The international platform of Ludwig-Maximilians University's Center for Economic Studies and the Ifo Institute

CESifo Working Papers
www.cesifo.org/wp

Timing of Kindness – Evidence from a Field Experiment

Axel Ockenfels
Dirk Sliwka
Peter Werner

CESIFO WORKING PAPER NO. 4885
CATEGORY 13: BEHAVIOURAL ECONOMICS
JULY 2014

CESifo
Center for Economic Studies & Ifo Institute

CESifo *Working Paper No. 4885*

Timing of Kindness – Evidence from a Field Experiment

Abstract

We conduct a field experiment in a naturally occurring labor environment and track whether the performance of workers responds to unexpected wage increases. Specifically, we investigate how the timing of wage increases affects efforts. We find that workers' performance is about 11% higher for the same total wage when their wage is increased in two steps as opposed to a single increase at the outset. Moreover, workers are more honest and are more willing to do voluntary extra work after surprising wage increases compared to a baseline condition without increases.

JEL-Code: C930, D640, J330, M520.

Keywords: gift exchange, positive reciprocity, field experiment, honesty.

Axel Ockenfels
University of Cologne
Department of Economics
Albertus-Magnus-Platz
Germany – 50923 Cologne
ockenfels@uni-koeln.de

Dirk Sliwka
University of Cologne
Seminar for Personnel Economics &
Human Resource Management
Albertus-Magnus-Platz
Germany – 50923 Cologne
dirk.sliwka@uni-koeln.de

*Peter Werner**
University of Cologne
Department of Economics
Albertus-Magnus-Platz
Germany – 50923 Cologne
peter.werner@uni-koeln.de

*corresponding author

This version: 26 June 2014

1. *Introduction and hypotheses*

Unexpected extra-payments trigger positive reciprocal responses by workers. For instance, in a field experiment conducted with library workers and solicitors, Gneezy and List (2006) observed a positive response to unexpected wage increases. Likewise, Bellemare and Shearer (2009) introduced surprise bonuses within a tree-planting firm that were also found to increase performance. However, little is known about the impact of the timing of such gifts. That timing may be an important issue is suggested by the fact that both studies mentioned above found that the positive effect on workers' efforts was not persistent. Gneezy and List (2006) report that the initially positive response basically vanished after only 1.5 to 3 hours, and Bellemare and Shearer (2009) find an effect only on the day of the gift, but not thereafter.[1]

Our study is the first attempt to investigate whether and how the timing of (unexpected) kindness matters. We investigate whether two unexpected wage increases, keeping total wage costs fixed, can boost workers' performance beyond what can be done with only one surprise. More specifically, we conduct a field experiment in a naturally occurring labor environment in a university library, comparing three different treatments. In the baseline treatment, the workers received a pre-announced wage (8 Euros per hour), which is equal to the usual market wage. In a second treatment, workers were informed about a substantial and unexpected wage increase before they started to work (12 Euros per hour). In our final treatment, the unexpected wage increase was split into two surprises: workers

[1] Bellemare and Shearer (2009) do not investigate whether behavior changes throughout a day. In a study by Landry et al. (2011), unconditional non-monetary rewards triggered performance increases, but to a lesser extent than rewards that were given conditional on reaching a performance target. Englmaier and Leider (2012) found more persistent positive reciprocal responses when the manager who assigned the wage increase was remunerated according to the performance of the agents. Combining laboratory and field experiments, Hennig-Schmidt, Rockenbach and Sadrieh (2010) also came to the conclusion that the assessment of the surplus an employee creates for the principal has an important impact on gift exchange. Cohn, Fehr and Götte (2012) identified positive performance effects of wage increases in a sample of Swiss newspaper distributors which, however, were found only among subjects with a high preference for reciprocity. Finally, two studies emphasized the relative importance of negative over positive reciprocity: Chemin, Delaat and Kurmann (2011) and Kube, Maréchal and Puppe (forthcoming) found distinct negative performance adjustments after surprise wage cuts whereas positive responses to surprise wage increases were only rather weak or insignificant. Fehr, Götte and Zehnder (2009) and Charness and Kuhn (2011) provide recent surveys of related laboratory studies. That kindness may trigger reciprocal responses have been formalized by influential theories, including Akerlof (1982), Rabin (1993), Levine (1998), Dufwenberg and Kirchsteiger (2004), and Falk and Fischbacher (2006).

received a moderately higher wage in the beginning (10 Euros per hour), and an additional wage increase in the middle of the working day (yielding 14 Euros per hour), holding the overall wage bill constant.

Standard theory, assuming purely selfish preferences, would suggest that effort levels are the same in all three treatments (our first Null Hypothesis), as wages are not conditioned on effort levels. However, laboratory and field evidence suggest that wage increases may be perceived as a kind act by the workers, which lead to positively reciprocal responses. This suggests that workers exhibit more effort in our treatments with extra-payments (Hypothesis 1). Regarding the timing of gifts, if the kindness sensation of workers is proportionally related to the extra-payment, distributing the extra-payment unevenly over time should not increase the overall effort level. Also, to the extent workers judge further kind acts as only confirming previously attributed kindness, there would be no value of splitting the extra-payment up in smaller portions. This suggests that 'additional surprises' cannot increase total efforts if total payments are held constant (our second Null Hypothesis). On the other hand, recent experiments by Kosfeld and Neckermann (2011) and Kube, Maréchal and Puppe (2012) suggest that the sensation of receiving a gift, rather than its monetary value, is the decisive aspect that triggers positive reciprocal responses. So, if in our context efforts are triggered by gifts *per se*, repeatedly increasing wages may further increase efforts even when the overall wage sum stays constant (Hypothesis 2).

Concerning the first hypothesis, and in line with previous evidence, we find that workers who receive an increase of their hourly wages perform better than workers who are paid their expected wage. In contrast to previous studies, the output increase is not immediate, though. Instead, workers in our gift treatments perform substantially better in the second half of the experiment, whereas the performance of workers in the control treatment decreases slightly on average.

Importantly, in line with Hypothesis 2, workers exert more effort for the same total wage bill if the wage is increased in two steps rather than in one step. Calculated over the project day, performance rises by about 11% when workers receive two pay increases

instead of one, even though this comes at no cost for the employer. Hence, more frequent wage increases may improve performance to a stronger extent than one wage increase in the beginning. Finally, workers in the gift treatments are willing to spend more extra-time working and exaggerate their performance reports less strongly than workers in the baseline condition.

Section 2 presents the environment in which our experiment was conducted and describes our experimental treatments that vary the timing of extra-payments. Our results are described in Section 3. Section 4 discusses our findings and concludes.

2. The Library Inventory Project

Our field experiment was carried out as part of a modernization project of the economics libraries at the University of Cologne in August 2012. One task of the project consisted of inserting adhesive labels with barcodes into the book stock (approximately 150,000 books) to enable automated borrowing procedures. A service agency specialized in recruiting temporary workers was commissioned to hire workers for a one-time job opportunity that would last for seven working hours in total and consisted of a library inventory task for a fixed hourly wage. Altogether 98 persons signed up for the job.

The working environment is well suited for our research purpose: As the agency is specialized in hiring agents for jobs on trade-fairs, providing jobs for one day only is a common practice and not unusual for the hired workers. All organizational issues including recruiting and the processing of payments are handled by the agency. Moreover, employees of trade fair agencies typically work as freelancers and apply for jobs at several agencies depending on which agency publishes a suitable job advertisement (Table A1 in the Appendix lists the demographic data that were collected from the agency). Interactions between employer and employees are mostly one-time and end after the project is carried out. Therefore, there is little room for reputation building in this environment.

The inventory project proceeded as follows: On each of six project days in August 2012, between 6 and 23 workers arrived at the University of Cologne in the morning and were guided to one of five libraries located at different places on campus. The inventory task would have been carried out irrespective of our experiment, and workers were not informed that the project was part of an experiment. When workers arrived, they were welcomed and, after explaining safety regulations, instructed how to carry out the working task. After the instruction phase (that lasted 30 minutes) and before the work started, every worker received rolls with 200 adhesive labels each and after that was assigned to a shelf with books in which the labels had to be affixed. When workers had completed a shelf or ran out of labels, they contacted the help desk where they would receive more rolls or be guided to a new shelf.

Workers were assigned to library rooms in which they worked alone; in cases where this was not possible (e.g., one library consisted of a very large lecture room), workers were assigned to shelves distant from each other to prevent communication. Moreover, libraries were open during the project, and workers were instructed to work silently in order not to disturb the students who were using the library.

Workers could shirk in several dimensions. First, as we paid a fixed hourly wage, there was no extrinsic incentive to exert effort, and, given that the working task was tedious and monotonous, it seems reasonable to assume that the intrinsic motivation of the workers was limited. Second, all workers decided on their own when and how long they would take breaks, although they were paid for the full seven hours irrespective of the time spent in breaks. Moreover, at the end of the day, workers were asked to kindly stay for another, yet unpaid 20 minutes and help the library team to sort an alphabetical library catalogue. Workers were free to leave if they did not want to participate in the extra task. Finally, workers had the possibility to cheat with respect to their performance reports. Shortly before the working day was over, student helpers approached the workers and told them that the organizers of the project wanted to reward the best performers. Therefore, when leaving, workers were asked to report their performance in hundreds of tags; it was announced that after the inventory project was over, performance of all workers would be compared and the best 30% would receive an additional bonus of 15

Euros to be paid together with the fixed wage. The bonus was announced at the end of the working day to make sure that it did not confound possible reciprocity effects in response to increases of the fixed hourly wage. Workers knew their working performance at the end of the day, because before the task started, they were asked to count the number of completed tag rolls throughout the working day and report the sum to the help desk before leaving in the evening. It was explained to the workers that this was done to facilitate the organization of the inventory project, in particular with respect to the assignment of workers to library rooms on subsequent project days. Finally, they were told that bonuses would be determined on the basis of the performance reports of all workers participating in the inventory project so that the incentive to cheat was not related to the number of workers in a particular shift. While we did not verify workers' statements prior to the bonus payments (which was apparent to the workers), we were able to detect systematic deviations between reported and actual performance as research assistants at the help desk secretly tracked how many rolls with labels each worker in fact received.[2]

We conducted three experimental treatments to explore the effect of the timing of gifts. Every worker within a particular shift was assigned to the same treatment, and none of the workers participated in more than one treatment. The assignment of workers to the treatments was determined by the project day for which the worker registered (Table A2 in the Appendix shows that our treatments were comparable with respect to the number of workers per session and the demographic structure of the workforce).[3] In the Baseline condition, workers received a fixed hourly wage of 8 Euros as was announced in the job advertisement and equal to a typical market wage for similar tasks. Similar to related field studies (see, for example, Gneezy and List 2006 and Kube, Maréchal and Puppe forthcoming) we introduced an unexpected monetary gift by increasing workers' wages

[2] Working performance and the time used for breaks was rigorously tracked by the research assistants during the project days. When a worker collected new rolls with adhesive tags at the help desk, the research assistant noted the number of tags and the time after the worker had left. Moreover, workers were asked to give notice at the help desk when they left their workplace to make a break and when they returned after the break.

[3] If we estimate simple probit models with the probability to participate in one particular treatment as the dependent variable and the age and gender of a particular person as independent variables, we do not find evidence for selection effects: in all three models, the independent variables are insignificant.

prior to the start of the working day in two other conditions. After workers had arrived and were instructed, they were privately approached by an experimental helper who informed them about the wage increase in the gift treatments.[4] Wage increases were communicated to the workers in a neutral way: It was announced that the project had obtained an extra-budget that would be passed on to the workers; no further explanation was given. Our treatments varied the way how the monetary gift was distributed. In the Gift_1 treatment, the full wage increase was implemented *initially*; workers were informed that they would receive an hourly wage of 12 Euros instead of 8 Euros before they started to work. In the Gift_2 treatment, however, an identical monetary gift was distributed *successively*. Workers received a first wage increase at the beginning of the working day. Here, the hourly wage was increased to 10 Euros. After three working hours were over (at this point, workers had been at the library for three and a half hours), they were approached and informed that the organizers of the projects had decided to pass on an additional amount of the extra-budget to the workers so that hourly wages would be increased again, now to 14 Euros for the remaining three and a half hours of the project day. This way, the total fix payment in both gift treatments is identical.

3. Results

Workers in the Baseline treatment inserted on average 1,031 tags into books on the project day. In line with Hypothesis 1, performance slightly increases by 3.9% to 1,071 tags in our treatment Gift_1 where workers received the full pay rise before they started to work. But this overall difference is insignificant (two-sided Mann Whitney U (MWU) test, $p = 0.353$).[5] However, workers who receive the pay rise in two steps (Gift_2) increase performance on average by 15.3% compared to Baseline to altogether 1,189 tags, and this difference is significant ($p = 0.025$). Moreover, total performance is also significantly higher (by 11.0%) in Gift_2 than in Gift_1 ($p = 0.047$), which confirms our

[4] The reason for addressing subjects privately was again to avoid communication among workers about the wage increase.

[5] Although our alternative Hypotheses 1 and 2 are one-sided, we employ two-sided tests everywhere in the paper.

Hypothesis 2. An overview of descriptive statistics on the treatment level can be found in Table A3 in the Appendix.[6]

To capture the dynamics of worker behavior, we compare performance across treatments for hours 1 to 3 and for hours 4 to 6 of the working day (see Figure 1). Our measure for output is the total number of labels a person has inserted in a particular time interval.[7] As starting times differed to some extent between working days and libraries, we measured the output per working hour after the start at the respective library on the respective day. We here compare performance within six hours after the start of the work, because the net working time is at least 6 hours for each library and day.[8]

Interestingly, as Figure 1 shows, performance does not differ in the first three hours. Average performance per worker for hours 1-3 accounts for 488, 447 and 486 tags in Baseline, Gift_1 and Gift_2, respectively.[9] However, a substantial positive time trend is found in both gift treatments: Compared to the first half of the experiment, workers in Gift_1 significantly increased their performance by 19.9% to 536 tags on average, in Gift_2 at 25.9% (to 612) the performance increase was even more substantial. In the control treatment, however, productivity (insignificantly) decreased by 5.2% to 462 tags.[10] As a result, comparing output levels in the second half of the experiment (hours 4-

[6] To make sure that the tags collected from the help desk were actually inserted into the books and thus reflected actual performance, we checked random samples of books for each person from the shelves he/she worked on during the project day. As it was tracked when a person worked on a particular shelf, the sample covered both the first and the second half of the working day. We did not encounter a single case where the tag was missing in a particular book, suggesting that this way of employee cheating was not relevant in our setting. Figure A1 in the Appendix displays the distribution of output across treatments.

[7] This was done with an exact protocol where it was (secretly) noted when each worker collected a new roll at the help desk. For the analyses in which performance is tracked for different time intervals, the 200 tags from the first roll within a new hour were split proportionally according to the share of the time since the last role was collected.

[8] Employees inserted labels in different libraries, but the instruction was carried out in a central room, so that working hours as well as the start und end times differed between locations. However, at every location, at least 6 working hours were achieved; the minimum working time accounted for 365 minutes or 6 hours and 5 minutes.

[9] There are no significant differences in the distributions of working performance between our gift treatments and the baseline (p-values of two-sided Mann Whitney U tests comparing performance to Baseline yield $p = 0.338$ for Gift_1 and $p = 0.798$ for Gift_2). The difference between performances in Gift_1 and Gift_2 is weakly significant ($p = 0.093$).

[10] Comparing individual changes within each treatment, two-sided Wilcoxon Matched Pairs Signed Rank (WMPSR) tests indicate that the increase in productivity is significant in Gift_1 and Gift_2 (both with $p = 0.001$), whereas the productivity decrease in Baseline is not ($p = 0.434$). The residual of total performance

6), we find that average performance in the gift treatments is significantly higher than in Baseline, as two-sided MWU tests reveal (p = 0.046 for Gift_1 versus Baseline and p = 0.002 for Gift_2 versus Baseline). In addition, we observe a (weakly) significant difference between Gift_1 and Gift_2 (p = 0.058) in the second half of the working day, again suggesting that the repeated experience of a wage increase triggers additional positive performance effects and thus leads to a higher persistence of reciprocal behavior.

Figure 1. Dynamics of performance per treatment (in mean number of tags)

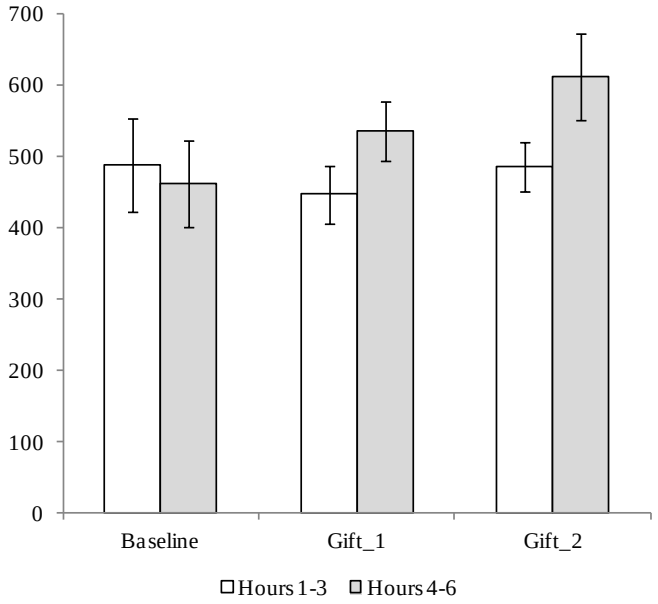


Figure 1 displays the mean number of tags per worker in hours 1-3 and 4-6, separately for each experimental treatment. The lines indicate 95%-confidence intervals for the mean values.

To further study time trends, we estimate regression models with a worker's performance in tags per particular working hour as the dependent variable. We control for subject specific heterogeneity by including random effects per experimental worker. Moreover, for all models we report robust standard errors clustered on the level of a shift (the group of persons who worked in a particular library on a particular project day). As further controls, we include the age and gender of a person and the number of workers in a particular shift. Table 1 shows the results.

---

and the sum of performance in Hours 1-3 and Hours 4-6 is the performance in the (incomplete) seventh working hour that is not comparable due to the differing end times across locations.

Table 1. Hourly performance (in number of tags)

Model No.	1	2	3
Dependent variable	Hourly Performance	Hourly Performance	Hourly Performance
Gift_1	5.283	-10.802	-14.979
	[8.426]	[12.707]	[11.702]
Gift_2	22.874**	-2.368	-3.562
	[11.302]	[13.188]	[12.484]
Hour		1.101	
		[2.429]	
Gift_1 X Hour		6.434**	
		[2.614]	
Gift_2 X Hour		10.097***	
		[3.604]	
Hours 4-6			-10.918
			[10.205]
Gift_1 X (Hours 4-6)			40.523***
			[11.319]
Gift_2 X (Hours 4-6)			52.872***
			[15.191]
Constant	200.554***	197.802***	206.013***
	[33.904]	[34.908]	[36.084]
Observations	582	582	582
Wald χ^2-value	13.0	93.9	69.8

The models are random effects specifications that use a worker's hourly performance in tags as the dependent variable. Robust standard errors are clustered on the level of a shift (all workers from a particular library at a particular project day). Control variables include the age and gender of a worker and a variable indicating the number of workers in the shift. ** and *** denominate significance on the 5%- and 1%-level, respectively. In Model 2, the variable Hour is scaled from 0 (hour 1) to 5 (hour 6) so that the treatment dummies reflect the differences in the first hour of the working day.

In Model 1, we only include the dummy variables for our experimental treatments. Similar to the descriptive statistics, the coefficient of Gift_1 is positive but insignificant, whereas the coefficient of Gift_2 is significant with $p = 0.043$. Moreover, the stronger overall performance effect observed in the Gift_2 treatment is corroborated in the model, as its coefficient is significantly larger than the coefficient of Gift_1 ($p = 0.050$, two-sided Wald-test). To capture the time trend of performance we include the working hour and the corresponding interaction terms with the treatment dummies in Model 2. Our variable for working hour is scaled from 0 (hour 1) to 5 (hour 6) so that the treatment dummies reflect the differences in the first hour of the working day. In line with the lack of a strong initial gift exchange effect found in the descriptive statistics, both treatment dummies are insignificant (the coefficients of Gift_1 and Gift_2 even turn negative). Yet

the interactions of treatment dummies with the number of hours are positive and significant in both cases, corroborating the notion that performance increases over time when workers receive a pay rise. But the coefficients of the two interaction terms do not differ from each other, as a two-sided Wald test indicates ($p = 0.196$). Finally, reflecting the roughly constant performance in Baseline, the coefficient of the working hour is insignificant.

In Model 3, we include a dummy for the second half of the working day (Hours 4-6) and interactions to the treatment dummies instead of the variable for hours. The dummy for specific effects in the second half of the experiment is important especially for the Gift_2 treatment, as here the workers are informed about the second wage increase at the beginning of the fourth working hour. The coefficients of the Gift_1 X (hours 4-6) and Gift_2 X (hours 4-6) interaction terms are both positive and significant, corroborating the positive effects of the wage increases in the second half of the working day. Similar to Model 2, while the estimated effect of Gift_2 X (hours 4-6) is more than 30% larger than the estimated effect of Gift_1 X (hours 4-6), the coefficients are not significantly different from each other ($p = 0.315$, two sided Wald test) in this specification. Yet, as we have seen before, accumulated over the whole day, *overall* performance is significantly larger in Gift_2. Finally, the coefficients of the demographic control variables and the number of workers per shift are all insignificant in Models 1 to 3.[11]

In the next step, we look for evidence in favor of reciprocal responses with respect to other measures. We first consider the overall productive working time achieved by the worker. As mentioned above, workers had some freedom during the project day to decide how to spend the time: They were free when to take a break, and after the end of the regular working day, they could decide to stay for another 20 minutes to help with an extra task. We calculate the net productive time of an employee by subtracting the time used for breaks (measured by the absence of the person from the shelf he was assigned

[11]We obtain very similar results if we estimate Models 1 to 3 with the natural logarithm of the hourly wage as the dependent variable to account for the skewness of the output distribution (see Figure A1 in the Appendix; models are not reported here).

to) from and, if the person decided to stay, adding the 20 minutes to her gross working time.

We find that gifts on average increase the productive working time. Compared to the 346.1 minutes achieved in the Baseline condition, average working time increases by 3.5% in Gift_1 (3.4% in Gift_2) to 358.2 minutes (357.9 minutes). Comparing productive working times across treatments yields weakly significant differences between Baseline and Gift_1 (p = 0.081, two-sided MWU-test) and between Baseline and Gift_2 (p = 0.074). At the same time, individual productive working times in Gift_1 and Gift_2 do not differ from each other (p = 0.850).[12]

Figure 2. Mean productive working time per treatment (in minutes)

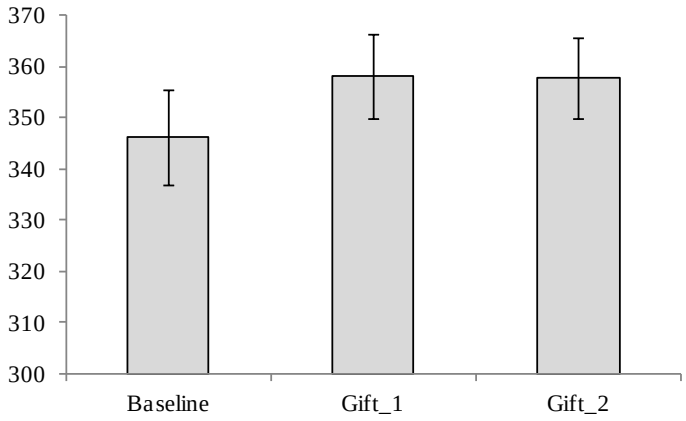


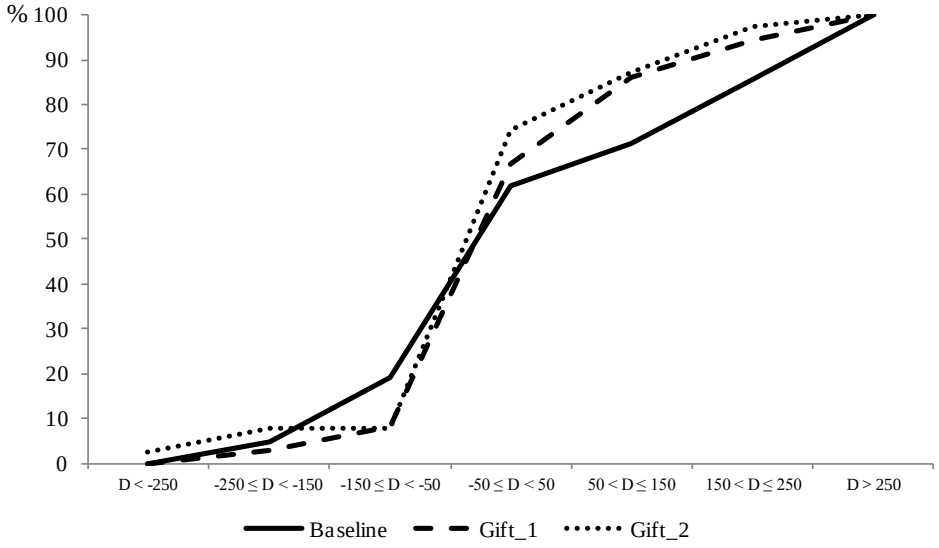
Figure 2 displays the mean productive working time in minutes, beyond 5 hours, separately for each experimental treatment. The lines indicate 95%-confidence intervals for the mean values.

Another category in which positive reciprocity might occur in our setting is employee honesty when reporting their performance for the bonus payment. Figure 3 plots cumulative distributions of the differences between reported and actual performance in the number of inserted tags (denoted by D).

[12] Altogether 4 observations had to be excluded, because the breaks of these workers were not tracked. On average, workers took a break for 43.1 minutes in Baseline, 39.1 minutes in Gift_1 and 34.6 minutes in Gift_2. Pairwise treatment comparisons with two-sided MWU tests yield p-values of p = 0.471 (Baseline versus Gift_1), p = 0.160 (Baseline versus Gift_2) and p = 0.451 (Gift_1 versus Gift_2). Moreover, 3 out of 21 workers in Baseline decided to stay for the unpaid extra task, compared to 16 (13) out of 38 (39) workers in the Gift_1 (Gift_2) treatment. Comparing treatments concerning the shares of workers who were willing to do the unpaid extra task with χ^2-tests yields p-values of p = 0.021 (Baseline versus Gift_1), p = 0.112 (Baseline versus Gift_2) and p = 0.427 (Gift_1 versus Gift_2).

The first observation from Figure 3 is that the number of reported tags typically differs quite substantially from the number of tags that the workers actually achieved. In only 37 of the 98 cases, reports match performance.[13] To distinguish a simple error from a deliberate overstatement of performance, we compare our treatments with respect to negative and positive deviations from the true value. First, we consider only cases where a worker reports her performance to be lower than what she in fact achieved. Mean values for the negative deviations from true performance (14 cases) account for -150, -125 and -150 in treatments Baseline, Gift_1 and Gift_2; the three treatments do not differ from each other as non-parametric tests indicate.[14]

Figure 3. Cumulative Distributions of deviations D between reported and actual performance across treatments (D = reported performance – actual performance)



[13] Comparing the shares of correct reports across treatments with two-sided χ^2-tests yields no significant differences. Two out of the 98 workers were not able to report their performance at the end of the project day. Moreover, the three treatments do not differ from each other with respect to the total distributions of deviations *D* from actual outputs, as all pairwise MWU tests yield significance levels of $p > 0.1$.

[14] Comparing the distributions of negative deviations of reported from true performance between treatments with two-sided MWU tests yields *p*-values of $p = 0.372$ (Baseline versus Gift_1), $p = 0.826$ (Baseline versus Gift_2), and $p = 0.912$ (Gift_1 versus Gift_2). In addition, the ratio of over-reported to under-reported output levels across treatments accounts for 2.13 in Baseline, 1.17 in Gift_1 and 0.76 in Gift_2, respectively. This suggests that workers not just became more attentive in keeping track of their output after receiving a gift. Instead, the fact that the ratio is substantially lower in the gift conditions than in Baseline is in line with the interpretation that accuracy is constant but workers become more honest after surprise wage increases.

Positive deviations, i.e. exaggerations of true performance, are observed much more frequently (45 cases) in the data than negative deviations.[15] Moreover, Figure 3 suggests that the absolute values tend to be larger in Baseline than in the gift treatments. Indeed the average overstatement of performance is with 319 tags in Baseline more than twice as high as in Gift_1 (146) and almost three times as high as in Gift_2 (114). Two-sided MWU tests confirm these differences: Comparing the distribution of deviations for positive values of D between Baseline and Gift_1 (Baseline and Gift_2) yields $p = 0.021$ ($p = 0.002$). Gift_1 and Gift_2 do not significantly differ from each other ($p = 0.109$). Hence, a further positive response to the wage increase in our setting is that it reduces the severity of worker cheating.

4. Discussion and Conclusion

We have conducted a field experiment where temporary workers carried out an inventory task in a naturally occurring labor environment. Our study is the first to investigate the impact of the timing of unexpected wage increases on the persistence of reciprocal worker responses.

First of all, in line with much of the previous literature our data provide clean evidence for positive reciprocity among the workforce: unexpected extra-payments resulted in an increase of the output (number of inserted tags), input (the time spent in a productive manner), and honesty (accurateness of self-reported performance).[16] Moreover, in contrast to Gneezy and List (2006), the performance increases in our setting, including the positive effect on voluntary extra-work and honesty, are realized in the second half and at

[15] Assuming that negative and positive deviations from the true performance occur with equal probability of 1/2, a two-sided Binomial test indicates that the large frequency of positive deviations in our sample is non-random ($p < 0.001$).

[16] We note that our results concerning worker honesty are related to findings from social psychology suggesting that employee misbehavior might be systematically related the perception of being treated fairly. In a famous field experiment, Greenberg (1990) found that workers became more prone to stealing from their employer after experiencing a temporary wage cut. We provide evidence that this mechanism might also work into the opposite direction: The experience of receiving a gift from the employer might reduce workers' inclination to behave opportunistically. Up to now, there is little work in economics exploring in which settings employee misbehavior might occur and through which channels it can be prevented (but see Gill, Prowse and Vlassopoulos 2012 and the references cited therein).

the very end of the working day, even when the wage increase has been announced right away in the morning. Our results thus show that, depending on the work environment, worker reciprocity may be a persistent phenomenon. A similar observation was made in the field experiment by Kube, Marèchal and Puppe (2012) where gifts also had a lasting positive effect on worker performance. Moreover, there is also evidence from psychology that positive affect caused, for example, by gifts, can increase subjects' persistency on a working task (Erez and Isen 2002).

A potential explanation for the delayed responses to gifts could be that workers required a certain time to become acquainted with the inventory task and therefore were not able to reciprocally increase performance at the beginning of the working day. Yet, if we estimate the models reported in Table 1 without the data from the first hour of the working day (not reported here), conclusions are very similar to the results reported above. Therefore, a more plausible explanation is that the delayed positive effect of wage increases in our setting seems to be driven by a higher perseverance of the workers in the second part of the day.

Second, our results indicate that the positively reciprocal increase in worker performance can be strengthened further when the wage increase is not given right away but split up over time. Overall performance is higher at the same labor costs when two increases are given sequentially as compared to one initial increase. This suggests that more frequently providing smaller wage increases may be preferable to less frequent but larger increases. We interpret this as evidence that the sensation of kindness triggers a positive response in addition to the purely material gain that comes with the gift. This observation is in line with the recent studies on non-monetary gifts by Kosfeld and Neckermann (2011) and Kube, Marèchal and Puppe (2012). Psychological research suggests that subjects might react in a more emotional way to unexplained events so that providing unexpected gifts repeatedly can become optimal from a principal's perspective to induce higher effort (see Wilson and Gilbert 2008). Moreover, a part of the additional positive response in our Gift_2 treatment might arise from the fact that workers do not consider their total wage when deciding about efforts, but look at the wage implemented in the particular working hour. As the result of such "narrow bracketing" (see, for example, Read, Loewenstein and

Rabin 1999) of effort choices to the current wage, workers might perceive the second half of the work day as a "12 Euro per hour" job in the Gift_1 treatment, compared to a "14 Euro per hour" job in the Gift_2 treatment, which might trigger higher performance in the latter treatment. Hence, it may be an interesting endeavor for future work to compare the importance of expected with unexpected wage increases of the same magnitude.

Interestingly, in our study, the timing of kindness significantly affects work intensity, but not worker honesty and overall working time. One might speculate that the relevant decisions on both, the voluntary extra-time spent in the library and the self-reported performance, are *explicit* and thus cognitively well-reflected, whereas work intensity is probably the result of a less deliberate and more affective choice. This observation is in line with the notion that the psychological mechanism behind the timing of positive reciprocity (fixing 'total kindness') at the workplace is best understood as resulting from what psychologists call "system 1", which is generally described as fast, automatic, emotionally charged, and requiring minimal cognitive resources (as opposed to "system 2", which is slow, deliberately controlled, analytical, affect free, and requires cognitive resources; see, e.g., Evans, 2008, and Kahneman, 2011).

Finally, we caution that if we consider the direct performance effects from an employer perspective, positive responses to wage increases may not be sufficient to make the gift profitable, which is in line with previous studies. The average wage cost per inserted tag accounts for 5.4 Euro cents in the Baseline treatment (the total wage of 56 Euros divided by the average performance of 1,031 tags) and is therefore still cheaper compared to 7.8 Euro cents in Gift_1 (84 Euros/1,071 tags) and 7.1 Euro cents in Gift_2 (84 Euros/1,189 tags). However, both the observation that reciprocity manifests over time and the positive impact on our other measures for reciprocity suggest that generous pay might become profitable when a longer time span is considered. Moreover, we have shown that a costless performance increase can be achieved simply by changing the timing in paying out the budget for wage increases. In our setting, this budget neutral performance increase accounts for some 11% compared to the case where the pay increase is paid out initially. This suggests that the timing of kindness is not only a promising research field, but also of considerable relevance for the engineering of incentive systems at the workplace.

Acknowledgments

We thank André Drost, Nadine Fritzler, Adrian Hopp, Margit Schön and Alfred Zimmermann from the Economics Library at the University of Cologne for their support of this study. We thank especially Andrea Fix and Johannes Wahlig for their excellent research assistance. Conference and seminar participants in Cologne, Maastricht, Paderborn and Zurich provided valuable comments and suggestions. Financial support of the German Research Foundation through the Leibniz-program and through the research unit "Design & Behavior" (FOR 1371) is gratefully acknowledged.

References

Akerlof, George. 1982. "Labor Contracts as a Partial Gift Exchange." Quarterly Journal of Economics, 97 (4): 543−569.

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

Chemin, Matthieu, Joost DeLaat, and André Kurmann. 2011. "Reciprocity in Labor Relations: Evidence from a Field Experiment with Long-Term Relationships." CIRPEE Working Paper No. 11-27.

Cohn, Alain, Ernst Fehr, and Lorenz Götte. 2012. "Fair Wages and Effort: Evidence from a Field Experiment." Working Paper.

Charness, Gary, and Peter J. Kuhn. 2011. "Lab Labor: What can Labor Economists Learn from the Lab?" In: Handbook of Labor Economics Vol. 4 Part A. ed. Orley Ashenfelter and David Card, 229−330. San Diego and Amsterdam: North-Holland.

Dufwenberg, Martin, and Georg Kirchsteiger. 2004. "A Theory of Sequential Reciprocity." Games and Economic Behavior, 47: 268−298.

Englmaier, Florian, and Stephen Leider. 2012. "Managerial Payoff and Gift Exchange in the Field." CESifo Working Paper No. 3707.

Erez, Amir, and Alice M. Isen. 2002. "The Influence of Positive Affect on the Components of Expectancy Motivation." Journal of Applied Psychology, 87(6): 1055−1067.

Evans, Jonathan St. B. T. 2008. "Dual-processing accounts of reasoning, judgment, and social cognition." Annual Review of Psychology, 59: 255-278.

Falk, Armin, and Urs Fischbacher. 2006. “A Theory of Reciprocity.” Games and Economic Behavior, 54: 293−315.

Fehr, Ernst, Lorenz Götte, and Christian Zehnder. 2009. “A Behavioral Account of the Labor Market: The Role of Fairness Concerns.” Annual Review of Economics, 1: 355−384.

Gill, David, Victoria Prowse, and Michael Vlassopoulos. 2012. “Cheating in the Workplace: An Experimental Study of the Impact of Bonuses and Productivity.” IZA Discussion Paper No. 6725.

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

Greenberg, Jerald. 1990. “Employee Theft as a Reaction to Underpayment Inequity: The Hidden Cost of Pay Cuts.” Journal of Applied Psychology, 75(5): 561-568.

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

Kahneman, Daniel. 2011. “Thinking, fast and slow.” Farrar, Straus & Giroux.

Kosfeld, Michael, and Susanne Neckermann. 2011. “Getting More Work for Nothing? Sysmbolic Awards and Working Performance.” American Economic Journal:Microeconomics, 3: 86−99.

Kube, Sebastian, Michel André Maréchal, and Clemens Puppe. 2012. “The Currency of reciprocity - gift-exchange in the workplace.” American Economic Review, 102(4): 1644−1662.

Kube, Sebastian, Michel André Maréchal, and Clemens Puppe. forthcoming. “Do wage cuts damage work morale? Evidence from a natural field experiment.” Journal of the European Economic Association.

Landry, Craig E., Andreas Lange, John A. List, Michael K. Price, and Nicholas G. Rupp. 2011. “Is there a ‘hidden cost of control’ in naturally-occurring markets? Evidence from a natural field experiment.” NBER Working Paper No. 17472.

Levine, David. 1998. “Modeling Altruism and Spitefulness in Experiments.” Review of Economic Dynamics, 1: 593-622.

Rabin, Matthew. 1993. “Incorporating Fairness into Game Theory and Economics.” American Economic Review, 83: 1281−1302.

Read, Daniel, George Loewenstein, and Matthew Rabin. 1999. “Choice Bracketing.” Journal of Risk and Uncertainty, 19(1−3): 171−197.

Wilson, Timothy D., and Daniel T. Gilbert. 2008. “Explaining away a model of affective adaptation.” Perspectives on Psychological Science, 3(5): 370−386.

Appendices for online publication only.

A1. Additional results

Table A1. Demographic data

Gender	Obs.	in %
Female	67	68.4
Male	31	31.6
Age	Obs.	in %
< 20 years	10	10.2
20 - 30 years	76	77.6
> 30 years	11	11.2
n/a	1	1.0

Table A2. Descriptive statistics (experimental treatments)

	Baseline	Gift_1	Gift_2
No. of workers	21	38	39
Avg. number of workers per shift	5.3	9.5	6.5
Avg. age in years	26.6	24.4	25.2
Share of female workers in %	66.7	71.1	66.7
Avg. gross working time in minutes	386.4	387.3	385.8

Table A3. Descriptive statistics (performance per treatment)

	Total Performance (in tags)
Baseline	1031.0
Gift_1	1071.1
Gift_2	1188.5
	Productive Working Time (in minutes)
Baseline	346.1
Gift_1	358.2
Gift_2	357.9
	Performance reports (in tags)
Baseline	1123.8
Gift_1	1112.1
Gift_2	1229.5

Figure A1. Distribution of output (rounded to 100s of tags) – in % of workers per treatment