differences between treatment groups in relation to productivity in session 2 were not contaminated by preceding giving behavior.

4.3.2. Accounting for the quality of output

In addition to investigating the treatment effects on the quantity of output, I explored the effects on output quality. Panel C of Fig. 2 shows the mean percentage of *correct* responses in task 3 by treatment group. This is a composite measure of work performance that accounts for both the quantity and quality of work. Consistent with the observed differences across treatment groups in the quantity of output, the mean percentage of correct responses was highest in the Rise-2 group (95.63%), followed by the Rise-1 group (93.56%) and the HighWage group (87.97%).

To assess these differences in a regression framework, I reestimated Eq. (9) using the proportion of correct responses in task 3 as the dependent variable. Column 5 in Table 2 displays the results. Consistent with the treatment effects on the quantity of output, the proportion of correct responses was significantly higher in the Rise-1 and Rise-2 treatment groups than it was in the HighWage treatment group ($p = 0.007$ for the Rise-1 group and $p = 0.001$ for the Rise-2 group, one-tailed t-tests). Furthermore, the increase in the Rise-2 treatment group (6.9 percentage points) was larger than that in the Rise-1 treatment group (5.1 percentage points) ($p = 0.048$, one-tailed t-test). Again, the contribution of mood to the increase in productivity in session 2 was approximately 25% ((6.9–5.1)/6.9 = 0.261).

4.3.3. Accounting for serial correlation in performance across sessions

I also explored the robustness of the results to accounting for potential serial correlation in performance across sessions at the individual level. Pooling observations from both sessions, I estimated the following regression model using least squares:

$$y_{it} = \alpha_0 + \alpha_1 R1_i + \alpha_2 R2_i + \alpha_3 \mathbb{1}_{t=3} + \alpha_4 \left(R1_i \times \mathbb{1}_{t=3}\right) + \alpha_5 \left(R2_i \times \mathbb{1}_{t=3}\right) + X_i' \alpha_6 + \epsilon_i. \quad (10)$$

The dependent variable y_{it} is the proportion of responses recorded by subject i in task t, where $t \in \{1, 2, 3\}$. $R1_i$, $R2_i$, and X_i are the same as for Eqs. (8) and (9). $\mathbb{1}_{t=3}$ is a dummy variable that equals 1 if $t = 3$ and 0 otherwise. Tasks 1 and 2 combined, which took place in session 1, are the omitted category. Standard errors were clustered at the subject level. In this specification, α_1 (α_2) captures the mean difference in output in session 1 between the Rise-1 (Rise-2) and HighWage groups, whereas $\alpha_1 + \alpha_4$ ($\alpha_2 + \alpha_5$) captures the mean difference in output *in task 3* between the Rise-1 (Rise-2) and HighWage groups. Additionally, $\alpha_2 + \alpha_5 - \alpha_1 - \alpha_4$ captures the mean difference in output *in task 3* between the Rise-2 and Rise-1 groups.

Table 3 displays the results. First, the estimates of α_1 and α_2 are both statistically insignificant ($p = 0.172$ and $p = 0.370$, respectively, using two-tailed t-tests), complementing the evidence against self-selection effects discussed in Section 3.7. Second, the estimates of $\alpha_1 + \alpha_4$ and $\alpha_2 + \alpha_5$ were 0.045 and 0.060, respectively, and were virtually the same as those of β_1 and β_2 in Eq. (9) (see column 1 of Table 2). Third, and most important, accounting for potential serial correlation across sessions did not affect the statistical significance of the main results. Using one-tailed t-tests, the estimates of $\alpha_1 + \alpha_4$ and $\alpha_2 + \alpha_5$ were statistically significant at $p = 0.015$ and $p = 0.001$, respectively; and the estimate of $\alpha_2 + \alpha_5$ was larger than that of $\alpha_1 + \alpha_4$ at $p = 0.008$.

4.4. Addressing possible confounds

I conclude the results section by evaluating two potential confounds.

4.4.1. Contamination across treatments through social interaction

As mentioned earlier in Section 3.1, invitations to participate in this study were sent at various times to different random sub-samples of the subject pool. While this recruitment strategy helped to mitigate contamination across treatments, the latter was not completely eliminated by design. However, possible cross-contamination does not invalidate the results because it worked in the *opposite* direction of the predicted treatment effects and I found supporting evidence of these effects. Next, I elaborate on this argument.

First, consider how cross-contamination between sessions 1 and 2 could have affected the behavior of subjects in the Rise-1 treatment group compared with that of subjects in the HighWage group. The gift-exchange model discussed in Section 2 predicted that the pay rise announcement would trigger positive reciprocity in the Rise-1 group and hence increase output, because the workers would perceive the pay rise as a gift from their employer. However, if between sessions 1 and 2 subjects in the Rise-1 treatment group learned that their final wage was the same as the one the HighWage group had been offered directly, they might have perceived the pay rise as less kind than was predicted by the model. On the basis of the findings of Cohn et al. (2014), we would expect no *further* changes in performance in the Rise-1 group if

[17] Recall that the duration of each task was predetermined and held constant across treatments (see Section 3.1). Therefore, subjects in the Rise-1 and Rise-2 treatments improved their performance in task 3 by achieving more output in the same amount of time rather than by working for extra time.

[18] The contribution of mood was (5.8–4.4)/5.8 = 0.241 with regard to the proportion of responses recorded in task 3 and (11.5–8.7)/11.5 = 0.243 with regard to the probability of finishing this task.

Table 3
Treatment effects on productivity pooling all tasks.

	Responses
R1	−0.015
	(0.011)
R2	0.011
	(0.013)
$\mathbb{1}_{t=3}$	−0.018
	(0.021)
R1×$\mathbb{1}_{t=3}$	0.060***
	(0.023)
R2×$\mathbb{1}_{t=3}$	0.049**
	(0.023)
Constant	0.979***
	(0.052)
Control variables:	
Demographics	Yes
Self-reported preferences	Yes
Number of subjects	358
Number of observations	1041
Hypothesis tests (p-values):	
H_o: $\alpha_1 + \alpha_4 \leq 0$	0.015
H_o: $\alpha_2 + \alpha_5 \leq 0$	0.001
H_o: $\alpha_2 + \alpha_5 \leq \alpha_1 + \alpha_4$	0.008

Notes: This table shows the least-squares estimates of treatment effects on productivity pooling observations from both sessions. The econometric specification is given by Eq. (10). Demographic controls include gender, age, number of semesters of study, and a dummy variable that equals 1 if a subject is an economics/finance major and 0 otherwise. Self-reported risk, time, and pro-social preferences were elicited using a subset of questions from the streamlined preference survey module of Falk et al. (2016). Standard errors clustered by subject are shown in parentheses. One subject in the Rise-1 treatment group was excluded because of missing demographics data. Output in task 1 is missing for 29 subjects in the Rise-1 treatment group and 4 subjects in the Rise-2 group because their task 1 records were lost as a result of a technical glitch.
**$p < 0.05$, using a two-tailed t-test.
***$p < 0.01$, using a two-tailed t-test.

the subjects also learned that the Rise-2 group had not been promised a pay rise by the end of the first session.[19]

In summary, cross-contamination might have *attenuated* the influence of positive reciprocity, thereby biasing the average performance gap between the Rise-1 and HighWage groups downward compared to the prediction of the model. Nevertheless, the finding of a significant performance gap between these two groups despite such downward bias provides clear evidence that positive reciprocity played a significant role.

Now consider how cross-contamination between sessions 1 and 2 could have affected the behavior of subjects in the Rise-2 treatment group compared with that of subjects in the Rise-1 group. The gift-exchange model discussed in Section 2 predicted that the *unanticipated* pay rise announcement would boost the mood of subjects in the Rise-2 group and hence further increase output in session 2 compared with that of subjects in the Rise-1 group. However, if between sessions 1 and 2 subjects in the Rise-2 treatment group learned the final wage of subjects in the other groups, they might have anticipated a pay rise in session 2 (especially after learning about the pay rise in the Rise-1 group). This would have *attenuated* the impact of the pay rise announcement on mood, thus biasing the performance gap between the Rise-2 and Rise-1 groups downward. Nevertheless, the finding of a productivity gap between these two groups, despite the downward bias possibly induced by cross-contamination, indicates that mood played a role.

Overall, while contamination across treatments was not completely eliminated by design, it did not constitute an empirical problem in light of the clear evidence for the roles of positive reciprocity and mood in lifting worker productivity. If anything, cross-contamination might have attenuated the predicted treatment effects.

4.4.2. Mood influences versus reference-dependent reciprocity

In this study, I argued that the productivity gap between the Rise-1 and Rise-2 treatment groups observed in session 2 was driven by mood differences. A plausible alternative mechanism for such productivity gap, formalized by the theoretical model of Macera and te Velde (2018), is reference-dependent reciprocity. This mechanism relies on the premise that the Rise-1 and Rise-2 groups differed in their reference wages in session 2 (because they faced the rise announcement at different times) and, hence, in the intensity of reciprocity. However, as I explain next, the inclusion of the HighWage treatment group in the experiment allows me to rule out a significant influence of reference-dependent reciprocity on productivity.

The model of Macera and te Velde (2018) implies that workers reciprocate in response to a previous rise *only if* they did not have sufficient time to adapt to the new wage. The rationale is that because wage expectations adapt, "workers acclimatize to the new higher wage and effort reverts to baseline levels in the absence of further wage increases" (Macera and te Velde, 2018, p. 1). In the current setting, this hypothesis implies that productivity in session 2 should have been the same under the Rise-1 and HighWage treatments because the workers in the former group had sufficient time to adapt to the final wage, which coincided with the wage in the latter group. In contrast, in this study, I found that workers in the Rise-1 group were significantly more productive than those in the HighWage group even though they had learned about the pay rise a week earlier. This shows that reference-dependent reciprocity was not a primary driver in this study, strengthening the claim that the productivity gap between the Rise-1 and Rise-2 groups was driven by mood differences.[20]

5. Concluding remarks

In this study, I examined the productivity of student workers in Bogotá, who earned a fixed wage for data entry tasks in a one-off job. Subjects who received an unexpected 40% pay rise were more productive than those who were initially offered the same total final wage. This result replicated the key finding of some previous field studies (e.g., Gneezy & List, 2006, Bellemare & Shearer, 2009, Cohn et al., 2015, and Gilchrist et al., 2016) of a positive effect of an unconditional pay rise on productivity.[21] The distinctive contribution of the present study was to disentangle the influences of positive

[19] Cohn et al. (2014) conducted a field experiment in which workers were hired for a one-time sales promotion, were assigned to groups of two, and performed identical individual tasks. The individual workers received the same hourly wage during the first phase of the experiment. There were several treatment groups in the second phase of the experiment. In the baseline treatment, the hourly wage was not changed relative to the first phase. In the unilateral wage-cut treatment, only one group member's hourly wage was cut by 25%. The study found that workers in the unilateral wage-cut treatment whose wage was *not* cut but who witnessed their group member's pay being cut neither reduced nor increased their performance relative to the baseline treatment.

[20] Reference-dependent reciprocity might play a crucial role in a *repeated* game, as illustrated by the theoretical model of Macera and te Velde (2018). A previous pay rise announcement might induce the workers to expect additional rises *in the future*, which in turn would lead the workers to consider future pay rises as gifts only if they exceed expected rises. This would affect the extent of future reciprocation. As Macera and te Velde pointed out, testing this implication in future field experiments will help us to better understand the scope of gift exchange in longer-term regular jobs.

[21] More recently, DellaVigna et al. (2022) did not find any effect of an unconditional pay rise on productivity, but they found a sizeable positive impact of the pay rise on labor supply (as captured by subjects' willingness to work for extra time).

reciprocity and mood on gift exchange. While most of the increased effort immediately following the pay rise announcement was driven by reciprocity, the boost in mood also made a significant contribution, accounting for approximately 25% of the effect.

The results have significant implications for organizations. First, the finding that the announcement of an unconditional pay rise induced reciprocal behavior a week later, once the boost in mood had faded, suggests that a pay rise may increase worker productivity long after the rise is announced. Second, a shorter delay separating the pay rise announcement from the work shift will likely increase productivity further by capitalizing on the temporary boost in workers' mood.

Previous field tests of gift exchange at work, which were thoroughly reviewed by Esteves-Sorenson (2018) and DellaVigna et al. (2022), have yielded contradictory results. Some found increases in productivity, whereas others did not find any effect. Surprisingly, Esteves-Sorenson (2018) did not find any evidence of gift exchange in a productivity experiment similar to that of this study, wherein student workers entered data relating to academic articles in three 2-hour shifts using bibliographic software.

The discrepancy in these results is intriguing because Esteves-Sorenson implemented arguably stronger treatments, and several features of her study design enhanced the potential for gift exchange. First, she evaluated 50%, 67%, and 100% pay rises, rather than a 40% rise. Second, she explicitly framed the pay rise as a gift, while I did not.[22] Third, she paid the rise separately from the base wage at the beginning of each shift, whereas I paid all earnings in a single envelope at the end of the last shift. Fourth, although subjects in her experiment were told that the job was a one-off engagement, they did not remain anonymous; each subject negotiated with the research assistant regarding the time and place of the work and worked in isolation.

Clearly, research shedding more light on the reasons behind this conflicting evidence on gift exchange is needed. In addition to further exploring various aspects of pay, future research may systematically examine the reciprocity norm itself. While prominent anthropological and sociological theories view reciprocity as a universal social norm (Blau, 1964; Gouldner, 1960; Mauss, 1924), different populations may endorse it to varying degrees. A better understanding of such heterogeneity may help to reconcile some of the conflicting evidence on gift exchange.

Finally, an interesting question is whether the findings from this study regarding the roles of reciprocity and affect following a pay rise extend to situations in which workers suffer a wage cut. In two field experiments, Kube et al. (2013) and Cohn et al. (2014) found that wage cuts unrelated to performance had a significant detrimental effect on productivity, but they did not directly investigate the distinct roles of negative reciprocity and negative affect. (See also Lee & Rupp, 2007 for non-experimental evidence on workers' responses to wage cuts.) Because a wage cut represents a loss relative to the base wage, and losses generally loom larger than equal-sized gains (Kahneman & Tversky, 1979), the *affect-driven* response in terms of productivity to a wage cut is likely to be stronger than the response to an equivalent wage rise. In contrast, endorsement of negative reciprocity vis-à-vis positive reciprocity is less clear. Therefore, further empirical research is required to confirm the relative importance of negative reciprocity and negative affect following a wage cut.

Data availability

Data and Stata code for this study are available at https://osf.io/y3f2z.

[22] The subjects in her experiment were offered a wage rise "in an envelope embossed with the phrase 'A Gift for You' at the start of each shift" (Esteves-Sorenson, 2018, p. 4375).

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.socec.2022.101966.

References

Abeler, J., Altmann, S., Kube, S., & Wibral, M. (2010). Gift exchange and workers' fairness concerns: When equality is unfair. *Journal of the European Economic Association*, *8*(6), 1299–1324.

Akerlof, G. A. (1982). Labor contracts as partial gift exchange. *Quarterly Journal of Economics*, *97*(4), 543–569.

Akerlof, G. A., & Yellen, J. L. (1990). The fair wage-effort hypothesis and unemployment. *Quarterly Journal of Economics*, *105*(2), 255–283.

Ariely, D., Kamenica, E., & Prelec, D. (2008). Man's search for meaning: The case of Legos. *Journal of Economic Behaviour and Organization*, *67*, 671–677.

Bartlett, M. Y., & DeSteno, D. (2006). Gratitude and prosocial behavior: Helping when it costs you. *Psychological Science*, *17*(4), 319–325.

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

Blau, P. (1964). *Exchange and power in social life*. New York: Wiley.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment Manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49–59.

Carlson, M., Charlin, V., & Miller, N. (1988). Positive mood and helping behavior: A test of six hypotheses. *Journal of Personality and Social Psychology*, *55*, 211–229.

Charness, G., Cobo-Reyes, R., Lacomba, J. A., Lagos, F., & Sánchez, JMP. (2016). Social comparisons in wage delegation: Experimental evidence. *Experimental Economics*, *19*(2), 433–459.

Charness, G., & Kuhn, P. (2007). Does pay inequality affect worker effort? Experimental evidence. *Journal of Labor Economics*, *25*(4), 693–723.

Cialdini, R. B. (1993). *Influence: the psychology of persuasion*. New York: Harper Collins.

Cohn, A., Fehr, E., & Goette, L. (2015). Fair wages and effort provision: Combining evidence from a choice experiment and a field experiment. *Management Science*, *61*(8), 1777–1794.

Cohn, A., Fehr, E., Herrmann, B., & Schneider, F. (2014). Social comparison and effort provision: Evidence from a field experiment. *Journal of the European Economic Association*, *12*(4), 877–898.

DellaVigna, S., List, J. A., Malmendier, U., & Rao, G. (2022). Estimating social preferences and gift exchange with a piece-rate design. *American Economic Review*, *112*(3), 1038–1074.

Dufwenberg, M., & Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and Economic Behavior*, *47*(2), 268–298.

Esteves-Sorenson, C. (2018). Gift exchange in the workplace: Addressing the conflicting evidence with a careful test. *Management Science*, *64*(9), 4365–4388.

Falk, A., Becker, A., Dohmen, T., Huffman, D., & Sunde, U. (2016). *The preference survey module: a validated instrument for measuring risk, time, and social preferences*: *IZA discussion paper no. 9674*.

Fehr, E., Kirchsteiger, G., & Riedl, A. (1993). Does fairness prevent market clearing? An experimental investigation. *Quarterly Journal of Economics*, *108*(2), 437–459.

Gächter, S., & Thöni, C. (2010). Social comparison and performance: Experimental evidence on the fair wage-effort hypothesis. *Journal of Economic Behaviour and Organization*, *76*(3), 531–543.

Gagnon, N., & Noussair, C. N. (2020). Reciprocity under brief and long-time delays. *Economic Inquiry*, *58*(3), 1517–1530.

Gilchrist, D. S., Luca, M., & Malhotra, D. (2016). When 3+1>4: Gift structure and reciprocity in the field. *Management Science*, *62*(9), 2639–2650.

Gneezy, U., & List, J. A. (2006). Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments. *Econometrica*, *74*(5), 1365–1384.

Gouldner, A. W. (1960). The norm of reciprocity: A preliminary statement. *American Sociological Review*, *25*(2), 161–178.

Hennig-Schmidt, H., Sadrieh, A., & Rockenbach, B. (2010). In search of workers' real effort reciprocity–A field and a laboratory field experiment. *Journal of the European Economic Association*, *8*(4), 817–837.

Isen, A. M., Clark, M., & Schwartz, M. F. (1976). Duration of the effect of good mood on helping: Footprints on the sands of time. *Journal of Personality and Social Psychology*, *34*(3), 385–393.

Isen, A. M., & Levin, P. F. (1972). Effect of feeling good on helping: Cookies and kindness. *Journal of Personality and Social Psychology*, *21*(3), 384–388.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–292.

Kube, S., Maréchal, M. A., & Puppe, C. (2012). The currency of reciprocity: Gift exchange in the workplace. *American Economic Review*, *102*(4), 1644–1662.

Kube, S., Maréchal, M. A., & Puppe, C. (2013). Do wage cuts damage work morale? Evidence from a natural field experiment. *Journal of the European Economic Association*, *11*(4), 853–870.

Lee, D., & Rupp, N. G. (2007). Retracting a gift: How does employee effort respond to wage reductions? *Journal of Labor Economics*, *25*, 725–761.

Levin, P. F., & Isen, A. M. (1975). Further studies on the effect of feeling good on helping. *Sociometry*, *14*, 1–147.
Macera, R., & te Velde, VL. (2018). *On the power of surprising versus anticipated gifts in the workplace*: *Working paper*.
Malmendier, U., & Schmidt, K. M. (2017). You owe me. *American Economic Review*, *107*(2), 493–526.
Mas, A., & Moretti, E. (2009). Peers at work. *American Economic Review*, *99*(1), 112–145.
Mauss, M. (1924). *The gift: the form and reason for exchange in archaic societies. (translated by W.D. Halls)*. London: Routledge & Kegan Paul.
Pruitt, D. G. (1968). Reciprocity and credit building in a laboratory dyad. *Journal of Personality and Social Psychology*, *8*(2), 143–147.
Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review*, *83*(5), 1281–1302.
Sautua, S. I. (2022). Donation requests following a pay rise. *Journal of Economic Psychology*, *90*, Article 102518.
Smith, A. (1759–1809). *The theory of moral sentiments*. Glasgow: R Chapman.
Weyant, J. M. (1978). Effects of mood states, costs, and benefits on helping. *Journal of Personality and Social Psychology*, *36*(10), 1169–1176.
Wilke, H., & Lanzetta, J. T. (1970). The obligation to help: The effects of amount of prior help on subsequent helping behavior. *Journal of Experimental Social Psychology*, *6*, 488–493.