Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

# Fair Wages and Effort Provision: Combining Evidence from a Choice Experiment and a Field Experiment

Alain Cohn, Ernst Fehr, Lorenz Goette

To cite this article:
Alain Cohn, Ernst Fehr, Lorenz Goette (2015) Fair Wages and Effort Provision: Combining Evidence from a Choice Experiment and a Field Experiment. Management Science 61(8):1777-1794. http://dx.doi.org/10.1287/mnsc.2014.1970

MANAGEMENT SCIENCE
Vol. 61, No. 8, August 2015, pp. 1777–1794
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)

http://dx.doi.org/10.1287/mnsc.2014.1970


# Fair Wages and Effort Provision: Combining Evidence from a Choice Experiment and a Field Experiment

Alain Cohn, Ernst Fehr
Department of Economics, University of Zurich, CH-8006 Zurich, Switzerland
{alain.cohn@econ.uzh.ch, ernst.fehr@econ.uzh.ch}

Lorenz Goette
Faculty of Business and Economics, University of Lausanne, CH-1015 Lausanne, Switzerland,
lorenz.goette@unil.ch

The presence of workers who reciprocate higher wages with greater effort can have important consequences for firms and labor markets. Knowledge about the extent and determinants of reciprocal effort choices is, however, incomplete. We investigate the role of fairness perceptions and social preferences in a field experiment in which workers were hired for a one-time job. We show that workers who perceive being underpaid at the base wage increase their performance if the hourly wage increases, whereas those who feel adequately paid or overpaid at the base wage do not change their performance. Moreover, we find that only the workers who display reciprocity in a choice experiment show reciprocal effort responses in the field. The workers who lack reciprocity in the choice experiment do not respond to the wage increase, even if they feel underpaid at the base wage. Our findings suggest that fairness perceptions and social preferences are key in workers' performance response to wage increases. In our study, the wage increase affects effort mainly through the removal of perceived unfairness, i.e., the elimination of negative reciprocity toward the firm, rather than positive reciprocity. These results are the first direct evidence of the fair-wage effort hypothesis in the field and also help interpret previous contradictory findings in the literature.

Data, as supplemental material, are available at http://dx.doi.org/10.1287/mnsc.2014.1970.

*Keywords*: fairness perceptions; reciprocity; field experiment
*History*: Received March 26, 2013; accepted October 7, 2013, by Uri Gneezy, behavioral economics. Published online in *Articles in Advance* October 3, 2014.

## 1. Introduction

Throughout the history of economics, many prominent scholars pointed out that fairness concerns may influence labor-market outcomes (Hicks 1932, Marshall 1890, Slichter 1929).[1] The presence of fair-minded workers who reciprocate higher wages with greater effort may have important consequences for the functioning of labor markets. Concerns for fairness may give rise to noncompetitive wage premiums and involuntary unemployment (Akerlof 1982, Akerlof and Yellen 1990). Fairness concerns may also change firms' internal pay structure and wage setting over time (Cabrales et al. 2008), and they can influence the effects of labor-market policies such as minimum-wage legislations (Falk et al. 2006).

The nature of how wages affect effort is key to understanding important phenomena in the labor market, such as downward-wage rigidity as opposed to general rigidness of wages (Akerlof et al. 1996, Fehr and Goette 2005). In this paper, we directly test for the psychological mechanism underlying workers' response to a wage increase in a large-scale field experiment. There is thus far a lack of evidence of how fairness perceptions modulate the effort response to a wage increase. Understanding the psychological mechanism behind how wages affect effort has important consequences for theory. The simplest way of incorporating fairness concerns is to assume that workers generally consider higher wages to be fairer as in Akerlof (1984); this model predicts a general wage rigidity with respect to economic conditions. Akerlof and Yellen (1990) take a different approach based on intuitions derived from the social psychology and sociology literature. In their model, they assume that workers have a fair reference wage in mind. Wage increases up to the fair wage are assumed to induce workers to exert more effort because higher wages up to this wage limit are perceived to be fairer. In contrast, pay raises above the fair wage are assumed to have no impact on effort because workers already feel fairly treated. The key idea behind

[1] Slichter (1929) claimed, for example, that companies in the 1920s deliberately pursued a "high-wage" policy as they were convinced that paying workers a wage they consider to be fairer would increase productivity, reduce turnover, and prevent labor strikes (p. 431).

this model is that wage increases only affect workers' effort if they alleviate perceived underpayment. This asymmetry has important implications for the labor market. It predicts an asymmetry in how wages react to labor-market conditions. Although competitive upward pressures will always raise wages, firms may be reluctant to take advantage of deteriorating labor-market conditions because pushing the wage below what is perceived to be the fair wage may trigger effort reductions and thus offset the benefit of charging lower wages (Akerlof et al. 1996, Fehr and Goette 2005).

We conducted a controlled field experiment to examine how wage increases affect workers' fairness assessment of their pay and how this translates into changes in their effort. Our experiment took place in the context of a sales promotion. A publishing company hired workers for a one-time job. Their task was to distribute copies of a newly launched newspaper at train stations and other public places. We implemented two treatments in collaboration with the publishing company using a within-subjects design. All workers were recruited at an hourly base wage of CHF 22.[2] This is thus our baseline wage, and the subjects worked for this wage even before the experiment began. In a controlled way, we then increased workers' pay by CHF 5 per hour, to CHF 27. At different points in time, workers located in one sector received this wage increase, leaving workers in the other sector as the control group. A comparison of the productivity difference between the two treatments allows us to identify the average impact of the wage increase on workers' performance. We also conducted a follow-up survey two months after the experiment. In that survey, the workers also participated in a two-player choice experiment with monetary stakes, which allowed us to measure workers' reciprocal behavior. We can identify and interpret heterogeneous treatment responses by combining the additional data with the data from the field experiment.

We find that the wage increase has on average a positive and significant effect on workers' performance. However, the average effect hides individual heterogeneity in the extent to which workers reciprocated the wage increase. The wage increase raised the performance of workers who felt underpaid at the base wage. In contrast, adequately or overpaid workers did not respond to the wage increase. The effect is stable over time, and not driven by skill differences that may also determine fairness judgments. Our results imply that wage increases up to the level that workers consider to be fair lead to an increase in effort, whereas pay raises above this level have no impact on effort. In addition, we find that workers whom we identify as nonreciprocal individuals in the choice experiment also acted in a nonreciprocal manner in the field experiment; i.e., they did not respond to the wage increase, irrespective of whether or not they felt underpaid. In contrast, workers whom we identify as reciprocal types in the choice experiment provided higher effort in response to the wage increase if they felt underpaid at the base wage.

Our paper contributes to a recent empirical controversy. The evidence from laboratory experiments generally supports the notion that workers exert extra effort on average when they are paid a more generous wage (Brandts and Charness 2004, Charness 2000, Charness et al. 2004, Cooper and Kagel 2014, Fehr et al. 1993, Hannan et al. 2002). Results from recent field studies are more mixed (Bellemare and Shearer 2009; Gneezy and List 2006; Hennig-Schmidt et al. 2010; Hossain and List 2012; Kube et al. 2012, 2013; List 2006). These conflicting findings are sometimes taken as evidence that results from the laboratory do not generalize to natural environments, and that the psychological mechanisms identified in the lab may be irrelevant in the field (Camerer 2014, Levitt and List 2007, Stoop et al. 2012).

Our experiment further develops the methodology of labor-market field experiments in the realm of social preferences (Bandiera et al. 2005; Gneezy and List 2006; Kube et al. 2012, 2013). First, we measured workers' fairness perceptions during and after the experiment. This information is important because it helps us understand the relationship between wage changes and fairness perceptions and the link between fairness perceptions and effort choices. A common result in the studies mentioned above is that increases in hourly wages have a small and insignificant average effect on workers' performance. One reason for the lack of a significant effect in these studies may be that the experimenters paid a rather high *base* wage compared to the market wage for comparable jobs.[3] We know from our data that higher wages decrease the share of workers who feel underpaid. It is thus possible that only a relatively small number of workers felt underpaid at the base wage in the previous field

[2] CHF 22 is roughly the hourly wage that is paid in this area of Switzerland for simple tasks. For example, cleaning ladies in private households typically earn an hourly wage of CHF 20–25. Students who participate in laboratory experiments earn roughly CHF 25 per hour.

[3] In Kube et al. (2013) the base wage was EUR 15, whereas it amounted to EUR 12 in Kube et al. (2012). Students from the University of Karlsruhe (Germany) participated in the field experiment in both studies. The typical hourly wage for research assistants at this university is EUR 8, and the students earned on average EUR 10.5 per hour in previous employment relations. The hourly base wage in the library task of Gneezy and List (2006) was USD 12, whereas the typical market wage for similar activities was USD 10 (PayScale 2014). Thus, the base wage in these studies is already quite generous compared to typical alternative employment opportunities. It thus seems possible that only few workers felt underpaid at the base wage.

experiments and that a crucial condition for reciprocal effort choices may, therefore, have been absent.

Second, we measured workers' inclination toward reciprocal fairness by conducting a two-player, sequential-move game that we mailed to them. There is strong laboratory evidence suggesting that individuals' inclination toward reciprocity is heterogeneous—although many individuals exhibit such a behavioral regularity, a substantial share of subjects also acts rather selfishly and rarely (or never) returns a favor when doing so is costly (Bellemare et al. 2008, 2011; Charness and Rabin 2002; Fehr et al. 1993). Our "laboratory" measure of individuals' inclination toward reciprocity allows us to go beyond documenting an average effect of a wage increase and helps us understand the reasons behind heterogeneous effort responses.

Third, our study offers better precision due to a large sample. One reason for the lack of a significant effect in the other studies may be the limited statistical power to identify treatment effects due to the relatively small sample sizes with a maximum of 30 subjects per treatment. Although our average effect size is also modest, our larger sample of 196 workers allows us to reject the null hypothesis despite conservative calculations of standard errors that allow the residual to be correlated temporally as well as spatially. Another reason for our significant results may be that our measures for individuals' fairness perceptions, social preferences, and individual skill levels enabled us to control for individual heterogeneity. There is ample evidence of heterogeneity in individuals' preferences and skills (e.g., Bandiera et al. 2005, 2009, 2010; Bellemare et al. 2008; Fehr and Goette 2007; Mas and Moretti 2009), implying that the statistical power may be limited if one cannot control for these individual differences.

The remainder of this paper is organized as follows. Section 2 describes the design of our experiment and the setting in which it took place. Section 3 presents the empirical findings. Section 4 concludes the paper.

## 2. Experimental Setup

This section describes the institutional backdrop against which we conducted our field experiment. It also provides a first descriptive look at the data, performs randomization checks, and describes our empirical strategy.

### 2.1. Institutional Background

We conducted the field experiment in collaboration with a publishing company during the market launch of a novel daily newspaper. The publishing company commissioned a promotion agency to organize a sales promotion for its new product over a period of three months in the city of Zurich, Switzerland. Workers' task during the sales promotion was to distribute copies of the newspaper to passers-by at public places such as bus stops, train stations, and pedestrian areas. Workers could freely choose when to work, but they had to indicate their availability three to four weeks in advance, and once they had signed up for particular shifts they were not allowed to switch or cancel their chosen shifts. Workers had to sign up for blocks of three hours (4 P.M. to 7 P.M.) from Monday to Friday. Before the sales promotion began, all workers had agreed to work for CHF 22 per hour.

The workers were managed by team leaders with whom they met at the beginning and end of the shifts at the assigned locations. Because of the size of the sales promotion, the agency divided the city into two equally sized sectors (I and II) that covered about the same number of locations. Team leaders and workers were sometimes assigned to work in different locations but they always stayed within the same sector. During the shifts, team leaders visited the locations to ensure that the workers did not run out of newspaper copies. They also had to check in an unobtrusive way that workers would not throw copies away. No such incident was observed. Another important task of the team leaders was to record the number of copies each worker distributed during a shift. Team leaders knew how many copies each worker had received at the beginning and if necessary also during a shift and they counted the remaining copies when a shift ended. Neither workers nor team leaders knew that they were part of an experiment.[4]

We deliberately choose a work environment where explicit performance incentives were absent and where reputational or repeated game incentives were unlikely to play a role to cleanly identify whether workers voluntarily reciprocate higher wages with greater effort. Workers earned a performance-independent hourly wage. They also had no prospect for future employment at the publishing company.

### 2.2. Randomized Wage Increase

We implemented two treatments. In the baseline treatment, denoted by "CHF22," workers received their regular hourly wage of CHF 22. In the main treatment, denoted by "CHF27," the publishing company raised workers' hourly wage by CHF 5.[5] Treatments were communicated to the workers shortly before the beginning of a shift with both a postcard and a text message. We used two communication channels to ensure that every worker was informed about the relevant wage. Text messages were sent a few minutes

[4] Because subjects were in their natural environment and they did not know that they were part of an experiment, our study is a natural field experiment according to the classification scheme outlined in Harrison and List (2004) and Al-Ubaydli and List (2014).

[5] To maintain a certain wage hierarchy, team leaders received a pay raise of CHF 5 per hour for the entire duration of the experiment, i.e., the raise was independent of the respective wage level of their workers. The workers were not informed about the team leaders' pay level.

prior to a shift and postcards were handed out by the team leaders at the start of the shifts. The message in treatment CHF22 reminded the workers to retain the straps from the distributed newspaper bundles. The request to keep the straps allowed team leaders to collect the numbers of distributed copies more quickly. In treatment CHF27 the message included the straps reminder and additionally informed the workers about the pay raise. Importantly, the message made it clear that the publishing company, and not the promotion agency, was paying the workers the higher wage: the messages concluded with the name of the publishing company and the postcards bore its logo. Because most workers worked on an irregular basis (e.g., on Monday in one week but on Wednesday and Thursday in another week) we communicated the relevant wage daily, shortly before the beginning of a shift.

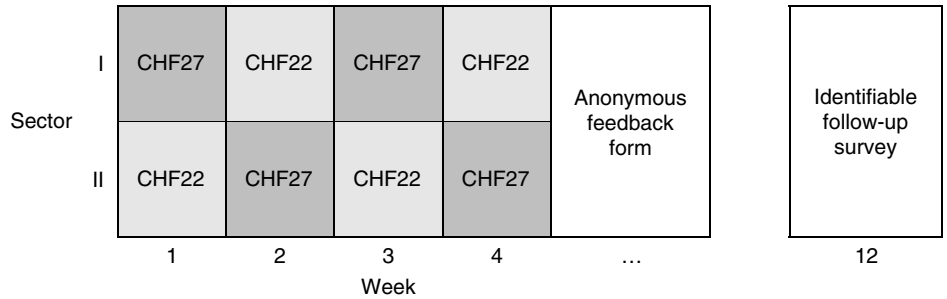
The field experiment took place over a four-week period toward the end of the sales promotion. We exploited the spatial division of the city into sectors between which the workers could not move to randomize the two treatments across the two sectors. We chose to randomize the wages on a sector-week level (see Figure 1). Thus, in each of the four experimental weeks the workers in one sector received the higher wage (treatment CHF27), whereas those in the other sector served as a control and were paid the base wage (treatment CHF22). The weekly rotation of the treatments served two purposes. First, compared to a less-frequent rotation, a weekly rotation allows for a more robust identification of confounding time effects, which could have been large if the newspaper had become better known over time. Second, we expected a strong turnover after the second week of the experiment because summer vacation in the schools was beginning at this time. A weekly rotation therefore also helped generate within-subjects variations in pay, enabling us to control for individual–specific factors that affect workers' performance. Although the majority of the workers experienced both wages, it is unlikely that the workers were able to anticipate the wages they were paid during the experimental period. First, communication between workers from different sectors was unlikely. Second, the weekly rotation of the wages during the experimental period was repeated only once and constituted a small part of the entire sales promotion. Finally, workers had highly irregular workdays making it implausible that they could see through the wage pattern of the experimental period.

Toward the end of the sales promotion, the promotion agency sent a feedback form to the workers, asking them to state their opinion about the working conditions. One question of particular interest to us asked the workers to rate the fairness of the wages: "I consider the regular (higher) hourly wage of CHF 22 (CHF 27) for doing this job to be [1 = very unfair, 2 = moderately unfair, 3 = neither . . . nor, 4 = moderately fair, 5 = very fair]." This information serves us as a manipulation check, as it allows us to examine the effectiveness of our fairness manipulation. To get truthful answers, workers were asked to fill in the form anonymously.

### 2.3. Follow-Up Survey and Choice Experiment

Two months after the conclusion of the field experiment, we conducted a follow-up survey among the workers. The survey asked a variety of questions related to temporary work (see Online Appendix A; online appendices available at https://sites.google.com/site/alaincohn/home/research). Participants were paid CHF 7 for completion. They first had to indicate up to three employers from the previous six months (which covered the time period of the sales promotion), and then answer several questions relating to each of their listed employers. The questions of key interest to us asked the participants to state the wage they were effectively paid and the wage they considered to be fair for their work: "What hourly wage did you earn at employer *X*?" and "What hourly wage would you find appropriate for doing this job at employer *X*?"

**Figure 1** **Timing of Events**




*Notes.* This figure visualizes the two phases of the study. The first phase consists of the randomized wage increase that took place during the four experimental weeks of the sales promotion and an anonymous feedback form that was administered toward the end of the experimental period. We randomized wages on a sector-week level. Thus, for a given week the workers in one sector received the higher wage (treatment CHF27), whereas those in the other sector served as a control and were paid the base wage (treatment CHF22). In a second phase, we invited the workers to participate in a survey on part-time work that allowed us to measure both their perceived underpayment at the base wage and their preference for reciprocal fairness. The survey respondents were identifiable by a code, which enabled us to link their answers to their individual performance data.

By subtracting the answers of the first question from the second, we are able to construct an individual measure of perceived underpayment. The workers had no reason to believe that the survey was in any way connected to the field experiment conducted two months earlier. They did not know that they had been part of an experiment and the survey was administered by researchers whom they perceived to be completely unrelated to the publishing company.[6]

At the end of the survey we measured workers' inclination toward reciprocal fairness by inviting them to participate in a simplified version of the moonlighting game played for real money (Abbink et al. 2000). This two-player sequential-move game has the important feature that both kind and unkind actions are feasible for both players. In our version of the game, the first movers could divide CHF 24 in three different ways: they could choose between (i) an *unfair* allocation (CHF 18 for the first mover and CHF 6 for the second mover), (ii) an *equitable* allocation (CHF 12 for both players), or (iii) a *generous* allocation (CHF 6 for the first mover and CHF 18 for the second mover). The second movers could then reward or punish the first movers by assigning up to two positive or negative points, respectively; they could also decide not to assign any points at all. The reward and punishment technology was designed in a way such that one positive (negative) point cost the second mover CHF 2 and increased (decreased) the first mover's payoff by CHF 6. The game was played one-shot; second-mover behavior therefore captures reciprocal fairness without repeated game effects. For the purposes of this study, we assigned all workers to the role of the second mover and elicited their decisions using the strategy method.[7] We classify the workers as reciprocal if they returned more positive points in the generous allocation than in the equitable allocation, or more positive points in the equitable allocation than in the unfair allocation. Thus, to be counted as a reciprocal type, the assignment of positive points needs to be a (weakly) increasing function of the second mover's payoff, with a strictly positive slope in at least one of the two comparisons (unfair vs. equitable, equitable vs. generous).[8] This classification does not distinguish between distributional motives (e.g., Fehr and Schmidt 1999) and intention-based reciprocity (e.g., Falk and Fischbacher 2006, Rabin 1993). The purpose of this measure is to provide us with a way of testing whether individuals with a tendency toward reciprocal fairness in the choice experiment responded differently to the wage increase we implemented, and thus needed to be parsimonious. Evidence from laboratory experiments suggests that intention-based reciprocity is probably more important in explaining variation across individuals in such experiments (Charness and Levine 2007, Falk et al. 2003), though we cannot rule out that distributional preferences play a role too.[9] We also experimented with trying to distinguish between positive and negative reciprocity in the spirit of Dohmen et al. (2009), but in our game, positive and negative reciprocity are highly correlated and do not allow us to distinguish them from each other.

### 2.4. Descriptive Statistics and Randomization Checks

Table 1 presents the summary statistics of the data used in this paper. We observed 196 individuals during the four experimental weeks of the sales promotion. More than 70% of the workers were female. Workers were between 16 and 42 years old, with an average age of 22.5 years. Most of them were Swiss citizens. On average, they completed 6.5 shifts during the experimental period and distributed 228 newspaper copies per hour. Table 1 also shows that the experimental conditions were well balanced within individuals. Workers' median exposure to the higher wage was 50% of the shifts; the 25th percentile of the distribution was 42%, and the 75th percentile was 66%.

Turning to the fairness perception of wages, we asked in the follow-up survey what wage workers would consider appropriate for this type of job. The average reply was CHF 1 more than the CHF 22 paid in the baseline treatment. This judgment, however, is not uniform, as can be seen in the histogram in Figure 3 (see §3.1). It shows the distribution of the difference between what workers considered to be the appropriate wage and the base wage of CHF 22. Slightly less than half of the workers considered the base wage as the appropriate wage. The majority of the workers (53%) perceived themselves to be underpaid at the base wage of CHF 22, with a sizable group of workers (20%) perceived CHF 25 to be the appropriate wage.

[6] The survey respondents were also informed that their individual data would be treated confidentially and never revealed to their employer. To match the survey data with individual performance data, we put the same code on the survey that we received from the promotion agency to identify the individual performance data.

[7] The choices of the first movers in the moonlighting game were collected through a survey mailed to a random sample of students at the University of Zurich around the same time.

[8] We used the strategy method because it provided us with well-defined behavioral types. The downside of the strategy method is that punishment levels might have been less pronounced than with the direct-response method (Brandts and Charness 2011). As a consequence, we might have misclassified weakly reciprocal individuals.

[9] Because distributional relationships are rather unclear in the field, distributional preferences may be even less important in the field than in the laboratory. For example, it is likely that the workers did not have a precise idea of how wealthy the publishing company was and how valuable their effort was for the firm. We thank an anonymous referee for pointing this out to us.

**Table 1 Descriptive Statistics**

| Variable | Mean | SD | *N* |
|---|---|---|---|
| *Hourly copies distributed* | 227.760 | 83.995 | 1,269 |
| *Number of shifts* | 6.474 | 4.248 | 196 |
| *Fraction of shifts worked in CHF27* | 0.526 | 0.234 | 196 |
| *Fairness rating of CHF22* | 2.858 | 0.844 | 113 |
| *Fairness rating of CHF27* | 3.894 | 0.910 | 113 |
| *Perceived underpayment* (in CHF) | 1.097 | 2.056 | 119 |
| *Identification with the firm* | −0.235 | 0.963 | 119 |
| *Age* (in years) | 22.465 | 4.698 | 114 |
| *Male* | 0.272 | 0.447 | 114 |
| *Foreigner* | 0.132 | 0.340 | 114 |
| *Number of siblings* | 1.439 | 0.912 | 114 |
| *Secondary school* | 0.614 | 0.489 | 114 |
| *Apprenticeship/vocational school* | 0.281 | 0.451 | 114 |
| *Additional, further education* | 0.211 | 0.409 | 114 |
| *Baccalaureate* | 0.675 | 0.470 | 114 |
| *Technical school* | 0.237 | 0.427 | 114 |
| *University* | 0.228 | 0.421 | 114 |
| *Points returned if 1st mover proposed* (18, 6) | −0.559 | 1.050 | 118 |
| *Points returned if 1st mover proposed* (12, 12) | 0.254 | 0.730 | 118 |
| *Points returned if 1st mover proposed* (6, 18) | 0.890 | 0.885 | 118 |
| Reciprocal individuals: | | | |
| *Points returned if 1st mover proposed* (18, 6) | −0.922 | 0.807 | 77 |
| *Points returned if 1st mover proposed* (12, 12) | 0.273 | 0.599 | 77 |
| *Points returned if 1st mover proposed* (6, 18) | 1.247 | 0.632 | 77 |
| Nonreciprocal individuals: | | | |
| *Points returned if 1st mover proposed* (18, 6) | 0.122 | 1.122 | 41 |
| *Points returned if 1st mover proposed* (12, 12) | 0.220 | 0.936 | 41 |
| *Points returned if 1st mover proposed* (6, 18) | 0.220 | 0.909 | 41 |

*Notes.* This table describes the data used in this paper. The data come from three different sources: the sales promotion, the feedback form, and the follow-up survey that includes the moonlighting game. During the four experimental weeks, 196 workers were observed; 113 of them returned the feedback form and 119 of them completed the follow-up survey. In the follow-up survey, five workers did not answer the questions on personal characteristics and one worker did not participate in the moonlighting game.

**Table 2 Randomization Check for Worker Characteristics**

| | Treatment | | | | |
|---|---|---|---|---|---|
| | CHF22 | | CHF27 | | |
| Variable | Mean | SD | Mean | SD | *p*-value |
| *Perceived underpayment* (in CHF) | 1.088 | 2.099 | 1.081 | 2.142 | 0.69 |
| *Age* (in years) | 23.370 | 5.257 | 23.344 | 5.397 | 0.77 |
| *Male* | 0.281 | 0.450 | 0.267 | 0.443 | 0.68 |
| *Foreigner* | 0.161 | 0.368 | 0.172 | 0.378 | 0.70 |
| *Number of siblings* | 1.376 | 0.849 | 1.367 | 0.854 | 0.91 |
| *Secondary school* | 0.648 | 0.478 | 0.633 | 0.483 | 0.69 |
| *Apprenticeship/vocational school* | 0.331 | 0.471 | 0.308 | 0.462 | 0.52 |
| *Additional, further education* | 0.248 | 0.432 | 0.242 | 0.429 | 0.85 |
| *Baccalaureate* | 0.618 | 0.487 | 0.658 | 0.475 | 0.27 |
| *Technical school* | 0.251 | 0.434 | 0.211 | 0.409 | 0.22 |
| *University* | 0.245 | 0.431 | 0.211 | 0.409 | 0.29 |
| *Points returned if 1st mover proposed* (18, 6) | −0.651 | 1.017 | −0.663 | 1.004 | 0.98 |
| *Points returned if 1st mover proposed* (12, 12) | 0.251 | 0.726 | 0.248 | 0.684 | 0.88 |
| *Points returned if 1st mover proposed* (6, 18) | 0.811 | 0.902 | 0.857 | 0.891 | 0.46 |

*Notes.* Because of the within-subject design, most workers participated in both treatments. It was possible, however, that some workers worked $x$ times (shifts) in treatment CHF22 and $y$ times (shifts) in treatment CHF27. Therefore, our randomization check takes this into account, i.e., the characteristics of this worker count $x$ times for treatment CHF22 and $y$ times for treatment CHF27. This is a very conservative randomization check because showing insignificant differences in worker characteristics across treatments would be much easier if we were to count each worker only once. The first four columns in this table show the treatment averages and standard deviations of worker characteristics. The last column contains the $p$-values ($X^2$ tests for binary variables and Mann–Whitney tests for nonbinary variables) for the null hypothesis of perfect randomization.

The last panel in Table 1 displays the average number of points returned in the moonlighting game. In response to the generous allocation of "CHF 6 for the first mover, CHF 18 for the second mover," the workers spent an average of 0.89 positive points to increase the first mover's payoff. They spent on average 0.25 positive points if the first mover chose the equitable allocation (CHF 12, CHF 12). If the first mover chose the unfair allocation (CHF 18, CHF 6), the workers spent 0.56 negative points to reduce the first mover's payoff. Table 1 also displays the results separately for the reciprocal ($N = 77$) and the nonreciprocal ($N = 41$) participants. There are large quantitative differences between these two groups. This is because many of the nonreciprocal individuals acted strictly selfish, never spending any points. Others classified as nonreciprocal displayed a concern for efficiency and always transfered the same number of positive points.

We present various checks that verify that we successfully randomized subjects to treatment conditions. As explained above, the assignment of workers to shifts and locations was organized in a way that made it nearly impossible for them to selectively pick shifts in treatment CHF27. Nevertheless, we need to verify whether worker characteristics in the two treatments were identical, because imbalances in these characteristics could also create treatment differences in effort. It is therefore important to know whether worker characteristics are balanced across treatments. Table 2 shows that this is the case. Worker characteristics in the two treatments are statistically identical for every dimension we measured (e.g., underpayment judgments, age, gender, etc.).[10]

Most importantly, we need to establish that there was no selectivity with respect to the number of shifts worked as a function of the treatment, as this would complicate our empirical analysis below. As we explained earlier, the rule was that the workers had to sign up for shifts well in advance. This made it difficult for workers to select into treatments. The results in Table 3 confirm that there was no selectivity. The table reports the results from a regression where the number of shifts a worker completed in each treatment is the dependent variable. The results show that the workers worked on average the same number of shifts in the

[10] All $p$-values reported in this paper are two sided.

**Table 3 Participation at the Individual Level During the Experiment**

| | (1) | (2) |
|---|---|---|
| Dependent variable: | *Number of shifts per treatment* | |
| *CHF27* | 0.189 | 0.189 |
| | (0.130) | (0.184) |
| *Intercept* | 3.143*** | 3.143*** |
| | (0.162) | (0.092) |
| Fixed effects | | |
| Worker | No | Yes |
| $R^2$ | 0.002 | 0.846 |
| *N* | 392 | 392 |

*Notes.* OLS estimates. Standard errors in parentheses are clustered on the individual level. The unit of observation is a worker in each treatment. The dependent variable is the number of shifts per treatment and *CHF27* is an indicator variable for treatment status.
***$p < 0.01$.

**Table 4 Participation at the Sector-Day Level During the Experiment**

| | (1) | (2) |
|---|---|---|
| Dependent variable: | *Fraction of unfilled shifts per day* | |
| *CHF27* | −0.003 | −0.003 |
| | (0.031) | (0.032) |
| *Intercept* | 0.226*** | 0.228*** |
| | (0.016) | (0.025) |
| Fixed effects | | |
| Sector | No | Yes |
| Day | Yes | Yes |
| $R^2$ | 0.532 | 0.533 |
| *N* | 40 | 40 |

*Notes.* OLS estimates. Standard errors in parentheses are clustered on the day level. The unit of observation is a day in each treatment. The dependent variable is the fraction of unfilled shifts per day and *CHF27* is an indicator variable for treatment status.
***$p < 0.01$.

two treatments. They worked 3.1 shifts in treatment CHF22; the dummy for treatment CHF27 is small and far from significant: it is equal to 0.19 and has a standard error of 0.13, as shown in column (1). Thus, we would have been able to detect even small differences of 0.26 shifts across groups, which is less than 5% of the 6.5 shifts worked altogether. We consider this a "tightly estimated zero": the number of shifts worked in the two treatments is statistically indistinguishable. This result also holds if we control for individual fixed effects (see column (2) in Table 3).

In Table 4 we check the randomization to treatments from a different angle. The promotion agency had difficulties filling the available shifts; shifts remained vacant in roughly 20% of cases. It would have been easier for the company to fill the high-wage shifts if workers had self-selected into treatment CHF27. Table 4 shows, however, that this was not the case. In this table the number of unfilled shifts in each sector for each day of the study is the dependent variable. The results show that the coefficient of the *CHF27* dummy is close to zero and insignificant, indicating that treatment CHF27 had no effect on the share of unfilled shifts. Thus, there was no selectivity into the number of shift worked.

We also examine whether the experimental conditions affected our underpayment measure and workers' willingness to reciprocate in the moonlighting game. It is important to check whether a higher exposure to one of the treatments influenced our underpayment measure (i.e., the difference between the wage deemed to be appropriate and the base wage) and our measure of reciprocity. If they did, we could not use these measures as independent variables in our regressions described below. Column (1) in Table 5 shows that the fraction of shifts during which a worker is exposed to treatment CHF27 has no effect on our underpayment variable.[11] Thus, the small differences in exposure to treatment CHF27 did not affect subsequent fairness judgments of wages. Columns (2)–(4) in Table 5 examine whether the exposure to treatment CHF27 affected workers' behavior in the moonlighting game for the three possible transfers of the first mover. Since the choices in the three cases are probably highly correlated, this needs to be taken into account when calculating the standard errors. We therefore estimate the three equations as seemingly unrelated regressions, thus allowing arbitrary correlation in the residual across the three choices. Taken individually, the results in each of the columns show no significant influence of treatment CHF27 on workers' second-mover behavior. We also perform a joint test of the hypothesis that the exposure to treatment CHF27 did not affect points returned in any of the three cases by estimating all three regressions using a seemingly unrelated regression model. The *p*-value of the hypothesis of no effect in all three equations is 0.58. We thus conclude that variations in the exposure to the higher wage neither influenced our underpayment measure nor workers' behavior in the choice experiment. This, therefore, allows us to use these measures as independent variables in our empirical analysis.

### 2.5. Empirical Strategy

Approaching and addressing passers-by requires effort. Therefore, the number of copies distributed is our measure of individual effort. Since we have shown

[11] We ran an alternative specification where we additionally include a dummy variable for the workers who worked only under the high wage during the experimental period of the sales promotion, and another dummy variable for those who experienced only the base wage. Roughly 10% of the workers who participated in the follow-up survey experienced only the high wage, and about 5% only the base wage. We find that higher exposure to treatment CHF27 during the experiment is again not related to fairness perceptions that were measured two months later. We thank an anonymous referee for suggesting this possibility.

**Table 5 Randomization Check for Outcomes Measured After the Field Experiment**

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Dependent variable: | *Perceived underpayment* | *Points returned if 1st mover proposed* (18, 6) | (12, 12) | (6, 18) |
| *Fraction of shifts in CHF27* | 0.190 (0.684) | 0.008 (0.431) | 0.180 (0.299) | 0.497 (0.360) |
| *Intercept* | 0.992** (0.431) | −0.564** (0.256) | 0.155 (0.178) | 0.617*** (0.214) |
| $R^2$ | 0.000 | 0.000 | 0.003 | 0.016 |
| *N* | 119 | 118 | 118 | 118 |

*Notes.* Column (1) reports OLS estimates with robust standard errors in parentheses, and columns (2)–(4) report the estimates of seemingly unrelated regressions. Throughout all columns, the independent variable is workers' exposure to treatment CHF27 indicated as the fraction of shifts they worked under the higher wage. In column (1) the dependent variable is the perceived underpayment at the base wage, and in columns (2)–(4) the dependent variable is the number of points returned by the workers in the moonlighting game.
$^{**}p < 0.05$; $^{***}p < 0.01$.

in Tables 3 and 4 that our treatment did not affect the number of shifts worked, this implies that there is no selectivity into work with respect to the treatment. We can focus directly on labor supply conditional on having chosen to work. We first estimate the average treatment effect of the wage increase on workers' performance using the following regression model:

$$\log(y_{ikt}) = \beta_0 + \beta_1 I(CHF27)_{kt} + \lambda_k + \delta_t + \epsilon_{ikt}, \quad (1)$$

where the dependent variable is the log of the number of hourly copies distributed by worker $i$ at a specific location $k$ on day $t$. Our key exogenous variable is the wage condition in which a worker was working, which is represented by the treatment indicator *CHF27* for the high-wage condition. Its coefficient $\beta_1$ (multiplied by 100) can thus be interpreted as the percentage change in workers' performance that is caused by the wage increase. We include an exhaustive set of fixed effects $\lambda_k$ as controls for the location at which a worker $i$ worked on day $t$. We also include day fixed effects $\delta_t$ for each day of the experiment.[12]

Our experimental wage increase guarantees that the treatment indicator is uncorrelated with the residual $\epsilon_{ikt}$ and that the ordinary least squares (OLS) estimator will satisfy the conditional-mean independence needed for convergence to the true parameter. It does not, however, imply that the residual is uncorrelated across observations, thus raising potential issues with the standard errors of the estimated coefficients. Standard OLS procedures require that the residual is independent between any two observations. Two potential sources of correlation for the residual in our application are serial correlation within an individual, and correlations among the residuals of spatially close observations on the same day. Temporary, but auto-correlated shocks to an individual's motivation may induce serial correlation in an individual's performance on the job and thus render the assumption of independence of the residuals within an individual invalid. The residuals may also be spatially correlated because one could imagine that if a swarm of commuters heads in one direction, this affects the pick-up rate of the newspaper of spatially close workers in a similar way. This introduces a spatial correlation into the residuals of geographically close workers (Conley 2008). In this case, because the residual is uncorrelated with the treatment, the treatment effect can still be estimated consistently using OLS. However, the standard errors provided by standard OLS routines will be wrong, and, potentially, biased downward (Angrist and Pischke 2009). We use Hsiang's (2010) nonparametric correction for spatial and temporal correlation to correct for the two potential problems and to obtain the correct standard errors without imposing further restrictions on the covariance matrix. This procedure is very similar to multiway clustering in that it allows for any form of spatial and serial correlation, but uses a kernel estimate of the covariance matrix in which spatially close observations can be defined. We chose a radius of 3 kilometers in our application, and place no constraints on the number of lags for the individual.[13]

We also estimate a variant of Equation (1) with individual fixed effects $\nu_i$ included:

$$\log(y_{ikt}) = \beta_0 + \beta_1 I(CHF27)_{kt} + \nu_i + \lambda_k + \delta_t + \epsilon_{ikt}. \quad (2)$$

Including the individual fixed effects is not strictly necessary, because the treatments are randomized. However, it may be interesting for a different purpose: our experiment creates between-subject as well as within-subject variation in wages, and it is not clear whether within-subject manipulations make the treatment effect stronger or weaker (Charness et al. 2012), though one could certainly suspect so. The OLS

[12] We also estimated a model where we included team-leader fixed effects. The estimation results remain the same.

[13] An omnibus way to fix this problem would be to cluster observations along two dimensions: individuals and week × sector. However, this correction is valid asymptotically, as the number of week $x$ sector cluster gets large. We only have eight-week × sector cells and thus this is unfeasible. (We are grateful to an anonymous referee for pointing this out.) Our correction allows arbitrary correlations within an individual, and spatially correlated residuals across different individuals on the same day. The only difference between our correction and the one suggested here is only that ours does not allow the residual of individual $i$ on day $t$ to be correlated with the residual of another individual on a different day in that week. It is difficult to imagine how one newspaper distributed today by one worker could have an effect on the number of newspapers another worker would distribute in the future. Thus, we are comfortable restricting ourselves to the slightly more restrictive formulation we have chosen here.

estimate from Equation (1) used between-subject as well as within-subject variation in the wage to identify the treatment effect in approximately equal proportions. By also estimating Equation (2), the treatment effect in this column is identified entirely within subject. This allows us to asses if between subject and within subject create different treatment effects.

In a second step of the analysis, we investigate individual heterogeneity in the treatment effect. Specifically, we examine how workers' perceived fairness of the base wage affected their performance response to the wage increase. In keeping with our specifications above, we estimate the following regression model:

$$\log(y_{ikt}) = \beta_0 + \beta_1 I(CHF27)_{kt} + \beta_2 I(CHF27)_{kt} \times \Delta_i + \beta_3 \Delta_i + \lambda_k + \delta_t + \epsilon_{ikt}, \quad (3)$$

where $\Delta_i$ measures worker $i$'s perceived underpayment, i.e., the difference between what worker $i$ considered to be a fair wage for this job and the wage he was paid in the baseline treatment. The estimate of the coefficient $\beta_3$ should be interpreted with caution: $\Delta_i$ is not randomly assigned and may well be correlated with the error term, and therefore bias the estimate of $\beta_3$. For example, if unproductive individuals also feel entitled to a higher wage, there would be a downward bias in the estimate of $\beta_3$. However, $\beta_1$ (multiplied by 100) can be causally interpreted as the percentage change in workers' performance that is triggered by the wage increase if the worker feels adequately paid in the baseline treatment, i.e., if $\Delta_i = 0$. Because the wage increase is randomized and the regression controls for $\Delta_i$, the treatment variable is uncorrelated with the regression residual. Similarly, the coefficient of the interaction term $\beta_2$ indicates the extent to which underpaid workers responded to the wage increase and can also be given a causal interpretation for the same reason. Intuitively, this is as if one estimated the impact of the regression on the different subgroups of $\Delta_i$: because the treatment is randomized with respect to $\Delta_i$, these are the true treatment effects in each of the groups.[14] We also estimate the same equation including individual fixed effects rather than a control for $\Delta_i$:

$$\log(y_{ikt}) = \beta_0 + \beta_1 I(CHF27)_{kt} + \beta_2 I(CHF27)_{kt} \times \Delta_i + \nu_i + \lambda_k + \delta_t + \epsilon_{ikt}. \quad (4)$$

[14] To see this more formally, assume that the residual in Equation (3) is correlated with $\Delta_i$ and that, for the sake of the argument, $\epsilon_{it} = \eta_{it} + \gamma(\Delta_i - \Delta)$, but that the standard assumption are satisfied with respect to $\eta_{it}$, i.e., $E(\eta_{it} \mid T_{it}, \Delta_i) = 0$. The conditional mean function in that case is given by $E(y_{it} \mid T_{it}, \Delta_i) = \beta_0 - \gamma\Delta + \beta_1 T_{it} + \beta_2 T_{it}\Delta_i + (\beta_3 + \gamma)\Delta_i$. Thus, the regression equation $y_{it} = \beta_0 - \gamma\Delta + \beta_1 T_{it} + \beta_2 T_{it}\Delta_i + (\beta_3 + \gamma) \cdot \Delta_i + \eta_{it}$ defines a valid regression that satisfies all assumptions with respect to the residual $\eta_{it}$. Therefore, OLS estimates of Equation (3) will converge to the parameters shown in the previous equation: the estimates of the treatment effects including the interaction with $\Delta_i$ will be consistent, but the estimates of the impact of $\Delta_i$ itself and the constant should not be given a structural interpretation.

This specification allows for fully flexible control for possibly nonlinear effects of $\Delta_i$ on performance.

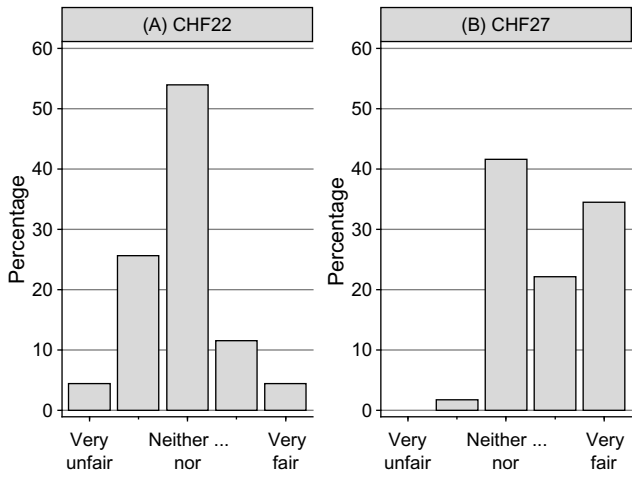
## 3. Results

The literature on social preferences suggests that the extent to which higher wages lead to higher effort depends on whether and how higher wages are associated with an increase in the perceived fairness of the wage. We therefore begin the exposition of our results by presenting the effect of the wage increase on workers' perceived fairness of their wage. We then turn to the average treatment effect of the wage increase on workers' performance and assess the robustness of this result. Finally, we probe deeper into the psychological mechanisms that underlie the observed pattern in the data.

### 3.1. Impact of the Wage Increase on Workers' Perceived Fairness of the Wage

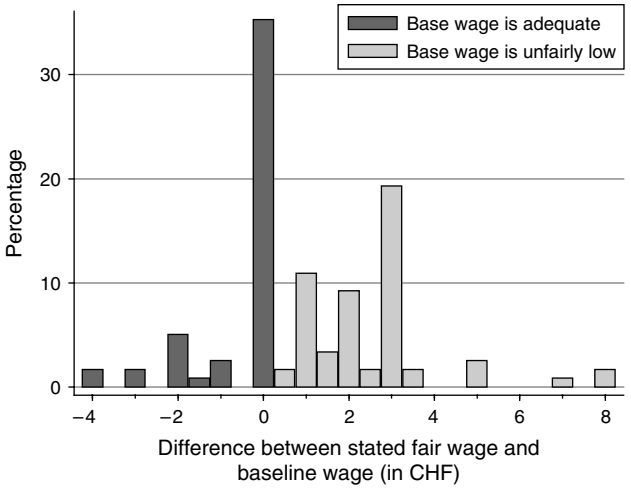
Data from the feedback form collected toward the end of the field experiment point to a strong impact of the wage increase on workers' perceived fairness of their pay. Panel (A) in Figure 2 shows that roughly 30% of the workers perceived the base wage as unfairly low, whereas panel (B) in this figure indicates that less than 2% reported the same for the higher wage. Moreover, comparing the two panels reveals that there is a clear shift to a fairer assessment of pay in treatment CHF27 ($p < 0.01$, Wilcoxon signed-rank test). However, there is strong heterogeneity in the fairness perception of the base wage. The strongest shift in the distribution of fairness assessments comes from the bottom end of the distribution, meaning that the higher wage had the largest impact on those workers who considered the

**Figure 2 Workers' Fairness Assessments of the Base Wage and Higher Wage**



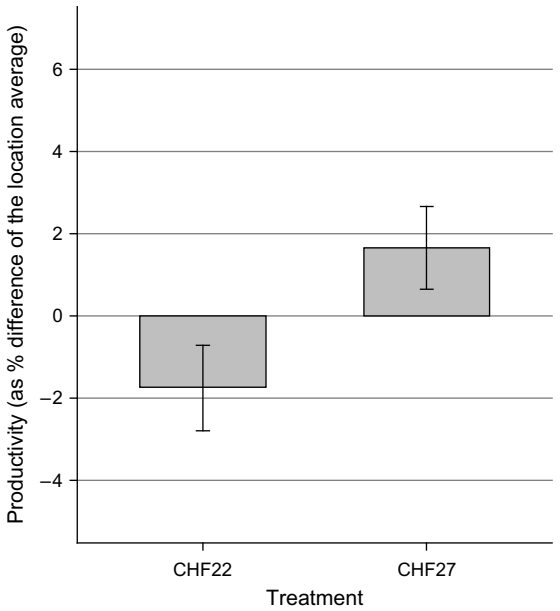

*Note.* This figure shows workers' fairness ratings of the base wage (CHF22) and the higher wage (CHF27) on a five-point scale using data from the anonymous feedback form.

**Figure 3 Workers' Perceived Underpayment at the Base Wage**




*Notes.* This figure shows a histogram of the difference between what the workers considered to be the appropriate wage and the base wage of CHF 22 per hour. A positive number indicates that a worker felt underpaid at the base wage, whereas a negative number indicates that the worker perceived to be overpaid for this job. Bins are in steps of CHF 0.5.

**Figure 4 Average Treatment Effect of the Wage Increase on Workers' Performance**




*Notes.* This figure shows workers' average performance and standard error of the mean in each treatment. Performance is measured as the hourly copies distributed normalized by the average hourly copies distributed at the location.

base wage to be unfairly low. There is also, however, a large group of workers for whom the higher wage could not improve their perceived fairness of the wage; those workers already perceived the base wage as rather fair and considered the higher wage to be equally fair.

As discussed earlier, we find similar individual heterogeneity in fairness perceptions in the follow-up survey conducted two months after the field experiment. Figure 3 shows that more than half of the workers perceived themselves to be underpaid at the base wage, implying that a large number of the workers accepted the job even though they thought that pay was unfairly low. About one-third of the workers considered themselves to be adequately paid, whereas only few (12%) thought they were overpaid for this job. Because we know from every worker the wage that is considered as fair ("appropriate") we can also calculate how many workers felt underpaid at CHF 27. We find that only 3 out of 119 individuals felt underpaid at this wage. Thus, we have established the following result:

Result 1. The wage increase is associated with an increase in the perceived fairness of pay on average. However, there is strong heterogeneity in workers' fairness perceptions. The wage increase raises the perceived fairness of pay particularly among workers who evaluate the base wage as unfairly low.

### 3.2. Average Treatment Effect of the Wage Increase on Workers' Performance

We now turn to the behavioral effects of the wage increase and begin by investigating its average impact on performance. First evidence is shown in Figure 4, which displays the average performance in the two treatments. Because of the large variation in occupied locations across days, for each observation we subtract the mean productivity at the location in question to provide a meaningful comparison across treatments; a zero thus corresponds to the average performance in a particular location. Since we take the natural logarithm of these numbers, they can be interpreted directly as the percentage differences to the location mean. The figure suggests a positive average treatment effect. In response to the wage increase, workers increased their performance by about 4% on average. The standard errors of the means are rather small, and error bands do not overlap across treatments, possibly indicating that the effect is statistically significant. However, these standard errors do not exploit the other controls available to us and are possibly biased because of temporal and spatial correlations within the residual, as explained in §2.5. We thus turn to the regression estimates that resolve these issues. The first column of Table 6 shows the estimation results for Equation (1). The estimated average treatment effect of the wage increase on workers' performance is 3.7%. In column (2) we also estimate the model using individual fixed effects as specified in Equation (2). The results show that the average treatment effect is slightly lower, but still significant at the 5% level. Thus, in our experiment, there does not seem to be a within-subject amplification of the treatment effect. Overall, we have established the following result:

Result 2. The wage increase has a positive and significant impact on workers' performance on average.

**Table 6 Average Treatment Effect of the Wage Increase on Workers' Performance**

| | (1) | (2) |
|---|---|---|
| Dependent variable: | log(*Hourly copies distributed*) | |
| *CHF27* | 0.037** | 0.030** |
| | (0.017) | (0.014) |
| *Intercept* | 5.608*** | 5.269*** |
| | (0.090) | (0.084) |
| Fixed effects | | |
| Worker | No | Yes |
| Location | Yes | Yes |
| Day | Yes | Yes |
| $R^2$ | 0.594 | 0.718 |
| *N* | 1,269 | 1,269 |

*Notes.* OLS estimates. Standard errors in parentheses take account of serial correlation within an individual's residuals and spatial correlation among the residuals of spatially close observations on the same day (up to a distance of 3 km). The dependent variable is the logarithm of the number of hourly copies distributed and serves as our performance measure. The variable *CHF27* is an indicator variable for the treatment in which the workers were paid the higher wage.

**$p < 0.05$; ***$p < 0.01$.

The point estimates we obtain are roughly in line with the point estimates found in earlier studies that use comparable treatments (Gneezy and List 2006; Kube et al. 2012, 2013). However, the previous studies did not find significant effects, perhaps because of the smaller sample sizes.

We next examine whether exposure to treatment CHF27 changes the effectiveness of the treatment or changes the behavior once the workers returned to the normal wage of CHF 22. Our study allows us to go beyond earlier field experiments as we can identify such effects for up to two workweeks. As pointed out before, we observe our workers on an average of six days. However, 25% of them worked an average of 12 days, thus giving us solid identification of previous exposure to the higher wage on effort over an extended period. In column (1) of Table 7 we estimate the individual fixed effects model from Equation (2) that we augment by two variables: the number of previous shifts a worker was exposed to the CHF 27 wage, as well as an interaction of this variable with the treatment CHF27 indicator. The two variables allow us to examine two separate issues related to what previous exposure to the higher wage does to a worker's motivation. The interaction term allows us to examine whether repeated exposure to the CHF 27 wage reduces its effectiveness. Previous studies suggest that gift exchange has an emotional component that may dissipate over time with the consequence that the extra effort may wear off (Falk 2007, Gneezy and List 2006), and indeed, Gneezy and List (2006) find a decline in the effectiveness of gift exchange even within a day. Column (1) of Table 7 shows the results. The coefficient on treatment CHF27 can be interpreted as the impact of the wage increase on effort when workers encountered the increase for the first time. The estimate is positive and significant, and only somewhat larger than the estimate of the overall treatment effect. The coefficient on the interaction term is very small, and far from significant. Thus, previous exposure to the CHF 27 wage does not weaken workers' response to the higher wage in our case, in slight contrast to Gneezy and List (2006).[15]

**Table 7 Robustness Checks for Habituation and Reference Point Effects**

| | (1) | (2) | (3) |
|---|---|---|---|
| Dependent variable: | log(*Hourly copies distributed*) | | |
| *CHF27* | 0.044* | 0.046* | 0.076** |
| | (0.022) | (0.025) | (0.032) |
| *# previous CHF27 shifts* | 0.008 | 0.025* | |
| | (0.006) | (0.015) | |
| *CHF27* × *# previous CHF27 shifts* | −0.005 | −0.003 | |
| | (0.007) | (0.019) | |
| (*# previous CHF27 shifts*)$^2$ | | −0.002 | |
| | | (0.002) | |
| *CHF27* × (*# previous CHF27 shifts*)$^2$ | | −0.001 | |
| | | (0.002) | |
| *Intercept* | 5.262*** | 5.255*** | 5.383*** |
| | (0.083) | (0.082) | (0.115) |
| Fixed effects | | | |
| Worker | Yes | Yes | Yes |
| Location | Yes | Yes | Yes |
| Day | Yes | Yes | Yes |
| $R^2$ | 0.718 | 0.719 | 0.804 |
| *N* | 1,269 | 1,269 | 645 |

*Notes.* OLS estimates. Standard errors in parentheses take account of serial correlation within an individual's residuals and spatial correlation among the residuals of spatially close observations on the same day (up to a distance of 3 km). The dependent variable is the logarithm of the number of hourly copies distributed and serves as our performance measure. The variable *CHF27* is an indicator variable for the treatment in which the workers were paid the higher wage. The variable *# previous CHF27 shifts* captures the number of previous shifts a worker had already received the higher wage. Column (3) replicates the estimation results in column (2) of Table 6 while excluding observations (and also the subsequent ones) that occurred when workers switched back to the base wage.

*$p < 0.10$; **$p < 0.05$; ***$p < 0.01$.

Conditional on the *CHF27* dummy being equal to zero, the exposure variable alone in column (1) of Table 7 measures how the performance of a worker changes if, having been exposed to a certain number of CHF27 shifts, he returns to the normal pay of CHF 22. One may suspect that workers are disappointed about falling back to the base wage after having encountered the higher wage, because this may have influenced their reference wage they expected to receive (Abeler et al. 2011, Fehr and Goette 2007). The estimated coefficient on the exposure variable is, however, mildly positive. Thus, if anything, workers were somewhat more productive, even when back on the CHF 22 wage,

[15] Gneezy and List (2006) find that effort declines within a workday. We have only one observation per workday and can thus not rule out that gift exchange becomes weaker over the course of the workday.

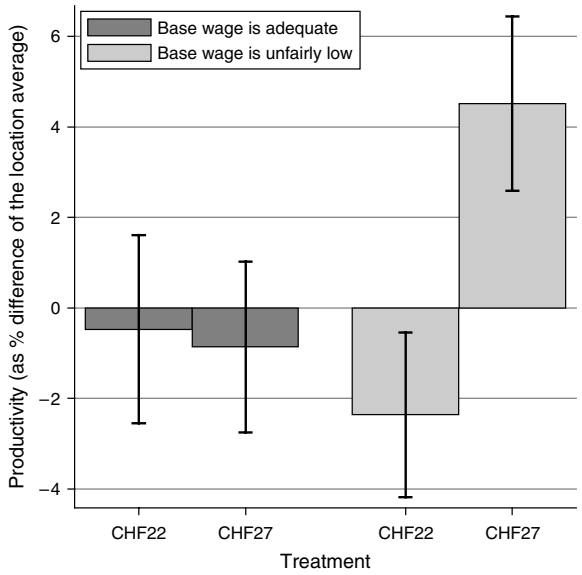
after having been exposed to CHF 27. The effect is, however, not significant. We also explore in column (2) of Table 7 whether exposure effects take perhaps a more nonlinear form. We choose a parsimonious specification and add a quadratic exposure term: if there are strong nonlinearities, the quadratic terms should pick this up. However, the coefficients on the quadratic terms are very small and not significantly different from zero, and an $F$-test cannot reject that all the exposure coefficients are equal to zero ($p = 0.31$).

An alternative way of looking at the issue of disappointment is to only use observations of workers until they go back from the higher wage to the normal wage of CHF 22, as is done in column (3) of Table 7. The estimate of treatment CHF27 is positive and significant, and somewhat larger than the estimated coefficient for the entire sample, though it comes with a larger standard error. Thus, again, this indicates that the difference in effort created by the wage increase is not because of worker disappointment once back at the normal wage of CHF 22.[16] Taken together, these results show no evidence that the wage increase is fading in effectiveness as exposure to the high wage increases or that the effect is strongly driven by a reduction in effort after a higher wage, rather than by an increase in effort as the higher wage is paid.

### 3.3. Heterogeneous Treatment Effects of the Wage Increase on Workers' Performance

In a next step, we attempt to provide direct evidence on the psychological mechanisms behind the responses to the wage increase. We do this by directly testing differential predictions regarding which subsample should respond to the wage increase by different models. In contrast to reciprocity models, a key ingredient of the fair wage-effort hypothesis by Akerlof and Yellen (1990) is that only the workers who feel underpaid at the base wage should respond to the higher wage with higher effort. We therefore investigate whether heterogeneous fairness perceptions lead to heterogeneous effort responses to the wage increase. Figure 5 displays the treatment effect separately for two types of workers: the workers who considered the base wage to be fair and those who considered it to be unfairly low. The figure shows that the workers who felt adequately paid at the base wage (including the few who felt overpaid) did not react to the wage increase at all, whereas those who perceived the base wage as inadequately low increased their performance by about 7%. Judging by the error bands, the figure suggests that the impact of the wage increase is significant for the second group, but not for the first. To better understand the heterogeneity in the treatment effect we estimate Equation (3) where we include an interaction term between treatment CHF27 and the underpayment measure, denoted by $\Delta_i$. Because we need to restrict the sample to individuals who completed the follow-up survey, we first replicate the average treatment effect without taking account of the underpayment perceptions to show that the workers who completed the follow-up survey responded in exactly the same way to the wage increase as all others did. As can be seen in column (1) of Table 8, the point estimate on treatment CHF27 is identical to that in column (1) of Table 6. Column (2) of Table 8 now presents the estimates of Equation (3) that includes the interaction of treatment CHF27 with $\Delta_i$. The estimate of $\beta_1$ is not significant, implying that we cannot reject the hypothesis that the workers who felt adequately paid at the base wage (i.e., $\Delta_i = 0$) did not respond to the wage increase. However, workers who felt underpaid responded quite strongly. The estimate of 0.019 for $\beta_2$ is highly significant and implies that workers' performance response to the wage increase gets stronger with every Swiss franc that a worker felt underpaid at the base wage.[17] We also estimate the model with individual fixed effects as described

**Figure 5** **Heterogeneous Treatment Effect of the Wage Increase on Workers' Performance**




*Notes.* This figure displays workers' average performance and standard error of the mean in each treatment, shown separately for the workers who felt adequately paid at the base wage and those who felt underpaid. Workers' performance is the number of hourly copies distributed normalized by the average productivity at the location.

[16] This is not to say that wage cuts do not lead to effort reductions (e.g., Cohn et al. 2014). Keep in mind that in our setting, the increase was explicitly announced as a bonus coming from the publisher, not the regular employer. This may take some of the sting out of going back to CHF 22.

[17] The coefficient estimate of the interaction term between treatment CHF27 and underpayment perceptions remains the same when we treat the few overpaid workers in the same way as the adequately paid workers, i.e., when we force $\Delta_i$ to be zero for the overpaid workers. We thank an anonymous referee for bringing this to our attention.

**Table 8 Heterogeneous Treatment Effect of the Wage Increase on Workers' Performance**

| | (1) | (2) | (3) |
|---|---|---|---|
| Dependent variable: | log(*Hourly copies distributed*) | | |
| *CHF27* | 0.039** | 0.018 | 0.005 |
| | (0.019) | (0.023) | (0.020) |
| $CHF27 \times \Delta_i$ | | 0.019** | 0.018** |
| | | (0.008) | (0.007) |
| $\Delta_i$ | | 0.001 | |
| | | (0.006) | |
| *Intercept* | 5.619*** | 5.633*** | 5.317*** |
| | (0.113) | (0.108) | (0.130) |
| Fixed effects | | | |
| Worker | No | No | Yes |
| Location | Yes | Yes | Yes |
| Day | Yes | Yes | Yes |
| $R^2$ | 0.596 | 0.601 | 0.725 |
| *N* | 722 | 722 | 722 |

*Notes.* OLS estimates. Standard errors in parentheses take account of serial correlation within an individual's residuals and spatial correlation among the residuals of spatially close observations on the same day (up to a distance of 3 km). The dependent variable is the logarithm of the number of hourly copies distributed and serves as our performance measure. The variable *CHF27* is an indicator variable for the treatment in which the workers were paid the higher wage. The variable $\Delta_i$ is the difference between the wage a worker considered to be fair and the base wage. The interaction term $CHF27 \times \Delta_i$ thus measures the treatment effect as a function of workers' perceived underpayment.

**$p < 0.05$; ***$p < 0.01$.

in Equation (4). In this specification, reported in column (3) of Table 8, the point estimate of $\beta_1$ is again close to zero and insignificant. However, the point estimate of $\beta_2$ is positive and significantly different from zero ($p = 0.01$). It implies that for every Swiss franc that a worker felt underpaid, workers' response to the higher wage increases by 0.019. Thus, for a worker who considered the fair wage to be CHF 25 (as many did), the response is $0.005 + 3 \times 0.019 = 0.062$ or 6.2%; this effect is almost twice as large as the average treatment effect. Thus, the workers who felt underpaid at the base wage responded very differently to the wage increase than those who felt adequately paid, a finding that highlights the importance of heterogeneous fairness perceptions in our understanding of the mechanism behind the motivational effects of pay raises. We summarize these finding in the following result:

Result 3. There is significant heterogeneity in workers' response to a wage increase. Workers who perceive themselves to be underpaid at the base wage raise their performance significantly when they are paid a higher wage, whereas workers who feel adequately paid or overpaid at the base wage do not respond to a wage increase.

In the following, we present two robustness checks for Result 3. The first check examines the role of workers' ability in the treatment effect. We are particularly interested in the hypothesis that high-ability workers have different underpayment perceptions. If, for example, these workers are more likely to perceive themselves to be underpaid, the interpretation of the positive impact of perceived underpayment on the performance response to the wage increase becomes more ambiguous—perhaps it is workers' ability, and not their fairness perceptions that drive the underpayment effect. We use worker fixed effects $\nu_i$ from Equation (2) as a measure of worker ability.[18] Workers with a high $\nu_i$ were, on average, more productive than those with a low $\nu_i$. We find that the correlation between individual ability and the underpayment measure is close to zero and insignificant (Spearman rank correlation: 0.029, $p = 0.75$).[19] This provides a first hint that ability is an unlikely explanation for the underpayment effect. We test this more formally in Table 9. For ease of comparison, column (1) of this table presents again the underpayment effect without controlling for ability. To examine the role of ability, we divide the workers into two groups: workers with a $\nu_i$ above the median (= high-ability), and those with a $\nu_i$ below the median (= low-ability). We then reestimate Equation (1) and include a dummy variable for the high-ability workers as well as an interaction of this dummy variable with the indicator for treatment CHF27. Column (2) of Table 9 shows that the coefficient of the interaction term is insignificant, meaning that we cannot reject the hypothesis that high- and low-ability workers responded similarly to the wage increase. We additionally estimate a model that includes the underpayment variable, the ability variable, as well as their interactions with the treatment indicator. Column (3) of Table 9 shows again that workers' abilities do not drive the underpayment effect. Whereas the point estimate of the underpayment interaction remains robustly positive and significant ($p = 0.01$), the ability interaction is still insignificant.

The second robustness check explores whether the heterogeneous reciprocal responses are caused by differences in how much workers cared about their employer, rather than by differences in fairness perceptions. One could hypothesize that the workers who valued the firm's profit most were also those who reacted most strongly to the higher pay. We use the question "How strongly do you identify yourself with your employer?" from the follow-up survey with answers categories

[18] The individual fixed effects $\nu_i$ serve as a proxy for skills. However, they may capture other worker-specific factors that influence performance, such as intrinsic motivation or conscientiousness. We thank an anonymous referee for pointing this out to us.

[19] We investigated several other potential instruments for perceived wage fairness, such as gender or the difficulty of finding a job. We find no strong predictors of fairness perceptions.

**Table 9 Robustness Checks for Ability and Identification with the Firm**

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Dependent variable: | log(*Hourly copies distributed*) | | | | |
| *CHF27* | 0.018 | 0.022 | 0.007 | 0.031* | 0.013 |
| | (0.023) | (0.026) | (0.028) | (0.018) | (0.022) |
| *CHF27* × $\Delta_i$ | 0.019** | | 0.021** | | 0.017** |
| | (0.008) | | (0.008) | | (0.008) |
| $\Delta_i$ | 0.001 | | −0.006 | | 0.003 |
| | (0.006) | | (0.006) | | (0.006) |
| *CHF27* × *High-ability* (= 1) | | 0.030 | 0.016 | | |
| | | (0.045) | (0.046) | | |
| *High-ability* (= 1) | | 0.232*** | 0.238*** | | |
| | | (0.032) | (0.033) | | |
| *CHF27* × *Identification* | | | | −0.058*** | −0.049** |
| | | | | (0.020) | (0.019) |
| *Identification* | | | | 0.048*** | 0.048*** |
| | | | | (0.013) | (0.013) |
| *Intercept* | 5.633*** | 5.410*** | 5.427*** | 5.635*** | 5.645*** |
| | (0.108) | (0.120) | (0.117) | (0.113) | (0.108) |
| Fixed effects | | | | | |
| Worker | No | No | No | No | No |
| Location | Yes | Yes | Yes | Yes | Yes |
| Day | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.601 | 0.657 | 0.660 | 0.602 | 0.608 |
| *N* | 722 | 722 | 722 | 722 | 722 |

*Notes.* OLS estimates. Standard errors in parentheses take account of serial correlation within an individual's residuals and spatial correlation among the residuals of spatially close observations on the same day (up to a distance of 3 km). The dependent variable is the logarithm of the number of hourly copies distributed and serves as our performance measure. The variable *CHF27* is an indicator variable for the treatment in which the workers were paid the higher wage. The variable $\Delta_i$ is the difference between the wage a worker considered to be fair and the base wage. *High-ability* is an indicator variable for worker ability, which takes the value of 1 if a worker's fixed effect is above the median and 0 otherwise. *Identification* measures to what extent workers identified with the firm, ranging from −2 "very weakly" to 2 "very strongly." The interaction terms *CHF27* × $\Delta_i$, *CHF27* × *High-ability*, and *CHF27* × *Identification* measure the treatment effect as a function of workers' perceived underpayment, their ability, and their identification with the firm, respectively.

*$p < 0.10$; **$p < 0.05$; ***$p < 0.01$.

ranging from −2 "very weakly" to 2 "very strongly" as a measure of how closely connected workers felt to their employer. We again estimate Equation (1) and include the identification variable as well as its interaction with the indicator for treatment CHF27. Column (4) of Table 9 shows that the workers who identified more strongly with their employer provided more effort in the baseline condition ($p < 0.01$). As discussed in Footnote 14, one should be careful with interpreting this coefficient in a causal way though. In contrast to the above hypothesis, these workers reciprocated the higher pay to a lesser extent than those who did not care much about their employer ($p < 0.01$). Moreover, whereas this negative interaction effect remains robust to the inclusion of the underpayment variables, it cannot account for the underpayment effect. Column (5) of Table 9 shows that the point estimate of the underpayment interaction remains almost unchanged ($p = 0.03$). Thus, controlling for identification with the employer does not affect the relationship between perceived underpayment and workers' performance response to the higher wage.

The way fairness perceptions mediate the behavioral responses to the higher wage is consistent with the fair-wage effort hypothesis (Akerlof and Yellen 1990). It stipulates that reciprocal workers respond to an improvement toward the fair wage by increasing effort up to the point where they feel treated fairly. By the same token, this theory also predicts that individuals' behavior displays reciprocity in the sense that unfairly low wages are reciprocated with low effort, whereas high wages are reciprocated with normal effort. To further corroborate this interpretation of our data, our strategy is to use a behavioral measure of reciprocity from a different context, i.e., the choice experiment, and to examine if it predicts behavior consistent with the fair-wage effort hypothesis in the field experiment. This measure, explained in detail in §2.3, shows that there is considerable heterogeneity in reciprocity concerns between individuals. Seventy-seven workers can be classified as reciprocal in the moonlighting game. By contrast, 41 workers cannot be classified as reciprocal: their points returned do not depend on the generosity of the first mover and they are often zero for all three first-mover allocations. Thus, our measure of reciprocal fairness from the choice experiment captures individuals' general inclination toward reciprocity and allows us to test whether it predicts effort responses to the wage increase. We keep in mind the caveat that our reciprocity measure does not distinguish between outcome-based and intention-based models of social preferences, as we discussed before.

We test this by estimating Equation (4) separately for the reciprocal and the nonreciprocal workers.[20] Table 10 displays the results. We first turn to the estimates for the reciprocal individuals in column (1). As before, we find no positive effect of the wage on performance when workers perceived themselves to be adequately paid or overpaid. The coefficient $\beta_1$ on the main effect of treatment CHF27 is small and insignificant. However, we find that fairness perceptions play an even stronger role in workers' performance response to the wage increase. The coefficient of the interaction term, $\beta_2$, is now 0.028, implying that workers' performance

[20] We checked whether underpayment perceptions differ between the two groups. The test results show that we cannot reject the hypothesis that reciprocal workers and nonreciprocal workers do not differ in their underpayment judgments (Kolmogorov-Smirnov test, $p = 0.92$). We also estimated interaction effects for reciprocal and nonreciprocal workers, but we decided to report the results in separate equations for expositional simplicity. Estimating interaction effects yields the same conclusions, but the presentation of the results is more cumbersome.

**Table 10 Effect of the Wage Increase on the Performance of Reciprocal vs. Nonreciprocal Workers**

| | (1) | (2) |
|---|---|---|
| Sample: | Reciprocal workers | Nonreciprocal workers |
| Dependent variable: | log(*Hourly copies distributed*) | |
| *CHF27* | 0.000 | 0.042 |
| | (0.023) | (0.036) |
| $CHF27 \times \Delta_i$ | 0.028** | −0.010 |
| | (0.012) | (0.009) |
| *Intercept* | 5.634*** | 5.928*** |
| | (0.189) | (0.231) |
| Fixed effects | | |
| Worker | Yes | Yes |
| Location | Yes | Yes |
| Day | Yes | Yes |
| $R^2$ | 0.767 | 0.793 |
| *N* | 466 | 243 |

*Notes.* OLS estimates. Standard errors in parentheses take account of serial correlation within an individual's residuals and spatial correlation among the residuals of spatially close observations on the same day (up to a distance of 3 km). The dependent variable is the logarithm of the number of hourly copies distributed and serves as our performance measure. The variable *CHF27* is an indicator variable for the treatment in which the workers were paid the higher wage. The variable $\Delta_i$ is the difference between the wage a worker considered to be fair and the base wage. The interaction term $CHF27 \times \Delta_i$ thus measures the treatment effect as a function of workers' perceived underpayment. Column (1) shows the estimates for reciprocal workers, whereas column (2) shows the same for nonreciprocal workers.
**$p < 0.05$; ***$p < 0.01$.

response to the higher wage increases by 0.028 with every Swiss franc that a worker felt underpaid at the base wage. Thus, an individual who felt underpaid by CHF 3 at the base wage displayed a performance response of 8.4%, which is almost three times as large as an average individual. For reciprocal individuals, the performance response to a wage increase, therefore, strongly depends on their fairness perceptions. By contrast, the estimates for the nonreciprocal individuals in column (2) show no evidence that their performance response depends on fairness perceptions. The point estimate is even negative but not significant. The data reject the hypothesis that fairness perceptions mediate the response to the wage increase the same way for reciprocal and nonreciprocal individuals, i.e., that $\beta_2^R = \beta_2^{NR}$ ($p = 0.01$). Notice, however, that the nonreciprocal group exhibits a more blurred estimate of the response to the wage increase in the absence of underpayment. The estimate of $\beta_1^{NR} = 0.042$ with a standard error of 0.036. Yet, even when taking this blurriness into account, we reject the hypothesis that $(\beta_1^R \beta_2^R)' = (\beta_1^{NR} \beta_2^{NR})'$ in a joint $F$-test ($p = 0.04$). Furthermore, we also reject the hypothesis that the wage increase did not affect the response of the reciprocal workers ($p = 0.02$), whereas we cannot reject the null of no response for the nonreciprocal workers. Thus, only the reciprocal workers responded to the wage increase, and this response takes the form of the fair-wage effort hypothesis. This pattern is statistically different from what we observe for the nonreciprocal workers, even though we cannot characterize the nonreciprocal workers' behavior with much precision. For this last group, our simplest interpretation is that they did not respond to the wage increase. We summarize these findings in the following results:

Result 4. Underpaid reciprocal workers strongly increase their performance when they are paid a higher wage, whereas the pattern is significantly different for nonreciprocal workers: even when feeling underpaid, nonreciprocal workers do not respond to a wage increase.

## 4. Discussion and Conclusion

This paper presents a study that combines a field experiment examining the impact of a wage increase on workers' performance with both a survey measure of workers' fairness perceptions and a "laboratory" measure of workers' inclination toward reciprocity. The combination of these tools enables us to acquire deeper insights into the role of fairness perceptions and reciprocity in workers' response to a wage increase. Our results show that the perceived fairness of the base wage plays an important role in mediating the effort response to a wage increase. Workers who perceive being underpaid at the base wage increase their effort in response to a wage increase, whereas those who feel fairly paid or overpaid do not respond to the higher wage. We also find that only those workers whom we classify as reciprocal individuals in a two-player choice experiment respond to the wage increase with higher performance, whereas those whom we classify as nonreciprocal and selfish leave their performance unchanged irrespective of whether or not they feel underpaid. Taken together, our evidence suggests that the response to the wage increase is mainly driven by the elimination of perceived unfairness of the baseline pay, i.e., the elimination of negative reciprocity, rather than positive reciprocity because of the higher wage paid. This corresponds closely to the assumptions in Akerlof and Yellen (1990).

In our setup, the workers knew from the outset that the job for the newspaper's launch was a one-time job that would last only a few weeks, and that the wage increase came from the publishing house not the promotion agency. It is thus unlikely that reputational or repeated game effects affected the results. In principle, the assumption of fair-minded workers is not necessary to explain a positive relationship between wages and effort (MacLeod and Malcomson 1989, 1998; Shapiro and Stiglitz 1984). These alternative theories, however, are based on the assumption of long-term employment

relations, which is difficult to believe in our setting.[21] In addition, these theories predict that even selfish workers without any inclination toward reciprocal fairness respond in a reciprocal manner to the wage increase. This is because in a repeated game, the selfish types also have an incentive to raise effort as it may pay to do so if one considers the employer's response in the future. In our data, however, we find no evidence that the selfish workers respond to a wage increase. Theories based on repeated interactions are thus hard pressed to explain why only the workers who exhibit an inclination toward reciprocal fairness (in the choice experiment) responded to the wage increase, and why this reaction was different from that of selfish workers.

So far we have studied the statistical significance of a wage increase and tried to understand the psychological mechanism behind it. Studying these effects and understanding the psychological mechanisms is important in its own right. But another question concerns the economic significance of our results. Recall that in response to a 20% wage increase, effort increased on average by roughly 4%. For a reciprocal worker who felt that CHF 27 was the appropriate wage for this job, the increase in effort would have been 14% (see Table 10). Are these increases in effort profitable for a company? As we show analytically in Online Appendix B, it depends on the production function. If effort and workers are perfect substitutes, then the firm can always increase output by 10% by hiring 10% more workers, raising costs by 10%. In this case, raising the wage by 10% is only profitable if it raises effort—and thus output—by at least 10%. In our setting, production is independent and this may be the relevant benchmark. Paying higher wages is most likely not profitable and this may also explain why the firm paid the low wage of CHF 22. If, however, effort and workers are not perfect substitutes and are combined in the manner of Cobb–Douglas into output, this is no longer the case. In particular, if the elasticity of output is greater with respect to effort than with respect to workers (i.e., if the Cobb–Douglas exponent on effort is greater than 0.5), the firm can increase output more efficiently by boosting effort than by increasing the number of workers. This could be the case in work settings where one worker's effort also raises the productivity of others, as it is realistic in many settings. As we show in Online Appendix B, if the Cobb–Douglas parameter is 0.59 for effort and 0.41 for workers, then raising the wage is profitable if a 20% raise yields 14% more effort, which is roughly corresponds to the magnitude we find for individuals who feel severely underpaid. Thus, the magnitudes we find in our context are nonnegligible and may even be profitable in other business environments with only slight departures from perfect substitutability.

Our results also indicate that laboratory and field evidence are not in conflict with each other. If we control for workers' fairness perceptions, workers in the field reciprocate on average higher wages with higher performance, as they do in the laboratory. Likewise, workers who display reciprocal responses in the choice experiment are those who also show reciprocal responses in the field. Our results thus also lend partial support to the notion that laboratory results generalize to field settings. In fact, as we mentioned above, our choice-based measure of reciprocity helps us better understand the heterogeneity in field responses. However, our results also strongly suggest that the increase in effort came exclusively from workers who felt underpaid at the baseline wage, and were therefore mediated through the removal of negative reciprocity. Workers who felt adequately paid or overpaid did not raise their effort. This is in contrast to laboratory results that routinely also find strong effects of positive reciprocity (Charness 2004, Fehr et al. 1993).

The insights gained from this study also point to new directions for future research. Our study shows that the composition of the work force in terms of perceived underpayment and reciprocity crucially determines the effectiveness of higher wages. This suggests that much can be gained by studying the determinants of workers' fairness perceptions and how firms can shape them through wage and nonwage instruments. A study by Greenberg (1990), for example, suggests that a sensitive and careful explanation of the pay level may be an effective and inexpensive way to help prevent that workers feel underpaid when they are paid a low wage. An alternative way to mitigate underpayment feelings could be to make coworker wages less transparent. Similarly, Ockenfels et al. (2015) show that low-paid workers are particularly unsatisfied with their job and also reduce performance when they have a rather good knowledge about their relative standing in the wage hierarchy. Understanding of how compensation executives can "manage" workers' fairness perceptions in a productivity enhancing way seems to be a fruitful avenue for future research.

[21] Strictly speaking, these models are based on the assumption of infinitely repeated interactions between the firm and the workers. With strictly selfish preferences and common knowledge of rationality, this is a necessary condition to generate these effects (Kreps et al. 1982).

## Supplemental Material

Supplemental material to this paper is available at http://dx.doi.org/10.1287/mnsc.2014.1970.

## Acknowledgments

The authors thank three anonymous referees for helpful comments. The authors acknowledge the generous support of the Swiss National Science Foundation [Grant 100018_140734/1 on "The Distribution and Determinants of Social Preferences"].

## References

Abbink K, Irlenbusch B, Renner E (2000) The moonlighting game: An experimental study on reciprocity and retribution. *J. Econom. Behav. Organ.* 42:265–277.

Abeler J, Falk A, Goette L, Huffman D (2011) Reference points and effort provision. *Amer. Econom. Rev.* 101:470–492.

Akerlof GA (1982) Labor contracts as partial gift exchange. *Quart. J. Econom.* 97:543–569.

Akerlof GA (1984) Gift exchange and efficiency-wage theory—4 views. *Amer. Econom. Rev.* 74:79–83.

Akerlof GA, Yellen J (1990) The fair-wage effort hypothesis and unemployment. *Quart. J. Econom.* 105:255–283.

Akerlof GA, Dickens WT, Perry GL, Gordon RJ, Mankiw NG (1996) The macroeconomics of low inflation. *Brookings Papers Econom. Activity* 1996:1–76.

Al-Ubaydli O, List JA (2014) On the generalizability of experimental results in economics. Fréchette G, Schotter A, eds. *The Methods of Modern Experimental Economics* (Oxford University Press, Oxford, UK). Forthcoming.

Angrist JD, Pischke JS (2009) *Mostly Harmless Econometrics: An Empiricist's Companion* (Princeton University Press, Princeton, NJ)

Bandiera O, Barankay I, Rasul I (2005) Social preferences and the response to incentives: Evidence from personnel data. *Quart. J. Econom.* 120:917.

Bandiera O, Barankay I, Rasul I (2009) Social connections and incentives in the workplace: Evidence from personnel data. *Econometrica* 77(4):1047–94.

Bandiera O, Barankay I, Rasul I (2010) Social incentives in the workplace. *Rev. Econom. Stud.* 77:417–58.

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

Bellemare C, Kroeger S, Van Soest A (2008) Measuring inequity aversion in a heterogeneous population using experimental decisions and subjective probabilities. *Econometrica* 76:815–839.

Bellemare C, Kroger S, van Soest A (2011) Preferences, intentions, and expectation violations: A large-scale experiment with a representative subject pool. *J. Econom. Behav. Organ.* 78:349–365.

Brandts J, Charness G (2004) Do labour market conditions affect gift exchange? Some experimental evidence. *Econom. J.* 114:684–708.

Brandts J, Charness G (2011) The strategy versus the direct-response method: A first survey of experimental comparisons. *Experiment. Econom.* 14:375–398.

Cabrales A, Calvó-Armengol A, Pavoni N (2008) Social preferences, skill segregation, and wage dynamics. *Rev. Econom. Stud.* 75:65–98.

Camerer CF (2014) The promise and success of lab-field generalizability in experimental economics: A reply to Levitt and List. Fréchette G, Schotter A, eds. *The Methods of Modern Experimental Economics* (Oxford Univeristy Press, Oxford, UK). Forthcoming.

Charness G (2000) Responsibility and effort in an experimental labor market. *J. Econom. Behav. Organ.* 42:375–384.

Charness G (2004) Attribution and reciprocity in an experimental labor market. *J. Labor Econom.* 22:665–688.

Charness G, Levine DI (2007) Intention and stochastic outcomes: An experimental study. *Econom. J.* 117:1051–1072.

Charness G, Rabin M (2002) Understanding social preferences with simple tests. *Quart. J. Econom.* 117:817–869.

Charness G, Frechette GR, Kagel JH (2004) How robust is laboratory gift exchange? *Experiment. Econom.* 7:189–205.

Charness G, Gneezy U, Kuhn MA (2012) Experimental methods: Between-subject and within-subject design. *J. Econom. Behav. Organ.* 81:1–8.

Cohn A, Fehr E, Herrmann B, Schneider F (2014) Social comparison and effort provision: Evidence from a field experiment. *J. Eur. Econom. Assoc.* 12:877–898.

Conley TG (2008) Spatial econometrics. Durlauf SN, Blume LE, eds. *The New Palgrave Dictionary of Economics* (Palgrave Macmillian, London).

Cooper D, Kagel JH (2014) Other regarding preferences: A selective survey of experimental results. Kagel JH, Roth AE, eds. *Handbook of Experimental Economics*, Vol. 2 (Princeton University Press, Princeton, NJ). Forthcoming.

Dohmen T, Falk A, Huffman D, Sunde U (2009) Homo reciprocans: Survey evidence on behavioural outcomes. *Econom. J.* 119: 592–612.

Falk A (2007) Gift exchange in the field. *Econometrica* 75:1501–1511.

Falk A, Fischbacher U (2006) A theory of reciprocity. *Games Econom. Behav.* 54:293–315.

Falk A, Fehr E, Fischbacher U (2003) On the nature of fair behavior. *Econom. Inquiry* 41:20–26.

Falk A, Fehr E, Zehnder C (2006) Fairness perceptions and reservation wages—The behavioral effects of minimum wage laws. *Quart. J. Econom.* 121:1347–1381.

Fehr E, Goette L (2005) Robustness and real consequences of nominal wage rigidity. *J. Monetary Econom.* 52:779–804.

Fehr E, Goette L (2007) Do workers work more if wages are high? Evidence from a randomized field experiment. *Amer. Econom. Rev.* 97:298–317.

Fehr E, Schmidt KM (1999) A theory of fairness, competition, and cooperation. *Quart. J. Econom.* 114:817–868.

Fehr E, Kirchsteiger G, Riedl A (1993) Does fairness prevent market clearing? An experimental investigation. *Quart. J. Econom.* 108:437–460.

Gneezy U, List J (2006) Putting behavioral economics to work: Field evidence of gift exchange. *Econometrica* 74:1365–1384.

Greenberg J (1990) Employee theft as a reaction to underpayment inequity: The hidden cost of pay cuts. *J. Appl. Psych.* 75:561.

Hannan RL, Kagel JH, Moser DV (2002) Partial gift exchange in an experimental labor market: Impact of subject population differences, productivity differences, and effort requests on beha. *J. Labor Econom.* 20:923–951.

Harrison GW, List JA (2004) Field experiments. *J. Econom. Literature* 42:1009–1055.

Hennig-Schmidt H, Sadrieh A, Rockenbach B (2010) In search of workers' real effort reciprocity—A field and a laboratory experiment. *J. Eur. Econom. Assoc.* 8:817–837.

Hicks JR (1932) *The Theory of Wages* (Macmillan, London).

Hossain T, List JA (2012) The behavioralist visits the factory: Increasing productivity using simple framing manipulations. *Management Sci.* 58:2151–2167.

Hsiang SM (2010) Temperatures and cyclones strongly associated with economic production in the Caribbean and Central America. *Proc. Natl. Acad. Sci. USA* 107:15367–15372.

Kreps DM, Milgrom P, Roberts J, Wilson R (1982) Rational cooperation in the finitely repeated prisoners' dilemma. *J. Econom. Theory* 27:245–252.

Kube S, Maréchal MA, Puppe C (2012) The currency of reciprocity-gift-exchange in the workplace. *Amer. Econom. Rev.* 102:1144–62.

Kube S, Maréchal MA, Puppe C (2013) Do wage cuts damage work morale? Evidence from a natural field experiment. *J. Eur. Econom. Assoc.* 11:853–870.

Levitt S, List JA (2007) What do laboratory experiments measuring social preferences reveal about the real world? *J. Econom. Perspect.* 21:153–174.

List JA (2006) The behavioralist meets the market: Measuring social preferences and reputation effects in actual transactions. *J. Political Econom.* 114:1–37.

MacLeod W, Malcomson J (1989) Implicit contracts, incentive compatibility, and involuntary unemployment. *Econometrica* 57: 447–480.

MacLeod WB, Malcomson JM (1998) Motivation and markets. *Amer. Econom. Rev.* 88:388–411.

Marshall A (1890) *Principles of Economics* (Macmillan, London).

Mas A, Moretti E (2009) Peers at work. *Amer. Econom. Rev.* 99:112–145.

Ockenfels A, Sliwka D, Werner P (2015) Bonus payments and reference point violations. *Management Sci.* 61:1496–1513.

PayScale (2014) Temporary office assistant/clerical salary (United States). Accessed September 14, 2014, http://www.payscale.com/research/US/Job=Temporary_Office_Assistant_%2F_Clerical/Hourly_Rate.

Rabin M (1993) Incorporating fairness into game theory and economics. *Amer. Econom. Rev.* 83:1281–1302.

Shapiro C, Stiglitz JE (1984) Equilibrium unemployment as a worker discipline device. *Amer. Econom. Rev.* 74:433–444.

Slichter SH (1929) The current labor policies of American industries. *Quart. J. Econom.* 43:393.

Stoop J, Noussair CN, Van Soest D (2012) From the lab to the field: Cooperation among fishermen. *J. Political Econom.* 120: 1027–1056.