

Worker reciprocity and the returns to training: evidence from a field experiment*

September 2018

Jan Sauermann[†]

Abstract

Workers' reciprocal inclinations can be used to motivate why firms invest in training. This paper studies whether individuals reciprocate training investments by testing whether reciprocal individuals have higher returns to training. I use a field experiment with random assignment to training combined with survey evidence on workers' reciprocity and find that reciprocal workers have higher returns to training. This finding, which is robust to control for observed personality traits and unobserved characteristics, suggests that individuals reciprocate the firm's human capital investment with higher effort after the training, which is in line with theoretical models on gift exchange in the workplace.

JEL Codes: J24, M53, D03

Keywords: training, reciprocity, performance, field experiment, call centres

*The author gratefully acknowledges financial support from the Jan Wallanders och Tom Hedelius Stiftelse for financial support (Grant number I2011-0345:1). The author thanks Sebastian Butschek, Leonie Gerhards, Christian Grund, Thomas Kittsteiner, Ilyana Kuziemko, Matthew Lindquist, Jens Mohrenweiser, Arjan Non, Gerard Pfann, and seminar participants at SOFI Stockholm, HECER Helsinki, the COPE workshop in Vienna, the Behavioral Economics Network Workshop in Stockholm, and RWTH Aachen for valuable comments and suggestions.

[†]Swedish Institute for Social Research (SOFI), Stockholm University, 10691 Stockholm, Sweden; telephone: +46 8 16 2605; fax: +46 8 15 46 70; e-mail: jan.sauermann@sofi.su.se; Center for Corporate Performance (CCP); Institute for the Study of Labor (IZA); Research Centre for Education and the Labour Market (ROA).

1 Introduction

In labor markets challenged by globalization and technological change, training has been of interest to both policy makers and firms for a long time. In the US, for example, 70% of employees engage in work-related training and education programs (OECD, 2017). The costs of training investments are often borne by firms, either by directly financing training courses, or indirectly by training the work force during work hours (Leuven and Oosterbeek, 1999; Caliendo et al., 2015). In the context of human capital theory, training investments are usually explained by market imperfections, such as compressed wages Acemoglu and Pischke (1998, 1999). An alternative argument to explain why firms invest in human capital even if there is risk of poaching is based on reciprocity, which has been shown to be important for labor market outcomes (for example Fehr et al., 1993; Rabin, 1993; Fehr and Gächter, 2000; Falk and Fischbacher, 2006; Rotemberg, 2006; Dohmen et al., 2009). Positive reciprocity, that is an individual’s positive response to actions of others even in the absence of expecting material gains, can induce employees to reciprocate their employer’s investments, for example by reducing turnover, by providing greater effort, or by reducing wage demands (Leuven et al., 2005). These behavioral responses to training can result in heterogeneity in the returns to training and can thus be crucial for a firm’s decision to invest in the human capital of a worker.

The aim of this paper is test whether positively reciprocal workers exert greater effort after participation in firm sponsored training, which should then result in higher posttraining performance. The hypothesis that reciprocity is linked to higher training returns is tested using data containing direct measures of worker performance, collected in an inhouse call center of a multinational mobile network operator in the Netherlands. In this firm, I exploit the random assignment of workers to a training program to causally estimate the returns to training (De Grip and Sauermann, 2012).¹ Arguing that the provision of this firm sponsored and fully paid week long training program can be perceived as a positive action of the firm towards the worker, I show that reciprocal workers have higher returns to the training pro-

¹The study of De Grip and Sauermann (2012) estimates the causal effect of training participation on performance, as well as spillover effects from trained workers to untrained peers. The personality measures used in this study have not been employed before.

gram, that is, perform better after the training than workers with low reciprocal inclinations do after participation in the training program. This finding suggests that individuals with reciprocal attitudes return training investments by providing greater effort after participation in the training program. I show that this effect is robust and cannot be explained by other personality measures, including the Big Five and Locus of Control. When including worker fixed effects to account for any remaining unobserved worker specific characteristics, I find that the estimate of interest is still significant yet smaller in size, suggesting that unobserved worker characteristics are potentially also important.

Reciprocal inclinations have been shown to be important in various labor market settings. The underlying idea is that individuals directly receive utility from rewarding (positive reciprocity) or punishing (negative reciprocity), even if there is no material pay-off from this action (Rabin, 1993; Falk and Fischbacher, 2006). A large number of studies using laboratory and field experiments shows that gifts from the employer induce workers to provide greater effort than they would without the gift.² Most of these studies analyze the effect of a random treatment, that is a monetary or nonmonetary gift, on the subjects' effort provision, which allows to establish a causal link between size and type of the gift and a subject's response to the gift.

Fewer studies have used measures of reciprocal attitudes derived through experiments or surveys in combination with worker and firm level outcomes. Barr and Serneels (2009), for example, use experimentally derived measures of workers' reciprocal attitudes and measures of firm performance, and find that firms with a more reciprocal workforce are more productive than firms with a less reciprocal workforce. Cohn et al. (2015) show that performance increases after an unexpected raise in hourly wages can only be observed for workers who displayed reciprocal attitudes in a choice experiment. In environments, in which experiments are not feasible or possible, the implementation of survey questions on reciprocity in large scale surveys allows to test the importance of reciprocity for labor market outcomes (Perugini et al., 2003). Using a large representative subject pool for Germany, for example, Dohmen

²See, for example, Fehr, Gächter, and Kirchsteiger (1997), Charness (2004), Gneezy and List (2006), Hennig-Schmidt, Rockenbach, and Sadrieh (2010), Kube, Marechal, and Puppe (2012), Becker, Messer, and Wolter (2013), Cohn, Fehr, and Götte (2015), and DellaVigna et al. (2016).

et al. (2009) show that workers' reciprocal attitudes are linked to higher wages, and to working longer hours.

The provision of training opportunities might be one way through which firms can induce reciprocal responses. These responses could be triggered either by perceiving the training as a gift, for example if the training is of general nature and can improve outside options, or by perceiving the training as receiving the manager's attention (Dur, 2009). Indeed, Leuven et al. (2005) show that firms have an incentive to provide socially optimal levels of training if firms take workers' reciprocal inclinations into account, thereby reducing the holdup problem. Using crosssectional data for the Netherlands, Leuven et al. (2005) provide evidence that positive reciprocal individuals are more likely to participate in employer financed training courses than individuals with low reciprocal inclinations. Although other, nonexperimental studies have explored reciprocal reactions to training courses, it is not clear to which degree these results are driven by the potential endogeneity of training participation.³ In this study, randomized participation in a training program allows to estimate both the causal effect of training participation on performance, but also the indirect effect, that is the interaction between the training effect and the worker's reciprocal attitudes.

This study contributes to the general understanding of firms' investments in employees' human capital, and provides evidence for an alternative rationale for the firms' investments in training. Higher returns to training can make training investments beneficial, even if there is risk of poaching. This result helps to explaining the positive correlation between training incidence and reciprocal attitudes (Leuven et al., 2005), and more generally to explaining the large training investments made by firms (Leuven and Oosterbeek, 1999; Caliendo et al., 2015).

³By comparing pre and post training measures of organizational support for a military organization, Mullen et al. (2006) find that trainees react positively to training participation. Although merely comparing outcomes before and after participation in training, they find that perceived organizational support increases with training participation. For a large multinational company based in Germany, Kampkötter and Marggraf (2015) find that participation in on the job training is related to lower turnover and lower absenteeism, suggesting that this could be one possible channel through which workers reciprocate the firm's investment. Using representative survey data for the Dutch public sector, Montizaan et al. (2015b) show that firms' training investments are positively correlated with postponed entry to retirement and find that this effect is driven by individuals with (positive) reciprocal attitudes.

This paper also contributes to the literature on the importance of attitudes and personality in the labor market. An increasing number of studies has shown that personality traits are, amongst others, predictive for educational choices, job search, employment, and earnings. Besides reciprocity, examples are the Big Five personality traits (Bowles et al., 2001; Mueller and Plug, 2006; Gensowski, 2018), and Locus of Control (Caliendo et al., 2015, 2016). With regards to training, Offerhaus (2013) finds no effect of the Big Five on training participation using a representative sample for Germany, whereas Caliendo et al. (2016) provide evidence that higher internal Locus of Control is related to higher participation rates in general training course.⁴ To the best of my knowledge, this is the first study containing both objective and repeated (panel) information on worker performance and survey evidence on personality measures.

This paper is structured as follows. Section 2 provides details on the data, the field experiment, and the measures of reciprocity. The main results as well as robustness checks are presented and discussed in Section 3. Section 4 provides additional results and evidence on mechanisms. Section 5 summarizes and concludes.

2 Data and setting

2.1 Workplace, tasks, and performance measurement

This study makes use of a field experiment, which was organized in the inhouse call center of a multinational mobile network operator in the Netherlands from week 45/2008 to week 24/2009.⁵ This call center acts as a service center for current and prospective customers. I focus on the largest department, which only serves private customers with fixed cell phone contracts. Call agents in this department have only one task, to answer incoming customer phone calls, for example if customers have problems, complaints, or questions. Agents in

⁴In addition to these paper analyzing specific attitudes and personality, there are a number of studies analyzing cognitive and noncognitive determinants of experimental and labor market outcomes. See, for example, Bowles et al. (2001), Heckman et al. (2006), Lindqvist and Vestman (2011), and Gill and Prowse (2016).

⁵See De Grip and Sauermann (2012) for a more detailed description of data and the field experiment and the institutional background.

this department are not involved in sales or customer acquisition. All agents take part in a training course when entering the department, which enables them to handle basic types of calls. Throughout their careers, agents regularly participate in training programs, which typically focus on learning about promotional campaigns, improving communication and IT skills, as well as learning how to handle more complex calls.

Call agents are organized in 10 teams, each of which is led by a team leader. During the observation period, 26 agents work for each team. The main purpose of being assigned to a team leader is that workers can be more efficiently supervised and monitored. There is neither team specialization, team production, nor team based incentives. Although the firm collects large amounts of data on the performance of individual call agents, these are not explicitly used to incentivize the call agents. Agents' performance can influence wages only through an annual appraisal interview with their team leader in which agents are evaluated for the past year. Based on the outcome of this appraisal interview, agents receive an annual bonus as well as an annual wage increase. Otherwise, wages are fixed for agents.⁶

The data contain weekly information on performance outcomes, with average handling time being the most important measure for monitoring agent performance used in the firm. Average handling time is defined as the average time an agent needs to handle a customer call and is available for each individual agent and each working week.⁷ I use the inverse of average handling time multiplied by 100, which allows us to interpret high y_{it} as high performance. This performance outcome is observed for each week and each worker throughout the sample period.

2.2 The field experiment

In the department analyzed, the management introduced a new training program with the explicit aim to decrease the average time needed for handling calls. The training program

⁶Appraisal interviews are conducted in February and March, that is before the training started. Bonuses and wage increases are paid out from July onwards, that is after the end of the observation period used in this study.

⁷This measure is also used in Liu and Batt (2007), Murthy et al. (2008), and De Grip and Sauermann (2012). Agents with shorter average handling time are evaluated as performing well. The main argument for this approach is that shorter calls are cheaper for the firm. There is only limited evidence that short calls are associated with lower quality (cf. De Grip and Sauermann, 2012).

was designed as a week long program, held in the call center’s inhouse training center over 5 consecutive days from Monday to Friday. Call agents were paid the full wage for the training week. Roughly half the training time was reserved for group discussions, in which the group discussed skills they lacked in their tasks, how these skills could be improved, and how the agents could provide more help to each other. During the other half of the training time, training coaches assisted the agents in handling customer calls.

The 32 week long sample period between week 45/2008 and week 24/2009 consists of three periods: a pretraining period (weeks 45/2008-09/2009), a training period (weeks 10/2009-14/2009), and a posttraining period (weeks 15/2009-24/2009). During the pretraining period, in week 50/2008, 74 out of the total 177 agents were selected for participation in the field experiment. This nonrandomly selected group was then randomly assigned to treatment and control groups.⁸ Treatment group agents were consecutively trained in small training groups during the training period; control group agents were also trained, but only after the posttraining period, which is not part of the sample period used in this paper.

Agents in the field experiment were randomly assigned to treatment and control groups by first assigning half of the teams to be trained during the training period. Due to a constraint that not more than 10 agents could be trained at once, teams were also randomly split up into separate training groups. Overall, 34 agents were trained during the training period. Control group agents ($N = 40$) were trained from week 25/2009 onwards, that is from the first week after the end of sample period used in this paper.

In this paper, I make use of survey information on reciprocity and other individual specific characteristics, which was held in the beginning of the training program. Due to partial survey nonresponse, the sample used in this study reduces to $N = 63$, consisting of 30 agents in the treatment group, and 33 agents in the control group. Column 4 of Table 1 shows that observable characteristics are balanced across treatment and control groups.⁹

⁸The firm deliberately chose to train more experienced workers to avoid losing their training investment due to high turnover among agents with low tenure. Participants in the field experiment are *not* a random sample of the overall population, but are, on average, slightly older and have longer tenure with the firm (see Table 1 for agents in the field experiment and Table A.1 in the Online appendix for the full sample of agents). This selection does not, however, violate the randomization of the field experiment, which is based on the random assignment of agents selected for the field experiment.

⁹Table A.1 in the Online Appendix further shows that there are no significant differences between individuals for the experimental sample ($N = 74$) and those who participated in the survey ($N = 63$).

Only for one of the factors of the Big Five measures, extraversion, the difference between treatment and control group is significant at the 10% significance level. Overall, this shows that not only socioeconomic characteristics, but also personality traits are balanced across treatment and control groups, and that the treatment can be considered to be exogenous, conditional on being assigned to participate in the field experiment.

Column 1 of Table 1 show descriptive statistics for the field experimental sample. The majority of agents are women (70%), and are, on average, 36 years old. Agents have an average tenure of 4.2 years, and work as parttimers with an average of 17 weekly working hours.

2.3 Why is this setting useful to study reciprocal behavior?

There are three reasons that make this field experiment useful to study whether workers reciprocate the firm’s training investments. First, the random assignment of agents into treatment and control groups provides a setting, which allows to estimate the causal effect of the training program on performance, as well as its interaction with the measure of reciprocity. An important feature of this study is that agents in both the treatment and control groups are eventually trained and were always aware that they would eventually be trained. The estimates are identified, however, only on data *before* the control group is trained. Put differently, the effect of reciprocity is identified through variation in the timing of the training investment. Before the agents in the control group are trained, agents in the treatment group are observed for several after their training participation. This setting has two important implications for the interpretation of the results. First, although agents were neither informed about this randomization nor about the evaluation, management communicated that, due to capacity constraints in the training center, the training would be rolled out over the course of several months. For this reason, it is less likely to expect that agents in the control group perceived the treatment group’s training as unfair. Second, the fact that agents in the control group could observe or anticipate the employer’s investment

even implies that the estimates are likely to be a lower bound of the true effect of the interaction between the training and reciprocity.¹⁰

Second, during the training program, for which the agents were paid fulltime, agents were taken out of the usual work environment and placed in a training centre separated from the work floor. Results from an internal posttraining evaluation questionnaire conducted by management suggests that agents positively experienced and enjoyed the training. Agents give an overall grade of 8.1 on a scale from 1 (worst) to 10 (best); 84% of the respondents strongly agree with the statement that “the training investment was worth the effort”.¹¹ This suggests that agents perceived the training was perceived as a positive action of the firm, rather than a chore.

Third, the theoretical motivation of this paper is that workers could perceive the training as a gift if it is of general nature. While the content of the training certainly contains some rather firm specific elements, such as knowledge related to the IT infrastructure of the firm, skills such as efficient communication with customers can be interpreted as general contents that could also be productive in other call centres, of which there are several in the direct vicinity. The effect of reciprocity should only be stronger if the training has more general contents.

2.4 Measuring reciprocity, personality, and cognitive skills

During the field experiment, call agents participated in a survey on “working in call centers,” which included statements on reciprocal attitudes, personality measures, questions to elicit cognitive test scores, and questions on socioeconomic information that is not part of the firm’s personnel data. Individual information on reciprocity was gathered using the questions developed and experimentally validated by Perugini et al. (2003), which are, amongst others, implemented in the German Socio-Economic Panel (GSOEP, Dohmen et al., 2009). In this survey, respondents were asked to rate the six questions on a 5-point Likert scale from 1

¹⁰Because all agents were informed about the training in week 50/2008, it is not possible to further analyze anticipation effects.

¹¹This internal survey is only partly based on individuals of the sample used in this study, and cannot be merged to the data used in this study.

(‘does not apply to me at all’) to 5 (‘applies perfectly to me’), three of which are used to calculate a measure of positive reciprocity.¹² Although reciprocity measures derived from a gift exchange experiment are more reliable (for example Cohn et al., 2015), survey evidence on reciprocity shows to yield results that are consistent with theoretical predictions (Dohmen et al., 2009; Montizaan et al., 2015a). Figure 1 shows the distribution of positive reciprocity for the estimation sample. The distribution is fairly similar to populationwide distributions of reciprocity (cf. Figure A.1 in the Online Appendix and Dohmen et al. (2009)).

In contrast to measures of reciprocity that are elicited in choice experiments, a shortcoming of survey measures is that these might reflect other individual specific characteristics. To further explore whether other individual specific characteristics can drive the findings, I employ additional measures of personality, a measure of loyalty, and test scores for cognitive skills that were all gathered in the same survey. For personality, I make use of the Big Five (conscientiousness, extraversion, agreeableness, openness to new experience, and neuroticism), Locus of Control, and negative reciprocity, each of which has been shown to have importance for labor market outcomes (Bowles et al., 2001; Mueller and Plug, 2006; Montizaan et al., 2015a; Offerhaus, 2013; Caliendo et al., 2015, 2016; Gensowski, 2018). The Big Five are measured by the 15-item questionnaire, which is implemented, amongst others, in the German Socio-Economic Panel (Gerlitz and Schupp, 2005). Each of the five factors are standardized with a mean of 0 and a standard deviation of 1. To measure Locus of Control, I employ a set of three questions, implemented from Borghans et al. (2006).¹³ As a third measure for personality, I make use of the measure of negative reciprocity, which is defined as the average of the respondent’s answers to items (2), (3), and (5) from the

¹²The questions are (1) “If someone does me a favor, I am prepared to return it,” (2) “If I suffer a serious wrong, I will take my revenge as soon as possible, no matter what the costs,” (3) “If somebody puts me in a difficult position, I will do the same to him/her,” (4) “I do my best to help somebody who helped me before,” (5) “If somebody offends me, I will offend him/her back,” and (6) “I am ready to undergo personal costs to help somebody who helped me before.” Standardized values of the answers to items (1), (4), and (6) are used to calculate an average measure of positive reciprocal behavior. All results in this study are qualitatively similar when using measures of reciprocity, which are derived from Principal Component Analysis.

¹³The questions on Locus of Control are part of a 15-item question, which involve the concepts of Locus of Control, Anxiety, Self-image, Self-confidence, and Imagination. The questions used for Locus of Control are: “Setbacks are usually due to mistakes people make”, “Most people do not realize well to which extent their live is determined by chance”, and “Whether I reach the goals I have in my life is not a matter of luck”. All questions could be answered on a 5-point Likert scale from 1 (‘does not apply to me at all’) to 5 (‘applies perfectly to me’).

questions in Footnote 12.¹⁴ To measure cognitive skills, six questions on arithmetical and logical problems are implemented (CentERdata, 2007). The answers could be either right or wrong; the cognitive test score is computed as the average of correct answers.¹⁵ The measure of loyalty is based on six questions, implemented from Mowday et al. (1979).¹⁶

3 Results

To understand the role of reciprocity in the workplace, Table 2 shows correlation coefficients between positive reciprocity on the one hand, and agent specific characteristics, personality traits, and work outcomes on the other hand. The table shows that reciprocity is not significantly related to the agents' characteristics as measured by gender, age, and education. Among personality traits and survey measures in Column 2, only the Big Five personality measures are correlated with the measure of reciprocity, which is in line with Dohmen et al. (2008), who use representative survey data for Germany.¹⁷ The other survey based measures (cognitive test score, negative reciprocity, Locus of Control, loyalty) are not significantly correlated with reciprocity.

Among worker outcomes, only the variable leaver, which is defined as a dummy for whether an individual quits within a six months time frame after the end of the experiment,

¹⁴Note that these the measures of positive and negative reciprocity are constructed such that positive and negative reciprocity are not correlated.

¹⁵These questions are adapted from CentERdata (2007). The questions are (1) "There are two groups of tourists of 60 persons each. If 3/4 of the first group and 2/3 of the second group take the bus to the museum, how much larger is the first group than the second group?" (2) "A rubber ball jumps up half the distance which it fell down. If the ball falls down from a 18m high roof, how many meters does the ball travel before it touches the ground for the third time?" (3) "Four girls are 100cm, 150cm, 125cm, and 75cm, respectively. Debbie is the tallest of the four girls. Karin is the shortest. Emmy is taller than Sara. How tall is Sara?" (4) "A ball and a hat cost 1.10 Euro in total. The ball costs 1 Euro more than the hat. How much does the hat cost? Please give your answer in cents." (5) "If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?" and (6) "In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?". Due to partial nonresponse, the measure of cognitive skills is only available for 57 agents.

¹⁶The questions are (1) "In general, I'm telling positive things about [the firm], even if other people criticize it", (2) "Management of [the firm] does not sufficiently motivate me to increase my performance" (negative), (3) "I'm happy that I chose [the firm] instead of a job at a different firm", (4) "I have no interest in the development of [the firm]" (negative), and (5) "Management of [the firm] know what they do".

¹⁷With the exception of the measure of neuroticism, the signs of the correlation coefficients between Big Five elements and the measure of reciprocity are the same as in Dohmen et al. (2008).

is correlated with the measure of reciprocity (Column 3 of Table 2).¹⁸ Albeit only weakly significant at the 10%-level, the interpretation of this correlation coefficient is in line with theory: reciprocal individuals might be more inclined to staying longer in the firm.

3.1 The effect of reciprocity on the returns to training

To analyze how reciprocal attitudes are related to the returns to training, the full panel structure of the field experiment (63 agents) is employed, in which performance is observed in each week an agent is working both before and after the training period from week 45/2008 to week 24/2009. Because agents are randomly assigned to participation in the training course, the causal effect of participation in the training program on worker productivity and its interaction with the measure of reciprocity can be estimated from the following ordinary least squares regression:

$$(1) \quad \log(y_{it}) = \alpha + \tau_1 d_{it} + \tau_2 rec_i + \tau_3 d_{it} \cdot rec_i + \beta_1 X_{it} + \beta_2 t_t + \beta_3 X_t + u_{it}$$

where y_{it} is the measure of productivity of worker i in week t , which is based on average handling time and for which high levels of y_{it} are interpreted as high performance. The dummy d_{it} equals one in each week after agent i 's training participation, and rec_i is the survey measure of positive reciprocity. Following De Grip and Sauermann (2012), I include control variables for several characteristics to account for remaining individual heterogeneity (X_{it}), such as working hours in week t , and whether an agent works during peak hours with high customer load in week t . To control for trends in aggregate performance and overall work load, I also include a linear time trend t_t , and the overall number of calls divided by total number of fulltime equivalent agents (X_t).¹⁹ The idiosyncratic error term u_{it} is clustered at the team level to account for the team level randomization (cf. Section 2.2 and Abadie et al. (2017)).

¹⁸Note that it is only possible to observe whether individuals left the department, but not whether individuals left the firm entirely. Because there are only limited possibilities to be promoted from call agent to management positions, it is likely that most leaving agents are leaving the firm entirely.

¹⁹Because surveys were not all conducted in the same week, fixed effects are included to control for possible survey week effects.

Table 3 shows the results of estimating Equation 1. Column 1 shows that the causal effect of training participation on performance is 0.0872, that is agents are, after participating in the training program, on average 8.7% more effective in their main task, which is answering customer calls (cf. De Grip and Sauermann, 2012).²⁰ Column 2 shows that an agent’s reciprocal attitudes are not significantly related to the outcome variable, $\log(y_{it})$, which is in line with the pairwise correlation shown in Table 2 and shows that reciprocal agents are not more productive *per se*.

The coefficient of main interest in this study, $\hat{\tau}_3$, indicates whether the treatment effect is heterogenous with respect to the workers’ level of reciprocity. Column 3 shows that the interaction effect between reciprocity and the randomized treatment is positive and significant: a one standard deviation difference in worker reciprocity is related to a 5.4%-point difference in the estimated returns to training. The higher treatment effect of reciprocal individuals suggests that these individuals return the training with higher effort provision after the training, possibly to return the training investment with a favorable action to their employer.²¹

3.2 Do other (un)observed characteristics matter?

In this study, the individual workers’ measure of reciprocity is gathered from a set of experimentally validated questions (Perugini et al., 2003), which has been shown to yield results that are consistent with theoretical predictions (Dohmen et al., 2009; Montizaan et al., 2015a). A potential concern, however, is that the measure of reciprocity is correlated with other, observable and unobservable characteristics, such as other personality traits or cognitive ability, which might ultimately drive what is attributed to reciprocity.

²⁰Table A.2 in the Online Appendix shows that the reported treatment effect only slightly differs for the sample of all agents who participated in the field experiment ($N = 74$), and those who also participated in the survey ($N = 63$). This is in line with the results reported in Table A.1, which show that none of the observable characteristics differs significantly between the two groups.

²¹An alternative interpretation of this result would be that, within an existing gift exchange between firm and agent, the provision of productivity enhancing training provides an additional margin at which agents can reciprocate the employer’s actions. Because reciprocal workers are not more productive *ex ante* (cf. Table 2), this second interpretation seems to be less likely.

To illustrate how other individual specific characteristics affect the estimation of the parameters of interest in this study, Equation 1 is augmented with an unobserved individual specific characteristic μ_i (for simplification, subscript t and additional covariates are suppressed)

$$(2) \quad \begin{aligned} \log(y_i) &= \alpha + \tau_1 d_i + \tau_2 rec_i + \tau_3 d_i \cdot rec_i + \gamma \mu_i + \epsilon_i \\ &= \alpha + \tau_1 d_i + \tau_2 rec_i + \tau_3 d_i \cdot rec_i + \epsilon'_i \end{aligned}$$

The new error term ϵ'_i consists of the unobserved term μ_i and an i.i.d. element ϵ_i . If rec_i and μ_i are not correlated, that is $Cov(\mu_i, rec_i) = 0$, estimating Equation 2 would result in unbiased estimates of τ_1 , τ_2 and τ_3 . If $Cov(\mu_i, rec_i) \neq 0$, the effect of participating in the training program ($\hat{\tau}_1$) would still be estimated without bias due to the experimental randomization. The parameter of interest, $\hat{\tau}_3$, however, would be biased in this case.

To deal with this potential endogeneity problem, the remainder of this subsection shows two sets of additional results: first, I show that explicitly controlling for a wide range of personality traits and cognitive test scores, which have shown to be important for labor market outcomes, does not alter the effect of interest, $\hat{\tau}_1$. Second, I employ worker fixed effects to delete *any* time invariant worker specific effect that might bias the effect of interest.

Other personality traits and test scores To learn more about which individual specific characteristics are possibly correlated with reciprocity, additional survey questions are used that were gathered in the same survey as the measure of reciprocity. These include the Big Five, Locus of Control, a measure of negative reciprocity to account for personality, a measure of cognitive test scores, and a measure of loyalty. Most of these individual specific characteristics have been shown to be important drivers of economic behavior in a number of different settings. In the context of on the job training, Locus of Control has been shown to be relevant for both participation in training as well as the returns (Offerhaus, 2013; Caliendo et al., 2016).²² As shown in Table 2 and in line with Dohmen et al. (2008), all Big

²²See, for example, Mueller and Plug (2006), Lindqvist and Vestman (2011) and Gensowski (2018) for the importance of personality, Lindqvist and Vestman (2011) and Gensowski (2018) for cognitive skills, and Dohmen et al. (2009) and Montizaan et al. (2015a) for negative reciprocity.

Five personality elements are highly correlated with the measure of reciprocity. In contrast, the cognitive test score, Locus of Control, negative reciprocity, and the measure of loyalty are not correlated with the measure of reciprocity.

To test whether these characteristics may drive the coefficient of main interest, that is the interaction effect between reciprocity and the treatment ($\hat{\tau}_3$), Equation 1 is estimated by separately including each of the individual specific characteristics. For the Big Five elements, Columns 1 to 6 of Table 4 show that neither of the Big Five elements nor all five elements altogether significantly affect performance. Columns 7 to 10 show the corresponding estimates when including measures of the cognitive test score, negative reciprocity, Locus of Control, and loyalty, respectively. Although none of these variables are significantly correlated with the measure of positive reciprocity, one could argue that these characteristics could pick up similar treatment effect heterogeneity. Only the measure of loyalty is significantly related to the outcome variable. The coefficient of interest, the interaction between reciprocity and the treatment, always remains significant and stable throughout most regressions (0.045-0.061).²³

Including worker fixed effects An explicit way of dealing with unobserved characteristics would be to estimate Equation 2 in a fixed effects framework. By applying within transformation, the individual specific characteristic μ_i as well as the main effect of reciprocity is removed

$$\begin{aligned}
 \log(y_{it}) - \overline{\log(y_i)} &= \tau_1(d_{it} - \bar{d}_i) + \tau_2(rec_{it} - \overline{rec_i}) + \tau_3(d_{it} \cdot rec_{it} - \overline{d_i \cdot rec_i}) \\
 &\quad + \gamma(\mu_i - \bar{\mu}_i) + (\epsilon_i - \bar{\epsilon}_i) \\
 (3) \qquad \qquad \qquad &= \tau_1(d_{it} - \bar{d}_i) + \tau_3(d_{it} \cdot rec_{it} - \overline{d_i \cdot rec_i}) + \epsilon'_i
 \end{aligned}$$

in which $\hat{\tau}_1$ would be the estimated (main) treatment effect. Even though the experimental design does not require individual fixed effects to estimate the causal effect of training,

²³Table A.3 in the Online Appendix shows corresponding estimation results when also interacting the added characteristics with the treatment. Only when including the Big Five element of neuroticism and its interaction effect in the regression, or when including all five elements in one regression, the interaction effect between reciprocity and the treatment becomes insignificant (with p values of 0.134 and 0.120, respectively). Without affecting the interaction between reciprocity and the treatment, the table also shows that individuals with higher levels of loyalty, extraversion, and higher cognitive test scores have higher returns to training.

fixed effects can serve the purpose to control for unobserved (fixed) characteristics that are potentially correlated with the measure of reciprocity.

As a first step to explore how controlling for individual characteristics affects the estimation results, the sample is split by individuals with below median and above median levels of reciprocity, which allows to compare the results with and without fixed effects. Columns 1 and 2 of Table 5 show that, for individuals with low levels of reciprocity, the point estimate of the returns to the training program are very similar, albeit more precisely estimated when controlling for individual fixed effects. For individuals with high levels of reciprocity, Column 3 shows that participation in the training program leads to a 15.5% higher performance after the training. Including individual fixed effects reduces this estimate by 5%-points to 10.6%. There are two explanations for this diverging pattern of individuals with low levels and high levels of reciprocity. First, one could argue that the sample of treatment and control groups are not perfectly balanced for the sample of individuals with high levels of reciprocity. This could lead to a biased estimate of the returns to training, which could then be reduced by including individual fixed effects. Replicating the t -test from Table 1 for the sample of individuals with high levels of reciprocity ($N = 32$), however, does not show any significant differences.²⁴ Second, unobserved individual specific characteristics could matter more for individuals with high levels of reciprocity, than they do for individuals with low levels of reciprocity. This could also be the case, for example, if reciprocity is measured with measurement error in the lower part of the reciprocity distribution.

Column 5 of Table 5 shows the corresponding estimates for the full sample including worker fixed effects. In this regression, the estimated treatment effect of participating in the training is 2.5 percentage points lower compared to the baseline effect in Column 3 of Table 3 (11.1%). The interaction effect between the treatment and reciprocity, however, is slightly less than a third of the size in Table 3 (1.6% vs. 5.4%, cf. Column 3 of Table 3). This suggests that individual-specific characteristics that are not measured by the variables included in this study, partly explain the treatment effect heterogeneity with respect to reciprocity.

Results show that reciprocal individuals do have higher returns to training, despite being partly explained by unobserved individual specific characteristics μ_i . As the most

²⁴The results are available upon request.

conservative estimate, the fixed effect estimate of 1.6% can serve as an lower bound, whereas the OLS estimate of 5.4% shown in Table 3 can serve as an upper estimate of the true interaction effect of reciprocity. This result also shows the importance of appropriately dealing with unobserved characteristics that might be correlated with the variable of interest.

4 Additional results

4.1 Alternative channels of reciprocal behavior

The results so far have provided evidence for treatment effect heterogeneity with respect to workers' degree of positive reciprocity. Although the descriptive analysis provided only limited evidence for significant correlation between reciprocity and other worker outcomes, reciprocal individuals might have alternative, possibly competing, channels through which training investments are returned. Examples for these competing channels are turnover and fluctuation, the number of hours of absence, for example due to sickness, and the number of hours worked by the agent.²⁵

First, an alternative way to reciprocate the training investments would be lower labor turnover, so that the firm would not lose its training investments from workers leaving the firm. Table 2 shows that the measure of reciprocity is significantly correlated with a dummy variable whether or not an employee left the department within six months after participating in the survey: reciprocal individuals are less likely to have left within this period. Although this result does not provide any causal evidence, it suggests that turnover is a possible channel through which training investments can be reciprocated. Because *all* agents are trained eventually in this setting, it is not possible to estimate the causal effect of the training program and its interaction with reciprocity on turnover.

²⁵There are two alternative arguments, which cannot be tested with this setting. First, Leuven et al. (2005) put forward the argument that employees could reciprocate training with lower wage demands. In the context of this field experiment, however, this hypothesis is not testable, because there is no information on wage bargaining. Second, reciprocal agents might have higher returns to training because they perceive the training in a way that the firm pays attention to them, or even interpret the training investment as a signal that they would not be laid off in the near future. These arguments could be tested in other settings, in which the experimenter could, for instance, randomly vary layoff probabilities.

Second, reciprocal individuals could return the training investments by working longer hours. For the firm, this would result in a better utilization of labor. Column 2 of Table 6 shows that when estimating Equation 1 with the number of working hours as an outcome variable, the interaction variable is not significant. This suggests that working hours is not a likely channel to reciprocate the training.

Third, reciprocity might have a negative effect on hours of absence. If reciprocal individuals would have lower degrees of sickness absence after the training, this result would suggest that individuals reciprocate the training (Kampkötter and Marggraf, 2015). To address this channel, Equation 1 is estimated with an outcome variable *hours of absence*. Column 4 of Table 6 shows that the estimated coefficient $\hat{\tau}_3$ has the expected negative sign and is significantly different from zero. This suggests that, after participating in the training programme, agents with a one standard deviation higher level of reciprocity have 0.027 fewer hours of absence per week. This estimate seems relatively low, but has to be related to a mean of 0.111 hours of absence per agent and week.

4.2 Are agents returning the favor of being selected?

The results shown in Section 3 are in line with an interpretation of reciprocal inclinations generating higher returns to training. An alternative strategic motive is that agents selected to be part of the training program feel that they are part of an exclusive group and therefore feel more motivated.²⁶ This motive, which might be correlated with reciprocity, could trigger similar reactions and might explain why the fixed-effects estimates reduce the interaction effect between reciprocity and the treatment.

Agents in both treatment and control groups were selected to be part of the training program, but were then randomly assigned to either group. That is, the effects in this paper are identified from exogenous variation in participating in the training course, conditional on being part of the training program. Due to the randomization in the experiment, tenure, which was the main criteria of being part of the training program, is balanced across treatment and control groups. To shed more light on whether tenure affects the interaction

²⁶I thank an anonymous referee for highlighting this point.

effect between the treatment and reciprocity, Table A.4 estimates Equation 1 augmented by agents' tenure and an interaction effect between the treatment and tenure. The estimates show that neither tenure nor its interaction effect with the treatment dummy has a significant effect, whereas the interaction effect with reciprocity remains unchanged. Although it is not possible to rule out that being selected into the training program *did* affect motivation and created increased performance, it does not drive the interaction effect between the treatment and reciprocity.

4.3 Reciprocating during or after the training?

An important question is whether reciprocal individuals put more effort into their training, that is whether it is that reciprocal individuals are better in utilizing the training, or are they exerting greater effort *after* participation in training. Whereas the former, more efficient human capital acquisition, should lead to a more permanent increase in skills and thus performance, the latter might cause only a transitory effect on performance that fades out.

To test this hypothesis, Figure 2 shows the treatment effect and the interaction with reciprocity, separately for each week after the training. The black line shows that the treatment effect of participation in the training program reaches a peak in the fifth week after training, and decreases substantially afterwards. Despite being small in size, the interaction effect between treatment and reciprocity does not follow this decrease (gray line). Towards the end of the observation period (week 11 after the training period), the point estimate of the interaction effect has almost the same magnitude as the main effect, suggesting that reciprocal individuals have a small yet more permanent effect than nonreciprocal individuals. Although far from conclusive, this could be explained by higher effort provision during the training, resulting in more efficient human capital acquisition.

5 Conclusions

Although there is empirical evidence that workers with positive reciprocal inclinations participate more often in firm sponsored training (Leuven et al., 2005), there is no evidence on

whether and how reciprocal individuals return the firm’s training investments. This paper provides previously undocumented evidence that individuals reciprocate the participation in firm sponsored training, namely, whether individuals with positive reciprocal attitudes have higher returns to training. To test this hypothesis, personnel data with panel information on worker performance is combined with random assignment to training courses, and direct measures of reciprocal attitudes. Compared to survey datasets, this setting allows to causally identify treatment effects of participation in the training program, as well as its interaction with survey measures of reciprocity. The results show that individuals with reciprocal inclinations have higher returns to training, and that this cannot be explained by other personality traits, such as the Big Five or Locus of Control. Although using worker fixed effects reduces the effect of interest, taken together, the results suggest that workers reciprocate the firm’s training investments by greater effort provision, which translates into higher performance.

The finding that individuals with reciprocal inclinations have higher returns to training has important implications for the worker firm relationship: first, the finding provides an alternative channel through which the gift exchange between workers and firms can operate. Second, it is potentially important to consider alternative personality measures to account for worker-specific heterogeneity by including worker fixed effects as one otherwise runs the risk of misattributing effects. Third, it suggests an alternative argument for why firms invest in human capital, even if there is risk of poaching. Workers’ reciprocal attitudes increase the incentive for firms to invest in training, and to positively select workers in the hiring process (cf. Englmaier et al., 2016).

References

- Abadie, Alberto, Susan Athey, Guido Imbens, and Jeffrey Wooldridge (2017): “When Should You Adjust Standard Errors for Clustering?” Tech. rep.
- Acemoglu, Daron and Jörn-Steffen Pischke (1998): “Why Do Firms Train? Theory and Evidence,” *Quarterly Journal of Economics*, 113(1): 79–119.

- (1999): “The Structure of Wages and Investment in General Training,” *Journal of Political Economy*, 107: 539–72.
- Barr, Abigail and Pieter Serneels (2009): “Reciprocity in the workplace,” *Experimental Economics*, 12(1): 99–112.
- Becker, Sascha O., Dolores Messer, and Stefan C. Wolter (2013): “A Gift is Not Always a Gift: Heterogeneity and Long-term Effects in a Gift Exchange Experiment,” *Economica*, 80: 345–71.
- Borghans, Lex, Bart Golsteyn, and Andries De Grip (2006): “Meer werken is meer leren: Determinanten van kennisontwikkeling,” Report, CINOP, ’s-Hertogenbosch.
- Bowles, Samuel, Herbert Gintis, and Melissa Osborne (2001): “The Determinants of Earnings: A Behavioral Approach,” *Journal of Economic Literature*, 39(4): 1137–76.
- Caliendo, Marco, Deborah A. Cobb-Clark, Helke Seitz, and Arne Uhlenborff (2016): “Locus of Control and Investment in Training,” IZA Discussion Paper 10406, Institute of Labor Economics (IZA).
- Caliendo, Marco, Deborah A. Cobb-Clark, and Arne Uhlenborff (2015): “Locus of Control and Job Search Strategies,” *Review of Economics and Statistics*, 97(1): 88–103.
- CentERdata (2007): “Codeboek Lerend individu 2007,” Codebook, Tilburg University, Tilburg.
- Charness, Gary (2004): “Attribution and reciprocity in an experimental labor market,” *Journal of Labor Economics*, 22(3): 665–88.
- Cohn, Alain, Ernst Fehr, and Lorenz Goette (2015): “Fair Wages and Effort Provision: Combining Evidence from a Choice Experiment and a Field Experiment,” *Management Science*, 61(8): 1777–94.
- De Grip, Andries and Jan Sauermann (2012): “The Effects of Training on Own and Co-Worker Productivity: Evidence from a Field Experiment,” *Economic Journal*, 122(560): 376–99.

- DellaVigna, Stefano, John A. List, Ulrike Malmendier, and Gautam Rao (2016): “Estimating Social Preferences and Gift Exchange at Work,” Working Paper 22043, National Bureau of Economic Research.
- Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde (2008): “Representative trust and reciprocity: Prevalence and determinants,” *Economic Inquiry*, 46(1): 84–90.
- (2009): “Homo Reciprocans: Survey Evidence on Behavioral Outcomes,” *Economic Journal*, 119(536): 592–612.
- Dur, Robert (2009): “Gift Exchange In The Workplace: Money Or Attention?” *Journal of the European Economic Association*, 7(23): 550–60.
- Englmaier, Florian, Thomas Kolaska, and Stephen Leider (2016): “Reciprocity in Organizations: Evidence from the UK,” *CESifo Economic Studies*, 62(3): 522–46.
- Falk, Armin and Urs Fischbacher (2006): “A theory of reciprocity,” *Games and Economic Behavior*, 54(2): 293–315.
- Fehr, Ernst and Simon Gächter (2000): “Fairness and Retaliation: The Economics of Reciprocity,” *Journal of Economic Perspectives*, 14(3): 159–81.
- Fehr, Ernst, Simon Gächter, and Georg Kirchsteiger (1997): “Reciprocity as a Contract Enforcement Device: Experimental Evidence,” *Econometrica*, 65(4): 833–60.
- Fehr, Ernst, Georg Kirchsteiger, and Arno Riedl (1993): “Does Fairness Prevent Market Clearing? An Experimental Investigation,” *Quarterly Journal of Economics*, 108(2): 437–59.
- Gensowski, Miriam (2018): “Personality, IQ, and Lifetime Earnings,” *Labour Economics*, 51: 170–83.
- Gerlitz, Jean-Yves and Jürgen Schupp (2005): “Zur Erhebung der Big-Five-basierten Persönlichkeitsmerkmale im SOEP,” Research Notes 4, German Institute for Economic Research (DIW), Berlin.

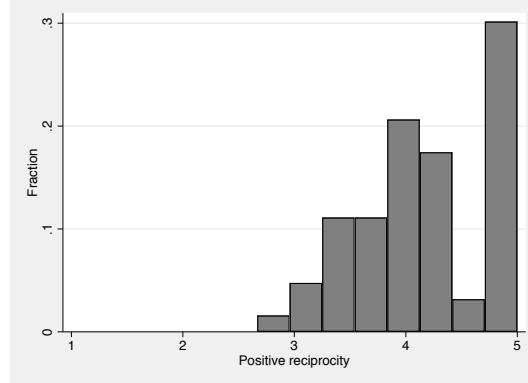
- Gill, David and Victoria Prowse (2016): “Cognitive Ability, Character Skills, and Learning to Play Equilibrium: A Level-k Analysis,” *Journal of Political Economy*, 124(6): 1619–1676.
- Gneezy, Uri and John A List (2006): “Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments,” *Econometrica*, 74(5): 1365–84.
- Heckman, James J., Jora Stixrud, and Sergio Urzua (2006): “The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior,” *Journal of Labor Economics*, 24(3): 411–82.
- 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): 817–37.
- Kampkötter, Patrick and Kathrin Marggraf (2015): “Do Employees Reciprocate to Intra-Firm Trainings? An Analysis of Absenteeism and Turnover Rates,” *International Journal of Human Resource Management*, 26(22): 2888–907.
- Kube, Sebastian, Michel André Maréchal, and Clemens Puppe (2012): “The Currency of Reciprocity: Gift Exchange in the Workplace,” *American Economic Review*, 102(4): 1644–62.
- Leuven, Edwin and Hessel Oosterbeek (1999): “Demand and supply of work-related training: Evidence from four countries,” in *Research in Labor Economics*, 18, 303–30.
- Leuven, Edwin, Hessel Oosterbeek, Randolph Sloof, and Chris Van Klaveren (2005): “Worker reciprocity and employer investment in training,” *Economica*, 72(285): 137–49.
- Lindqvist, Erik and Roine Vestman (2011): “The Labor Market Returns to Cognitive and Noncognitive Ability: Evidence from the Swedish Enlistment,” *American Economic Journal: Applied Economics*, 3(1): 101–28.
- Liu, Xiangmin and Rosemary Batt (2007): “The Economic Pay-Offs to Informal Training: Evidence from Routine Service Work,” *Industrial and Labor Relations Review*, 61(1): 75–89.

- Montizaan, Raymond, Frank Cörvers, Andries De Grip, and Thomas Dohmen (2015a): “The impact of negatively reciprocal inclinations on worker behavior: Evidence from a retrenchment of pension rights,” *Management Science*, 62(3): 668–81.
- Montizaan, Raymond, Andries de Grip, and Didier Fouarge (2015b): “Training Access, Reciprocity, and Expected Retirement Age,” IZA Discussion Paper 8862, Institute of Labor Economics (IZA).
- Mowday, Richard T, Richard M Steers, and Lyman W Porter (1979): “The measurement of organizational commitment,” *Journal of Vocational Behavior*, 14(2): 224–47.
- Mueller, Gerrit and Erik Plug (2006): “Estimating the Effect of Personality on Male and Female Earnings,” *Industrial and Labor Relations Review*, 60(1): 3–22.
- Mullen, Torrey Rieser, Christy Kroustalis, Adam W. Meade, and Eric. A. Surface (2006): “Assessing change in perceived organizational support due to training,” Conference paper, Paper presented at the 21st annual meeting of the Society for Industrial/Organizational Psychology, Dallas, TX.
- Murthy, Nagesh N., Goutam N. Challagalla, Leslie H. Vincent, and Tasadduq A. Shervani (2008): “The Impact of Simulation Training on Call Center Agent Performance: A Field-Based Investigation,” *Management Science*, 54(2): 384–99.
- OECD (2017): *OECD Skills Outlook 2017: Skills and Global Value Chains*, Paris: OECD Publishing.
- Offerhaus, Judith (2013): “The type to train? Impacts of personality characteristics on further training participation,” SOEPpapers on Multidisciplinary Panel Data Research 531, DIW Berlin, The German Socio-Economic Panel (SOEP), Berlin.
- Perugini, Marco, Marcello Gallucci, Fabio Presaghi, and Anna Paola Ercolani (2003): “The Personal Norm of Reciprocity,” *European Journal of Personality*, 17: 251–83.
- Rabin, Matthew (1993): “Incorporating Fairness into Game Theory and Economics,” *American Economic Review*, 83(5): 1281–302.

Rotemberg, Julio J. (2006): “Altruism, reciprocity and cooperation in the workplace,” in *Applications*, ed. by Serge-Christophe Kolm and Jean Mercier Ythier, Elsevier, vol. 2 of *Handbook of the Economics of Giving, Altruism and Reciprocity*, chap. 21, 1371–407.

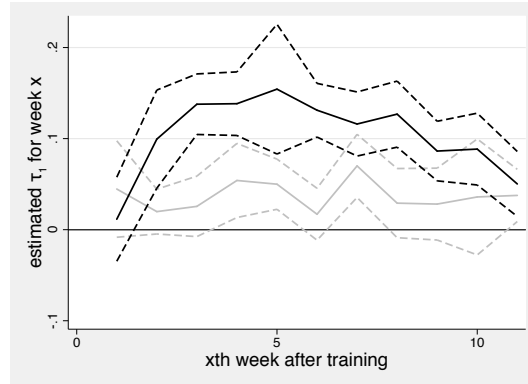
Figures

Figure 1: Distribution of reciprocity



Note: The figure shows the histogram of positive reciprocity for the estimation sample ($N = 63$). The underlying questions (see Section 2.4) could be answered on a scale from 1 ('does not apply to me at all') to 5 ('applies perfectly to me').

Figure 2: Treatment and interaction effects on performance over time



Note: This figure shows the estimated treatment effect on performance for each week after the training (solid black line) and the corresponding 95% confidence interval (dashed black lines). The gray lines shows the estimated interaction effect between treatment dummy and the measure of reciprocity (solid gray line) and the corresponding 95% confidence intervals (dashed gray lines). The estimates are based on a regression controlling for the level of reciprocity, working hours, the share of peak hours, calls per FTE, a linear time trend, and for the week the survey was taken (cf. the regression shown in Column 6 of Table 3). Week 0 denotes the training week.

Tables

Table 1: Descriptive statistics

Sample	(1) All	(2) Treat. group	(3) Control group	(4) Diff (4)–(5)
Gender (1=male)	0.3016 (0.4626)	0.3667 (0.4901)	0.2424 (0.4352)	0.1242 (1.0658)
Age	36.3504 (11.2356)	34.9622 (10.3689)	37.6125 (11.9876)	-2.6503 (-0.9341)
Tenure (in years)	4.1990 (3.9701)	4.4147 (3.9226)	4.0029 (4.0633)	0.4118 (0.4084)
High education	0.3103 (0.4667)	0.3571 (0.4880)	0.2667 (0.4498)	0.0905 (0.7348)
Performance	0.3629 (0.0837)	0.3673 (0.0727)	0.3589 (0.0935)	0.0085 (0.3985)
Number of Calls	196.4127 (115.7966)	187.0000 (119.1478)	204.9697 (113.8235)	-17.9697 (-0.6121)
Working hours	16.6508 (8.4799)	15.9667 (9.3199)	17.2727 (7.7309)	-1.3061 (-0.6074)
Share Peak-Hours	0.5328 (0.1935)	0.5386 (0.1817)	0.5276 (0.2063)	0.0110 (0.2242)
Absenteeism	0.1111 (0.3168)	0.1000 (0.3051)	0.1212 (0.3314)	-0.0212 (-0.2634)
Training incidence	0.1905 (0.3958)	0.1000 (0.3051)	0.2727 (0.4523)	-0.1727 (-1.7588)
Leaver	0.5238 (0.5034)	0.5667 (0.5040)	0.4848 (0.5075)	0.0818 (0.6412)
Positive reciprocity	4.2011 (0.6627)	4.1222 (0.7349)	4.2727 (0.5919)	-0.1505 (-0.8989)
Conscientiousness	12.6667 (1.4142)	12.5667 (1.3566)	12.7576 (1.4797)	-0.1909 (-0.5320)
Extraversion	12.0952 (1.8554)	12.6667 (1.5388)	11.5758 (1.9848)	1.0909* (2.4204)
Agreeableness	12.7619 (1.6821)	12.5667 (1.8696)	12.9394 (1.4987)	-0.3727 (-0.8767)
Openness to experience	10.3810 (1.9380)	10.5333 (1.9250)	10.2424 (1.9690)	0.2909 (0.5919)
Neuroticism	7.2857 (2.3721)	7.2667 (2.4344)	7.3030 (2.3517)	-0.0364 (-0.0603)
Cognitive test score	0.4561 (0.2450)	0.4770 (0.2736)	0.4345 (0.2144)	0.0425 (0.6511)
Negative reciprocity	2.4815 (0.8875)	2.5556 (0.8502)	2.4141 (0.9281)	0.1414 (0.6285)
Locus of Control	8.7143 (1.5390)	8.8667 (1.6132)	8.5758 (1.4797)	0.2909 (0.7466)
Loyalty	1.1270 (0.7042)	1.1467 (0.6453)	1.1091 (0.7634)	0.0376 (0.2099)
Observations	63	30	33	63

Note: Standard deviations are in parentheses in Columns 1 to 3; *t*-statistics are in parentheses in Column 4.

Table 2: Correlates of positive reciprocity

(1) Worker characteristics		(2) Personality measures		(3) Worker outcomes	
Gender (1=male)	0.1104	<i>Big Five</i>		Performance	-0.0295
Age	-0.0124	Conscientiousness	0.4020***	Tenure	0.1515
High skilled education	0.0929	Extraversion	0.0898	Working hours	0.0109
		Agreeableness	0.2404*	Share peak hours	0.0192
		Openness	0.2487**	Absenteeism	-0.0841
		Neuroticism	-0.2997**	Training incidence	-0.0466
		Cognitive test score	0.1907	Leaver	-0.2127*
		Negative reciprocity	-0.0115		
		Locus of Control	-0.0763		
		Loyalty	-0.1450		

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Figures in this table show the pairwise Pearson's correlation coefficients of workers' characteristics with the measure of positive reciprocity. The figures are based on the estimation sample and contains one observation per agent ($N = 63$, cf. Column 1 of Table 1). All time varying variables in Column 3 are averaged over the time between the start of the sample (45/2008) and the start of the field experiment (09/2009). Worker outcomes are defined as: performance (inverse of average handling time), tenure in years, weekly working hours, share of working hours during peak hours, absenteeism (share of weeks an agent reported being sick), training incidence (share of weeks in which an agent received training), leaver is defined whether a worker left within 6 months after the end of the experiment.

Table 3: The returns to training and interaction with reciprocity

	(1)	(2)	(3)
Treatment dummy	0.0872*** (0.0227)	0.1065*** (0.0161)	0.1114*** (0.0151)
Reciprocity		0.0137 (0.0255)	0.0022 (0.0268)
Treatment \times reciprocity			0.0540** (0.0224)
Calls per FTE	0.0001 (0.0001)	0.0001** (0.0001)	0.0001** (0.0001)
Working hours	0.0018 (0.0022)	0.0010 (0.0012)	0.0011 (0.0012)
Share peak hours	-0.2736*** (0.0813)	-0.2783*** (0.0659)	-0.2872*** (0.0594)
Time trend	0.0020 (0.0013)	0.0015* (0.0008)	0.0014* (0.0008)
Constant	-1.0833*** (0.1075)	-0.9946*** (0.0706)	-0.9879*** (0.0665)
Observations	1,672	1,672	1,672
Number of agents	63	63	63
R-squared	0.0768	0.2810	0.2895

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable: $\log(y_{it})$. All regressions also include fixed effects for the week in which agents participated in the survey. Standard errors are clustered at the team level.

Table 4: The effect of reciprocity and worker specific characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treatment dummy	0.1112*** (0.0157)	0.1099*** (0.0147)	0.1131*** (0.0147)	0.1117*** (0.0169)	0.1102*** (0.0141)	0.1182*** (0.0159)	0.1137*** (0.0157)	0.1155*** (0.0156)	0.1114*** (0.0147)	0.1118*** (0.0193)
Reciprocity	0.0043 (0.0253)	0.0018 (0.0269)	0.0004 (0.0290)	0.0081 (0.0279)	-0.0073 (0.0272)	-0.0008 (0.0313)	0.0056 (0.0250)	0.0022 (0.0268)	0.0027 (0.0278)	0.0070 (0.0208)
Treatment \times reciprocity	0.0544** (0.0234)	0.0545** (0.0218)	0.0537** (0.0217)	0.0500** (0.0223)	0.0512** (0.0207)	0.0451** (0.0185)	0.0612** (0.0243)	0.0560** (0.0223)	0.0539** (0.0225)	0.0460** (0.0185)
Conscientiousness	-0.0058 (0.0262)					-0.0042 (0.0285)				
Extraversion		0.0033 (0.0110)				-0.0151 (0.0097)				
Agreeableness			0.0082 (0.0150)			0.0053 (0.0176)				
Openness				-0.0432 (0.0307)		-0.0374 (0.0314)				
Neuroticism					-0.0295 (0.0174)	-0.0318 (0.0206)				
Cognitive test score							-0.0742 (0.0533)			
Negative reciprocity								-0.0161 (0.0163)		
Locus of Control									0.0052 (0.0172)	
Loyalty										0.0543** (0.0184)
Constant	-0.9818*** (0.0723)	-0.9926*** (0.0713)	-0.9853*** (0.0651)	-0.9955*** (0.0821)	-0.9991*** (0.0764)	-0.9791*** (0.0700)	-0.9334*** (0.0714)	-0.9759*** (0.0662)	-0.9833*** (0.0799)	-0.9509*** (0.0693)
Observations	1,672	1,672	1,672	1,672	1,672	1,672	1,531	1,672	1,672	1,672
Number of agents	63	63	63	63	63	63	57	63	63	63
R-squared	0.2897	0.2897	0.2904	0.3085	0.3034	0.3208	0.2355	0.2934	0.2899	0.3329

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable: $\log(y_{it})$. All regressions also include a constant, working hours, share in peak hours, calls per fulltime equivalents (FTE), a linear time trend, and fixed effects for the week in which agents participated in the survey. Standard errors are clustered at the team level. The F-Statistic and corresponding p -value corresponds to a t -test of the main and interaction effects of all added personality traits and cognitive test scores. Due to partial nonresponse, the sample used in Column 7 is based on 57 agents instead of 63 agents.

Table 5: The returns to training by low and high reciprocal individuals

Sample	(1) Below median reciprocity	(2) Above median reciprocity	(3) Below median reciprocity	(4) Above median reciprocity	(5) All
Treatment dummy	0.0632 (0.0483)	0.0608** (0.0208)	0.1551*** (0.0356)	0.1062*** (0.0201)	0.0863*** (0.0134)
Treatment \times Reciprocity					0.0155** (0.0058)
Constant	-0.9671*** (0.1165)	-0.9129*** (0.0825)	-0.9298*** (0.1804)	-1.0869*** (0.1057)	-0.9765*** (0.0627)
Observations	846	846	826	826	1,672
Number of agents (TG / CG)	31 (16/15)	31 (16/15)	32 (14/18)	32 (14/18)	63 (30/33)
R-squared	0.2343	0.6481	0.3953	0.6125	0.6284
Individual FE	No	Yes	No	Yes	Yes

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable: $\log(y_{it})$. All regressions also include working hours, share in peak hours, calls per fulltime equivalents (FTE), a linear time trend, and fixed effects for the week in which agents participated in the survey. Standard errors are clustered at the team level.

Table 6: The returns to training on sickness absence and working hours

Outcome variable	(1) Working hours	(2) Working hours	(3) Hours of absence	(4) Hours of absence
Treatment dummy	-1.1640* (0.6371)	-1.2687* (0.6398)	0.0178*** (0.0051)	0.0153** (0.0058)
Reciprocity	-0.0767 (0.9832)	0.1765 (1.0407)	-0.0162 (0.0145)	-0.0104 (0.0162)
Treatment \times reciprocity		-1.1909 (0.9924)		-0.0274* (0.0150)
Constant	22.0869*** (3.7114)	21.8534*** (3.6416)	0.2775** (0.0890)	0.2740** (0.0878)
Observations	1,672	1,672	1,672	1,672
Number of agents	63	63	63	63
R-squared	0.2185	0.2215	0.1230	0.1256

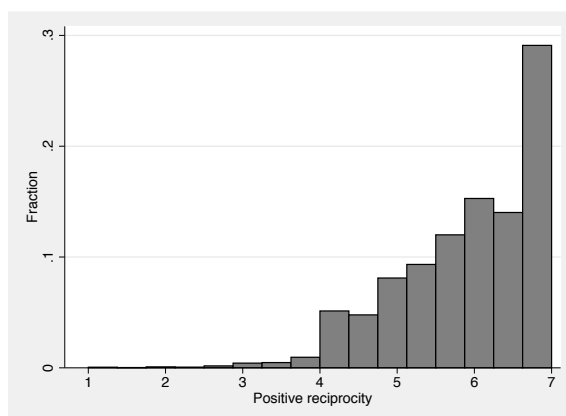
Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable: number of working hours (Columns 1 and 2), and the hours of absence (Columns 3 and 4). All regressions also include calls per FTE, a linear time trend, the number of working hours, the share of peak hours, and fixed effects for the week in which agents participated in the survey. Standard errors are clustered at the team level.

Online appendix to

“Worker reciprocity and the returns to training: evidence from a field experiment”

Jan Sauermann (Friday 28th September, 2018)

Figure A.1: Distribution of reciprocity in the German Socio-Economic Panel



Note: The figure shows the histogram of positive reciprocity taken from the German Socio-Economic Panel, Wave 2005 ($N = 16,056$). The underlying questions (see Section 2.4) could be answered on a scale from 1 ('does not apply to me at all') to 7 ('applies perfectly to me'). See Dohmen et al. (2009).

Table A.1: Descriptive statistics of agents in the call centre and the field experiment

Sample	(1)	(2)	(3)	(4)
	All agents	All	Field experiment With survey	Diff (2)–(3)
Gender (1=male)	0.2903 (0.4554)	0.3243 (0.4713)	0.3016 (0.4626)	0.1074 (0.9761)
Age	32.9498 (11.3471)	35.1918 (10.9330)	36.3504 (11.2356)	-2.0761 (-0.8122)
Tenure (in years)	2.7784 (3.6260)	3.9101 (3.8813)	4.1990 (3.9701)	0.2982 (0.3273)
Performance	0.3413 (0.1035)	0.3588 (0.0837)	0.3629 (0.0837)	0.0013 (0.0664)
Number of calls	200.9484 (99.8692)	194.0541 (111.9994)	196.4127 (115.7966)	-23.6059 (-0.9024)
Working hours	18.6516 (8.0078)	16.5946 (8.1508)	16.6508 (8.4799)	-1.6441 (-0.8632)
Share peak hours	0.5545 (0.1710)	0.5409 (0.1941)	0.5328 (0.1935)	0.0266 (0.5853)
Absenteeism	0.0710 (0.2576)	0.1216 (0.3291)	0.1111 (0.3168)	-0.0074 (-0.0951)
Training incidence	0.2065 (0.4061)	0.1892 (0.3943)	0.1905 (0.3958)	-0.1324 (-1.4497)
Leaver	0.6387 (0.4819)	0.5676 (0.4988)	0.5238 (0.5034)	0.0382 (0.3266)
Observations	155	74	63	

Note: The sample in Column 1 is defined as all agents working in the department during the observation period. Column 2 contains all agents who participated in the field experiment. Column 3 contains all agents who participated in the field experiment and who participated in the survey. Column 4 shows the difference between agents with survey participation and those without. Parentheses in Columns 1 to 3 contain standard deviations; parentheses in Column 4 contain t -statistics.

Table A.2: The returns to training and sample definition

Sample	(1)	(2)	(3)	(4)
	Field experiment		Surveyed agents	
Treatment dummy	0.0836*** (0.0217)	0.0882*** (0.0206)	0.0872*** (0.0227)	0.0860*** (0.0209)
Calls per FTE	0.0001 (0.0001)	0.0001 (0.0001)	0.0001 (0.0001)	0.0001 (0.0001)
Working hours	0.0015 (0.0022)	-0.0015 (0.0010)	0.0018 (0.0022)	-0.0017 (0.0011)
Share Peak-Hours	-0.2478*** (0.0588)	-0.3416*** (0.0749)	-0.2736*** (0.0813)	-0.3517*** (0.0794)
Time trend	0.0024 (0.0013)	0.0023** (0.0009)	0.0020 (0.0013)	0.0024*** (0.0009)
Constant	-1.1281*** (0.1027)	-1.0000*** (0.0888)	-1.0833*** (0.1075)	-0.9827*** (0.0931)
Observations	1,858	1,859	1,672	1,673
Number of agents	74	74	63	63
R-squared	0.0764	0.6165	0.0768	0.6274
Individual FE	No	Yes	No	Yes

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable: $\log(y_{it})$. The sample in Columns 1 and 2 is defined as all agents who participated in the field experiment ($N = 74$, cf. De Grip and Sauermann, 2012). The sample in Columns 3 and 4 is the estimation sample used in this study ($N = 63$), and is defined as agents from the field experiment who also participated in the survey. All regressions also include fixed effects for the week in which agents participated in the survey. Standard errors are clustered at the team level.

Table A.3: Reciprocity and other worker specific characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treatment dummy	0.1097*** (0.0164)	0.0866*** (0.0109)	0.1108*** (0.0149)	0.1137*** (0.0200)	0.1092*** (0.0135)	0.1006*** (0.0175)	0.0441 (0.0258)	0.1185*** (0.0180)	0.1123*** (0.0170)	0.1109*** (0.0189)
Reciprocity	0.0048 (0.0249)	0.0019 (0.0266)	0.0013 (0.0293)	0.0074 (0.0277)	-0.0063 (0.0269)	-0.0024 (0.0295)	0.0071 (0.0249)	0.0024 (0.0271)	0.0027 (0.0278)	0.0060 (0.0221)
Treatment \times reciprocity	0.0506** (0.0172)	0.0555** (0.0216)	0.0498* (0.0225)	0.0533* (0.0248)	0.0432 (0.0265)	0.0575 (0.0338)	0.0471* (0.0221)	0.0574** (0.0230)	0.0537** (0.0227)	0.0431** (0.0183)
Conscientiousness	-0.0075 (0.0297)					0.0010 (0.0332)				
Treatment \times Conscientiousness	0.0107 (0.0261)					-0.0204 (0.0312)				
Extraversion		-0.0022 (0.0102)				-0.0204** (0.0081)				
Treatment \times Extraversion		0.0552** (0.0229)				0.0496*** (0.0138)				
Agreeableness			0.0047 (0.0151)			0.0014 (0.0187)				
Treatment \times Agreeableness			0.0160 (0.0146)			0.0117 (0.0133)				
Openness to experience				-0.0394 (0.0291)		-0.0333 (0.0286)				
Treatment \times Openness				-0.0205 (0.0511)		-0.0174 (0.0528)				
Neuroticism					-0.0261* (0.0142)	-0.0327* (0.0159)				
Treatment \times Neuroticism					-0.0186 (0.0315)	0.0101 (0.0329)				
Cognitive test score							-0.0987* (0.0538)			
Treatment \times cog. test							0.1542*** (0.0349)			
Negative reciprocity								-0.0123 (0.0096)		
Treatment \times neg. reciprocity								-0.0220 (0.0322)		
Locus of Control									0.0064 (0.0183)	
Treatment \times Locus of Control									-0.0058 (0.0206)	
Loyalty										0.0476** (0.0182)
Treatment \times Loyalty										0.0392* (0.0214)
F-stat controls=0	0.0955 0.910	2.995 0.0956	0.651 0.542	1.050 0.385	1.723 0.228	758730 0.000	13.38 0.00149	1.161 0.352	0.0749 0.928	42.83 0.0001
P-value										
Observations	1,672	1,672	1,672	1,672	1,672	1,672	1,531	1,672	1,672	1,672
Number of agents	63	63	63	63	63	63	57	63	63	63
R-squared	0.2899	0.2956	0.2910	0.3093	0.3043	0.3255	0.2390	0.2946	0.2900	0.3367

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable: $\log(y_{it})$. All regressions also include a constant, working hours, share in peak hours, calls per fulltime equivalents (FTE), a linear time trend, and fixed effects for the week in which agents participated in the survey. Standard errors are clustered at the team level. The F-Statistic and corresponding p -value corresponds to a t -test of the main and interaction effects of all added personality traits and cognitive test scores. Due to partial non-response, the sample used in Column 7 is based on 57 agents instead of 63 agents.

Table A.4: Reciprocity and the effect of tenure on the returns to training

	(1)	(2)
Treatment dummy	0.1131*** (0.0156)	0.0940*** (0.0285)
Reciprocity	0.0035 (0.0248)	0.0040 (0.0244)
Treatment \times reciprocity	0.0535** (0.0216)	0.0530** (0.0212)
Tenure (in years)	-0.0021 (0.0051)	-0.0032 (0.0049)
Treatment \times tenure		0.0043 (0.0045)
Constant	-0.9700*** (0.0529)	-0.9659*** (0.0506)
Observations	1,672	1,672
Number of agents	63	63
R-squared	0.2901	0.2908

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable: $\log(y_{it})$. All regressions also include a constant, working hours, share in peak hours, calls per fulltime equivalents (FTE), a linear time trend, and fixed effects for the week in which agents participated in the survey. Standard errors are clustered at the team level.