

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

June 2020

Jan Sauermann<sup>†</sup>

## Abstract

Do reciprocal workers have higher returns to employer-sponsored training? Using a field experiment with random assignment to training combined with survey information on workers' reciprocal inclinations, the results show that reciprocal workers reciprocate employers' training investments by higher post-training performance. This result, which is robust to controlling for observed personality traits and worker fixed effects, suggests that individuals reciprocate the firm's human capital investment with higher effort, in line with theoretical models on gift exchange in the workplace. This finding provides an alternative rationale to explain firm training investments even with risk of poaching.

**JEL Codes:** J24, M53, D03

**Keywords:** on-the-job training, reciprocity, worker performance, field experiment

---

\*The author gratefully acknowledges financial support from the Jan Wallanders och Tom Hedelius Stiftelse (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, the coeditor, anonymous referees and seminar participants at SOFI Stockholm, HECER Helsinki, the COPE workshop in Vienna, the Behavioral Economics Network Workshop in Stockholm, KTH Stockholm, and RWTH Aachen for valuable comments and suggestions. The administrative data used in this paper contains sensitive information on the firm and its employees. Please contact the author when seeking to conduct a replication of the results presented in this paper. The field experiment is registered as AEARCTR-0004967 in the AEA RCT registry.

<sup>†</sup>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 long been of interest to both policymakers and firms. In the US, 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 during work hours (Leuven and Oosterbeek, 1999; Caliendo et al., 2015). Why firms invest in training despite the risk of poaching is often explained by market imperfections, such as compressed wages (Acemoglu and Pischke, 1998, 1999). An alternative argument is based on reciprocity, which may induce employees to reciprocate employer investments by, for example, reducing turnover, providing greater effort, or reducing wage demands (Leuven et al., 2005; Non, 2020). Reciprocating training investment thereby serves as a behavioral explanation for *why* firms invest in training.

This paper tests whether positively reciprocal workers respond to training investments by exerting greater effort in response to participation in firm-sponsored training, which should then result in higher post-training performance. The hypothesis that reciprocity is linked to higher training returns is tested using data on direct measures of worker performance, collected in an in-house call center of a multinational mobile network operator in the Netherlands. I exploit random assignment of workers to a training program in this firm to causally estimate the returns to training (De Grip and Sauermann, 2012).<sup>1</sup> Arguing that the provision of this firm-sponsored and fully paid, week-long training program can be perceived as a positive action by the firm towards the worker, I show that reciprocal workers have higher returns to the training program; that is, they perform better after training than do workers with low reciprocal inclinations after participation in the training. This is explained neither by other observed personality measures, nor by controlling for unobserved individual-specific effects.

Why should reciprocal inclinations matter for workplace training? Individuals with reciprocal inclinations directly receive utility from rewarding (positive reciprocity) or punish-

---

<sup>1</sup>De Grip and Sauermann (2012) study 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.

ing (negative reciprocity), even if there is no direct material pay-off from this action (Rabin, 1993; Falk and Fischbacher, 2006). Many studies using laboratory and field experiments show that gifts from the employer induce workers to provide greater effort than they would without the gift.<sup>2</sup> Most of these studies analyze the effect of a randomized treatment, that is, a monetary or nonmonetary gift, on subjects' effort provision, which allows establishment of a causal link between size and type of gift and a subject's response to the gift.

Fewer studies use explicit measures of reciprocal attitudes derived through experiments or surveys in combination with worker and firm-level outcomes. Barr and Serneels (2009) 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 increase in hourly wages can be observed only for workers who displayed reciprocal attitudes in a choice experiment. In environments in which experiments are not feasible or possible, implementation of survey questions on reciprocity in large-scale surveys allows testing the importance of reciprocity for labor market outcomes (Perugini et al., 2003). Englmaier and Leider (2020) provide evidence from a field experiment showing that positive responses to high (low) wages are driven mainly by individuals with high (low) reciprocal inclinations, which are derived from personality tests. Using a large representative subject pool for Germany, for example, Dohmen et al. (2009) show that workers' reciprocal attitudes are linked to higher wages and to working longer hours.

Provision of training opportunities can be seen as one way through which firms may induce reciprocal responses. These responses could be triggered either by perceiving the training as a gift (for example, if the training is of a 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 cross-sectional data for the Netherlands, Leuven et al. (2005) provide evidence that positive reciprocal individuals are more likely to participate

---

<sup>2</sup>See, 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).

in employer-financed training courses than individuals with low reciprocal inclinations. For Germany, Non (2020) provides corroborative evidence that reciprocal individuals are more likely to participate in training. Although this correlation could reflect unintended sorting of reciprocal individuals into training firms, the sorting pattern is also in line with the idea of strategic selection by firms to induce reciprocal incentives (Englmaier and Leider, 2012). Indeed, Englmaier et al. (2016) show that firms using personality tests when hiring are more likely to offer, among other benefits, on-the-job training. Non (2020) finds that reciprocity matters for training in slack labor markets only, suggesting that employers use training as a means of establishing gift-exchange relationships strategically.<sup>3</sup>

In the present study, randomized participation in a training program allows estimation of both the causal effect of training participation on performance, along with the indirect effect, that is, interaction between the training effect and the worker’s reciprocal attitudes. This not only helps to explain the positive correlation between training incidence and reciprocal attitudes (Leuven et al., 2005; Non, 2020), but also contributes to the general understanding of firm investments in employee human capital by offering an alternative rationale for firm investments in training even if there is risk of poaching (Acemoglu and Pischke, 1999; Leuven and Oosterbeek, 1999; Caliendo et al., 2015; Hoffman and Burks, 2017). This paper also contributes to the literature by addressing through which mechanisms reciprocity might affect the returns to training. Reciprocity could matter for the returns to training through increased effort *after* participating in the training, but could also lead to more efficient human capital acquisition *during* participation in the training course. Because more efficient human capital acquisition is likely to have more stable long-term effects, analyzing the mechanisms likely has important implications for firm investment in training.

Finally, this study also contributes to the literature on the returns to workplace training, the literature on personality in the labor market, and the role of personality in the heterogeneity of estimated returns to training. In the literature on training returns, causal interpretation of estimated returns to training has been a major issue addressed mainly by fixed effects approaches with observational data, and, to a lesser degree, with experimental

---

<sup>3</sup>Although other, nonexperimental studies explore reciprocal reactions to training courses, it is not clear to what degree their results are driven by potential endogeneity of training participation (Mullen et al., 2006; Kampkötter and Marggraf, 2015; Montizaan et al., 2015b).

variation (De Grip and Sauermann, 2012; Schwerdt et al., 2012; Adhvaryu et al., 2018; Prada et al., 2019). At the same time, ever more studies show that personality traits are predictive for, among other factors, educational choices, job search, employment, and earnings. Besides reciprocity, examples include the Big Five personality traits (Bowles et al., 2001; Mueller and Plug, 2006; Gensowski, 2018), which are survey-based measures containing factors on openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism, as well as locus of control, which is a survey-based measure of the degree to which individuals believe they have control over life outcomes (Caliendo et al., 2015, 2020). With regard to training, Offerhaus (2013) finds no effect of the Big Five on training participation using a representative sample for Germany, whereas Caliendo et al. (2020) provide evidence that higher internal locus of control is related to higher participation rates in general training courses.<sup>4</sup> 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, allowing analysis of how personality affects the returns to training.

This paper proceeds as follows. Section 2 details 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

The field experiment exploited in this study was organized in the in-house call center of a multinational mobile network operator in the Netherlands from week 45/2008 to week 24/2009.<sup>5</sup> This call center acts as a service center for current and prospective customers. I

---

<sup>4</sup>In addition to these papers analyzing specific attitudes and personality, a number of studies analyze cognitive and noncognitive determinants of experimental and labor market outcomes. See, for example, Bowles et al. (2001), Heckman et al. (2006), Mueller and Plug (2006) Lindqvist and Vestman (2011), Caliendo et al. (2015) and Gill and Prowse (2016).

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

focus on the largest department, which serves only private customers with fixed cell phone contracts. Call agents in this department have only one task, to answer incoming customer phone calls, for example, when customers have problems, complaints, or questions. Agents in 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 13 teams, each of which is led by a team leader. In each week, an average of nine agents works in each team. The main purpose of being assigned to a team leader is that workers can be more efficiently supervised and monitored. There is no team specialization, team production, nor team-based incentives. Calls are typically queued before they are assigned to an available agent, irrespective of the agent’s team membership. 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.<sup>6</sup>

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. This information is available for each individual agent and each working week.<sup>7</sup> 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.

---

<sup>6</sup>Appraisal 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. Assuming that assignment to treatment and control groups does not affect the appraisal outcomes, this should not invalidate the findings presented in this paper.

<sup>7</sup>This 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).

## 2.2 The field experiment

In the department analyzed, management introduced a new training program with the explicit aim of decreasing the average time needed to handle calls. The training was designed as a week-long program, held in the call center’s in-house training center over five consecutive days from Monday to Friday. Call agents were paid the full-time 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 agents could provide more help to each other. During the other half of the training time, training coaches assisted agents in handling customer calls.

The 32-week-long sample period between week 45/2008 and week 24/2009 consists of three periods: a pre-training period (weeks 45/2008-09/2009), a training period (weeks 10/2009-14/2009), and a post-training period (weeks 15/2009-24/2009). During the pre-training period, in week 50/2008, 74 out of the total 177 agents were selected for participation in the field experiment. This non-randomly selected group was then randomly assigned to treatment and control groups.<sup>8</sup> 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 the sample period used in this paper.

I make use of survey information on reciprocity and other individual-specific characteristics derived at 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.<sup>9</sup> For only one of the fac-

---

<sup>8</sup>The 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 random assignment of agents selected for the field experiment.

<sup>9</sup>Table 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 ( $N = 63$ ).

Table 1: Descriptive statistics

Sample	(1) All	(2) Treat. group	(3) Control group	(4) Diff (3)–(2)
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)
Positive reciprocity (std.)	0.0118 (1.0283)	-0.1105 (1.1403)	0.1230 (0.9184)	0.2335 (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

Notes: Standard deviations are in parentheses in Columns (1) to (3); *t*-statistics in Column (4). All variables are measured before the training intervention.



tors of the Big Five measures, extraversion, is the difference between treatment and control groups 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 exogenous, conditional on being assigned to participate in the field experiment.

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

## 2.3 Why this setting is useful to study reciprocal behavior

There are three reasons that make this field experiment useful in studying whether workers reciprocate the firm’s training investments. First, random assignment of agents into treatment and control groups provides a setting that allows estimation of 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 within-worker 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 weeks after their training participation. This setting has two important implications for results interpretation. 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 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 implies that the estimates are likely to be a lower bound of the true effect of the interaction between the training and reciprocity.<sup>10</sup>

---

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

Second, during the training program, for which the agents were paid their full-time wages, agents were taken out of the usual work environment and placed in a training center separated from the work floor. Results from an internal post-training 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 respondents strongly agree with the statement that “the training investment was worth the effort”.<sup>11</sup> This suggests that agents perceived the training as a positive action by the firm, rather than a chore.

Third, the theoretical motivation of this paper is that workers may perceive the training as a gift if it is of a 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 content that could also be productive in other call centers, of which there are several in the direct vicinity. The effect of reciprocity should be stronger only if the training has more general content.

## 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 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, for example, implemented in the German Socio-Economic Panel (GSOEP, Dohmen et al., 2009). In this survey, respondents were asked to rate six questions on a five-point Likert scale from 1 (“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.<sup>12</sup> Although reciprocity measures derived from

---

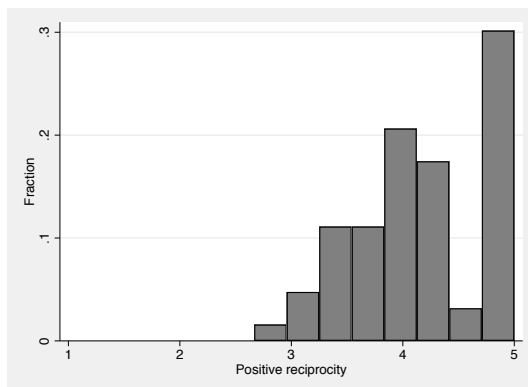
<sup>11</sup>This internal survey is based only partly on individuals from the sample used in this study, and cannot be merged into the data used in this study.

<sup>12</sup>See Table A.2 for the questions. All results in this study are qualitatively similar when using measures of reciprocity, which are derived from Principal Component Analysis.

a gift-exchange experiment are more reliable (e.g. Cohn et al., 2015), survey evidence on reciprocity yields 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, which is similar to population-wide distributions of reciprocity (see, e.g., Dohmen et al., 2009). This distribution is likely determined by a number of factors. While the distribution of reciprocity among applicants to this firm’s vacancies is unknown, personality tests applied in the hiring process can lead to an oversampling of reciprocal workers (cf. Englmaier et al., 2016). Furthermore, the negative correlation between reciprocity and turnover will affect the distribution of reciprocity in the firm (cf. Table 2). While the randomization is not affected by this selection, this suggests that workers might be more reciprocal than, for example, applicants or new hires to the firm.

Figure 1: Distribution of reciprocity



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

To analyze 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, all gathered in the same survey. For personality, I make use of the Big Five factors measuring conscientiousness, extraversion, agreeableness, openness to new experience and neuroticism, locus of control, which measures the degree to which individuals believe they have control over life outcomes, and negative reciprocity, defined as the tendency to retaliate

for negative experiences.<sup>13</sup> Each of these personality measures are 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, 2020; Gensowski, 2018). Table A.2 provides detailed information on measurement, questions used, and references for each of these measures.

## 2.5 Correlates of reciprocity

Table 2 contains 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 agent 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.<sup>14</sup> 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-month time frame after the end of the experiment, is correlated with the measure of reciprocity (Column (3) of Table 2).<sup>15</sup> 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 stay longer in the firm.

---

<sup>13</sup>The measure of negative reciprocity is based on the three other survey questions on reciprocity (see Table A.2). The measures of positive and negative reciprocity has often shown to be not correlated (see, e.g., Dohmen et al., 2009).

<sup>14</sup>With 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).

<sup>15</sup>Note that it is possible to observe only whether individuals left the department, 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 leave the firm entirely.

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		

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Figures in this table show pairwise Pearson's correlation coefficients of worker characteristics with the measure of positive reciprocity (standardized). The figures are based on the estimation sample and contain one observation per agent ( $N = 63$ , cf. Column (1) of Table 1). All personality measures are standardized with 0 mean and standard deviation of 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 as whether a worker left within 6 months after the end of the experiment.

### 3 Results

#### 3.1 Effect of reciprocity on 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. 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  $t$ , based on average handling time and for which high levels of  $y_{it}$  are interpreted as high performance. The dummy  $d_{it}$  equals 1 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 whether an agent works during peak hours with high customer load in week  $t$ .<sup>16</sup> 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 full-time equivalent agents ( $X_t$ ).<sup>17</sup> The idiosyncratic error term  $u_{it}$  is clustered at the team level to account for team-level randomization (cf. Section 2.2 and Abadie et al. (2017)).

Table 3 shows the results of estimating Equation 1. Column (1) shows that the causal effect of training participation on performance is 0.0861, that is, after participating in the training program, agents are, on average, 8.6% more effective in their main task, answering customer calls (cf. De Grip and Sauermann, 2012).<sup>18</sup> 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*.<sup>19</sup>

The coefficient of main interest in this study,  $\hat{\tau}_3$ , indicates whether the treatment effect is heterogeneous with respect to workers’ level of reciprocity. Columns (3) and (4) of Table 3 show 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 percentage-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, possibly to return the training investment with a favorable action to their employer.<sup>20</sup>

---

<sup>16</sup>In contrast to De Grip and Sauermann (2012), this paper refrains from using working hours as a control variable since the number of working hours is possibly affected by reciprocal behavior (see Section 4.1).

<sup>17</sup>Because surveys were not all conducted in the same week, fixed effects are included to control for possible survey week effects.

<sup>18</sup>Table A.3 in the Online Appendix shows that the reported treatment effect differs only slightly 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.

<sup>19</sup>Note that the strong increase in the adjusted  $R^2$  from Columns (1) to (2) of Table 3 is due to the little worker-specific information Column (1) controls for.

<sup>20</sup>An alternative interpretation of this result would be that, within an existing gift exchange between firm and agent, 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.

Table 3: Returns to training and interaction with reciprocity

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment dummy	0.0861*** (0.0229)	0.1052*** (0.0164)	0.1090*** (0.0158)	0.0753*** (0.0235)	0.0877*** (0.0150)	0.1138*** (0.0208)
Share Peak-Hours	-0.2501** (0.0949)	-0.2666*** (0.0602)	-0.2732*** (0.0540)	-0.2642*** (0.0571)	-0.3697*** (0.0381)	-0.3559*** (0.0814)
Calls per FTE	0.0001 (0.0000)	0.0001* (0.0001)	0.0001* (0.0001)		0.0001** (0.0000)	0.0001 (0.0001)
Time trend	0.0021 (0.0013)	0.0016** (0.0007)	0.0015* (0.0007)		0.0022*** (0.0006)	0.0014 (0.0008)
Reciprocity (std.)		0.0123 (0.0247)	0.0016 (0.0262)	0.0026 (0.0260)		0.0274 (0.0395)
Treatment $\times$ reciprocity			0.0514** (0.0222)	0.0500** (0.0219)	0.0151** (0.0059)	0.0430* (0.0206)
Constant	-1.0473*** (0.0938)	-0.9735*** (0.0729)	-0.9629*** (0.0689)	-0.7978*** (0.0349)	-1.0162*** (0.0541)	-0.8637*** (0.0791)
Observations	1,672	1,672	1,672	1,672	1,672	1,531
Number of agents	63	63	63	63	63	57
Adjusted R-squared	0.0706	0.2713	0.2790	0.3082	0.6109	0.3017
Worker FE	No	No	No	No	Yes	No
Week FE	No	No	No	Yes	No	No
Other measures of personality	No	No	No	No	No	Yes

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Dependent variable:  $\log(y_{it})$ . All regressions include fixed effects for the week in which agents participated in the survey. Other measures of personality include the Big Five elements, cognitive test score, negative reciprocity, locus of control, and loyalty. Standard errors are clustered at the team level.

A potential concern is that the measure of reciprocity is correlated with unobservable characteristics, such as other personality traits or cognitive ability, which might ultimately drive that which is attributed to reciprocity. An explicit way of dealing with any other characteristics would be to augment the error term in Equation 1 by an unobserved individual-specific characteristic  $\mu_i$  and to estimate it in a fixed effects framework, i.e. by estimating (for simplification, subscript  $t$  and additional covariates are suppressed)

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

in which  $\hat{\tau}_1$  is the estimated (main) treatment effect. Even though the experimental design does not require individual fixed effects to estimate the causal effect of training ( $\tau_2$ ), fixed effects serve the purpose of controlling for unobserved (fixed) characteristics that are potentially correlated with the measure of reciprocity.

Column (5) of Table 3 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 2.1 percentage points lower compared to the baseline effect in Column (3) of Table 3 (10.9%). The interaction effect between treatment and reciprocity, however, is slightly less than a third of the size shown in Table 3 (1.5% vs. 5.1%, cf. Column (5) of Table 3).<sup>21</sup>

The results show that reciprocal individuals do have higher returns to training, despite 70% of the original effect being explained by unobserved individual-specific characteristics  $\mu_i$ , i.e. resulting in  $Cov(\mu_i, rec_i) \neq 0$ . This result also shows the importance of appropriately addressing unobserved characteristics that might be correlated with the variable of interest.

### 3.2 Do other personality traits and test scores matter?

To learn more about which individual-specific characteristics are possibly correlated with reciprocity, additional survey questions 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.<sup>22</sup> Most of these individual-specific characteristics are 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 is shown to be relevant for both participation in training as well as returns (Offerhaus, 2013; Caliendo et al., 2020).<sup>23</sup> As shown in Table 2 and in line with Dohmen et al. (2008), all Big 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 treatment ( $\hat{\gamma}_3$ ), Equation 1 is estimated by including each of these measures. Column (6) of Table 3 shows that the effect is slightly

---

<sup>21</sup>Testing for the equality of the estimates in Column (3) and (5) of Table 3, the null hypothesis is rejected with a  $p$ -value of 10%.

<sup>22</sup>See Section 2.4 and Table A.2 for details on these measures.

<sup>23</sup>See, 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.



smaller than the estimate without controlling for these measures (Column (3)). This suggests that none of these variables causes the lower estimate when including individual-specific fixed effects.

To further provide evidence that the interaction effect between the treatment dummy and the measure of reciprocity is not merely driven by other individual-specific characteristics, Table A.4 augments the regression shown in Column (3) of Table 3 by each of the individual-specific characteristics and their respective interaction effects with the treatment dummy. The coefficient of interest, the interaction between reciprocity and the treatment, remains significant and stable throughout all regressions and varies between 0.042-0.056. Only when including the Big Five personality element of Neuroticism, the interaction effect becomes insignificant with a coefficient of 0.042.<sup>24</sup>

### 3.3 Reciprocating during or after training?

Having established that reciprocal individuals have higher returns to training, it is not clear *why* this is the case. There are two possible mechanisms that could drive this effect. First, reciprocal individuals could simply exert greater effort *after* participation in training, resulting in treatment effect heterogeneity by an individual's degree of reciprocity. Second, reciprocity could already matter earlier; that is, reciprocal individuals are better in utilizing the training when participating in the training. While it is difficult to find explicit tests to discriminate between these two mechanisms, higher effort during the training course should lead to more efficient human capital acquisition, as skills of reciprocal agents are higher than those of non-reciprocal agents. This should lead to a more permanent increase in skills and create rather stable effects on worker performance. If human capital acquisition is not affected by reciprocity, however, one might rather expect a transitory effect on performance that might fade out.<sup>25</sup>

---

<sup>24</sup>The results shown in Table A.4 also show that there are also heterogenous returns to training with respect to the Big Five personality element of Extraversion, the cognitive test score and the measure of loyalty. None of these, however, seems to affect the interaction between the treatment dummy and the measure of positive reciprocity.

<sup>25</sup>An alternative way of discriminating between these two mechanisms could be possible if one can observe workers moving across firms. If reciprocal workers are better in human capital acquisition during (general)

Figure 2 shows the treatment effect and the interaction with reciprocity, separately, for each week before and after the training. The estimates, which are taken from an estimation with the same specification as shown in Column (5) of Table 3, show that both the treatment effect and its interaction with reciprocity are zero in the weeks before the training. The treatment effect exhibits a dynamic pattern, i.e. it reaches a peak in the fifth week after training, and decreases substantially thereafter; this could be explained by spillover effects (De Grip and Sauermann, 2012), motivational effects, or even human capital depreciation. The solid gray line shows the interaction effect between the treatment dummy and reciprocity by week, after the training, and thereby corresponds to the interaction effect shown in Column (5) of Table 3. Although it is small in size, it does not follow the dynamics of the treatment effect. 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 non-reciprocal individuals. This could be, for example, explained by higher effort provision during training, resulting in more efficient human capital acquisition and more stable increases in performance.

## 4 Additional results

### 4.1 Alternative channels of reciprocal behavior

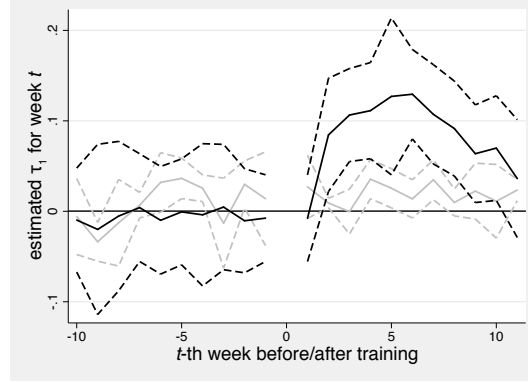
The results thus far provide evidence for treatment effect heterogeneity with respect to workers' degree of positive reciprocity. Although the descriptive analysis provides only limited evidence for significant correlation between reciprocity and other worker outcomes, reciprocal individuals might have alternative, possibly competing, channels through which training investments are returned. Examples for these competing channels are the number of hours worked by the agent and the number of hours of absence, for example, due to sickness.<sup>26</sup>

---

training, individual productivity should also increase performance after switching to other firms, whereas the behavioral mechanism of exerting higher effort after the training should not increase performance at other firms. I thank the coeditor for this suggestion.

<sup>26</sup>There are two alternative arguments, which cannot be tested in this setting. First, Leuven et al. (2005) put forward the argument that employees could reciprocate training with lower wage demands. In the

Figure 2: Treatment and interaction effects on performance over time



*Notes:* This figure shows the estimated treatment effect on performance for each week before and after the training (solid black line) and the corresponding 95% confidence interval (dashed black lines). The gray lines show 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 individual fixed effects, 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 (5) of Table 3). Week 0 denotes the training week.

First, reciprocal individuals could return the training investment by working longer hours. For the firm, this would result in better utilization of labor. Column (2) of Table 4 shows that when estimating Equation 1 with the number of working hours as an outcome variable, the interaction variable is not significant. When using the preferred specification including individual fixed effects to control for unobserved individual-specific characteristics that might be correlated with the measure of reciprocity, the estimate is significant and positive. This result suggests that workers with a one standard deviation higher reciprocity increase their working time by 0.4 hours per week (Column (3)).

Second, reciprocity might have a negative effect on hours of absence. If reciprocal individuals have lower degrees of sickness absence after the training, this would suggest that the training investment is reciprocated with lower absence. Kampkötter and Marggraf (2015) and Adhvaryu et al. (2018) show that training participation can affect these outcomes. To address this channel, Equation 1 is estimated with an outcome variable, *hours of absence*.

context of this field experiment, however, this hypothesis is not testable, because there is no information on wage bargaining. Second, workers might reciprocate training investments by reducing turnover, making investments therefore more beneficial for employers. In the setting of this study, however, only four agents left the department in the weeks between the training period and the end of the observation period, making it difficult to use turnover as an outcome variable.

Table 4: Returns to training on working hours and sickness absence

Outcome variable	(1)	(2)	(3)	(4)	(5)	(6)
	Working hours			Hours of absence		
Treatment dummy	-1.1631*	-1.2477*	-0.9891	0.0250***	0.0235**	-0.0044
	(0.6341)	(0.6189)	(0.9021)	(0.0070)	(0.0086)	(0.0120)
Reciprocity (std.)	-0.0860	0.1557		-0.0148	-0.0106	
	(0.9537)	(1.0073)		(0.0109)	(0.0128)	
Treatment $\times$ reciprocity		-1.1609	0.4089**		-0.0203	-0.0166
		(0.9827)	(0.1745)		(0.0139)	(0.0162)
Constant	22.0926***	21.8515***	22.9772***	0.1438*	0.1396*	0.1662**
	(3.7003)	(3.6349)	(4.2591)	(0.0730)	(0.0728)	(0.0568)
Observations	1,672	1,672	1,672	1,672	1,672	1,672
Number of agents	63	63	63	63	63	63
Adjusted R-squared	0.2095	0.2121	0.5059	0.0649	0.0659	0.1691
Individual FE	No	No	Yes	No	No	Yes

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Dependent variable: number of working hours (Columns (1) to (3)) and hours of absence (Columns (4) to (6)). All regressions also include calls per FTE, a linear time trend, 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.

With or without including individual fixed effects in the estimation, the results show that  $\hat{\tau}_3$  has the expected negative sign, but is not significantly different from zero.

## 4.2 Do agents return 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.<sup>27</sup> 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 selected to be part of the field experiment were primarily selected on their tenure. Table A.1 in the Online Appendix shows that agents in the field experiment have much longer tenure than the average for all workers. Being selected into the field experiment might be interpreted by agents as a signal that their tenure in the firm, i.e. their firm-specific human capital, is valued. To test whether there is systematic variation in the returns to training with respect to agent tenure, the estimation presented in Column (5) of Table 3 is augmented by tenure and its interaction with the treatment dummy. Although the experimental setup

<sup>27</sup>I thank an anonymous referee for highlighting this point.

ensures that tenure is balanced across treatment and control groups, variation in tenure can still be used to analyze this question. Stronger effects for more tenured agents would be in line with this hypothesis.

The estimates, however, show that the interaction effect of tenure with the treatment dummy is positive, yet not statistically significant (Table A.5). Although it is not possible to rule out that being selected into the training program *did* affect motivation and created increased performance, this does not drive the interaction effect between treatment and reciprocity.

### 4.3 Negative effects on non-trained workers

Given that the training was organized as a week-long training program, an important question is how removing up to 10 workers over one week from the work floor affects the remaining agents.<sup>28</sup> This implies that during training weeks, the headcount in training weeks is up to 10% lower at the department level, and up to 66% at the team level.

To test whether removing agents from teams for training purposes affects the remaining workers, Table A.6 tests whether the share of co-workers in training affects contemporaneous performance of the remaining agents. This table shows that having a higher share of co-workers participating in the training courses induces a modest negative effect on agent performance in the week that the co-workers are trained. This effect, however, is imprecisely estimated.

## 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 individuals reciprocate firm training investments. Using data from a field experiment in a call center in the Netherlands, this study is able to provide previously

---

<sup>28</sup>Note that all agents in the department do the same task, i.e. calls are randomly allocated to agents available in the department.

undocumented evidence that reciprocal individuals have higher returns to firm-sponsored training. To test this hypothesis, personnel data with panel information on worker performance are combined with random assignment to training courses and direct measures of reciprocal attitudes. Compared to survey datasets, this setting allows causal identification of treatment effects of participation in the training program, as well as its interaction with survey measures of reciprocity. Controlling for time-invariant individual-specific characteristics, the results show that individuals with reciprocal inclinations have higher returns to training. This suggests that reciprocal workers provide greater effort following firm training investments. Additional evidence suggests that this effort is provided during training, resulting in more efficient human capital acquisition.

The finding that individuals with reciprocal inclinations have higher returns to training has important implications for the worker–firm relationship. First, this finding provides an alternative channel through which the gift exchange between workers and firms can operate. Second, 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). Third, the result that individual fixed effects substantially reduce the estimated effect of interest shows that it is critical to consider alternative personality measures and individual fixed effects to account for worker-specific heterogeneity that might otherwise drive effects.

## References

- Abadie, A., Athey, S., Imbens, G., Wooldridge, J., 2017. When should you adjust standard errors for clustering? NBER Working Paper 24003, National Bureau of Economic Research.
- Acemoglu, D., Pischke, J.-S., 1998. Why do firms train? theory and evidence. *Quarterly Journal of Economics* 113 (1), 79–119.
- Acemoglu, D., Pischke, J.-S., 1999. The structure of wages and investment in general training. *Journal of Political Economy* 107, 539–72.

- Adhvaryu, A., Kala, N., Nyshadham, A., February 2018. The skills to pay the bills: Returns to on-the-job soft skills training. Working Paper 24313, National Bureau of Economic Research.
- Barr, A., Serneels, P., 2009. Reciprocity in the workplace. *Experimental Economics* 12 (1), 99–112.
- Becker, S. O., Messer, D., Wolter, S. C., 2013. A gift is not always a gift: Heterogeneity and long-term effects in a gift exchange experiment. *Economica* 80, 345–71.
- Borghans, L., Golsteyn, B., De Grip, A., March 2006. Meer werken is meer leren: Determinanten van kennisontwikkeling. Report, CINOP, 's-Hertogenbosch.
- Bowles, S., Gintis, H., Osborne, M., December 2001. The determinants of earnings: A behavioral approach. *Journal of Economic Literature* 39 (4), 1137–76.
- Caliendo, M., Cobb-Clark, D. A., Seitz, H., Uhlendorff, A., 2020. Locus of control and investment in training. *Journal of Human Resources* forthcoming.
- Caliendo, M., Cobb-Clark, D. A., Uhlendorff, A., March 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, G., 2004. Attribution and reciprocity in an experimental labor market. *Journal of Labor Economics* 22 (3), 665–88.
- Cohn, A., Fehr, E., Goette, L., 2015. Fair wages and effort provision: Combining evidence from a choice experiment and a field experiment. *Management Science* 61 (8), 1777–94.
- De Grip, A., Sauermann, J., 2012. The effects of training on own and co-worker productivity: Evidence from a field experiment. *Economic Journal* 122 (560), 376–99.
- DellaVigna, S., List, J. A., Malmendier, U., Rao, G., February 2016. Estimating social preferences and gift exchange at work. NBER Working Paper 22043, National Bureau of Economic Research.

- Dohmen, T., Falk, A., Huffman, D., Sunde, U., 2008. Representative trust and reciprocity: Prevalence and determinants. *Economic Inquiry* 46 (1), 84–90.
- Dohmen, T., Falk, A., Huffman, D., Sunde, U., 2009. Homo reciprocans: Survey evidence on behavioral outcomes. *Economic Journal* 119 (536), 592–612.
- Dur, R., 2009. Gift exchange in the workplace: Money or attention? *Journal of the European Economic Association* 7 (2-3), 550–60.
- Englmaier, F., Kolaska, T., Leider, S., 2016. Reciprocity in Organizations: Evidence from the UK. *CESifo Economic Studies* 62 (3), 522–46.
- Englmaier, F., Leider, S., May 2012. Contractual and organizational structure with reciprocal agents. *American Economic Journal: Microeconomics* 4 (2), 146–83.
- Englmaier, F., Leider, S., 2020. Managerial Payoff and Gift Exchange in the Field. *Review of Industrial Organization* 56, 259–280.
- Falk, A., Fischbacher, U., February 2006. A theory of reciprocity. *Games and Economic Behavior* 54 (2), 293–315.
- Fehr, E., Gächter, S., Kirchsteiger, G., 1997. Reciprocity as a contract enforcement device: Experimental evidence. *Econometrica* 65 (4), 833–60.
- Gensowski, M., 2018. Personality, IQ, and Lifetime Earnings. *Labour Economics* 51, 170–83.
- Gill, D., Prowse, V., 2016. Cognitive ability, character skills, and learning to play equilibrium: A level-k analysis. *Journal of Political Economy* 124 (6), 1619–1676.
- Gneezy, U., List, J. A., 2006. Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments. *Econometrica* 74 (5), 1365–84.
- Heckman, J. J., Stixrud, J., Urzua, S., July 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, H., Rockenbach, B., Sadrieh, A., 2010. In search of workers' real effort reciprocity—a field and a laboratory experiment. *Journal of the European Economic Association* 8 (4), 817–37.
- Hoffman, M., Burks, S. V., March 2017. Training contracts, employee turnover, and the returns from firm-sponsored general training. Working Paper 23247, National Bureau of Economic Research.
- Kampkötter, P., Marggraf, K., 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, S., Maréchal, M. A., Puppe, C., 2012. The currency of reciprocity: Gift exchange in the workplace. *American Economic Review* 102 (4), 1644–62.
- Leuven, E., Oosterbeek, H., 1999. Demand and supply of work-related training: Evidence from four countries. In: *Research in Labor Economics*. 18, pp. 303–30.
- Leuven, E., Oosterbeek, H., Sloof, R., Van Klaveren, C., 2005. Worker reciprocity and employer investment in training. *Economica* 72 (285), 137–49.
- Lindqvist, E., Vestman, R., January 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, X., Batt, R., 2007. The economic pay-offs to informal training: Evidence from routine service work. *Industrial and Labor Relations Review* 61 (1), 75–89.
- Montizaan, R., Cörvers, F., De Grip, A., Dohmen, T., 2015a. The impact of negatively reciprocal inclinations on worker behavior: Evidence from a retrenchment of pension rights. *Management Science* 62 (3), 668–81.
- Montizaan, R., de Grip, A., Fouarge, D., 2015b. Training access, reciprocity, and expected retirement age. IZA Discussion Paper 8862, Institute of Labor Economics (IZA).

- Mowday, R. T., Steers, R. M., Porter, L. W., 1979. The measurement of organizational commitment. *Journal of Vocational Behavior* 14 (2), 224–47.
- Mueller, G., Plug, E., 2006. Estimating the effect of personality on male and female earnings. *Industrial and Labor Relations Review* 60 (1), 3–22.
- Mullen, T. R., Kroustalis, C., Meade, A. W., Surface, E. A., 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.
- Murthy, N. N., Challagalla, G. N., Vincent, L. H., Shervani, T. A., 2008. The impact of simulation training on call center agent performance: A field-based investigation. *Management Science* 54 (2), 384–99.
- Non, A., 2020. Training participation and the role of reciprocal attitudes. *CESifo Economic Studies* 66 (1), 33–59.
- OECD, 2017. *OECD Skills Outlook 2017: Skills and Global Value Chains*. OECD Publishing.
- Offerhaus, J., 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, M., Gallucci, M., Presaghi, F., Ercolani, A. P., 2003. The personal norm of reciprocity. *European Journal of Personality* 17, 251–83.
- Prada, M., Rucci, G., Urzua, S., Jun. 2019. Training, Soft Skills and Productivity: Evidence from a Field Experiment. IZA Discussion Paper 12447.
- Rabin, M., 1993. Incorporating fairness into game theory and economics. *American Economic Review* 83 (5), 1281–302.
- Schwerdt, G., Messer, D., Woessmann, L., Wolter, S. C., 2012. The impact of an adult education voucher program: Evidence from a randomized field experiment. *Journal of Public Economics* 96 (7-8), 569–583.

# Appendix

Table A.1: Descriptive statistics of agents in the call center 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	

*Notes:* 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: Measurements of personality and cognitive test scores

Measure (Source)	Description and items
Reciprocity (Dohmen et al., 2009) (5-point Likert Scale)	<p>Average of items (1), (4) and (6) used to calculate positive reciprocity, (2), (3) and (5) for negative reciprocity</p> <ol style="list-style-type: none"> <li>(1) "If someone does me a favor, I am prepared to return it"</li> <li>(2) "If I suffer a serious wrong, I will take my revenge as soon as possible, no matter what the costs"</li> <li>(3) "If somebody puts me in a difficult position, I will do the same to him/her"</li> <li>(4) "I do my best to help somebody who helped me before"</li> <li>(5) "If somebody offends me, I will offend him/her back"</li> <li>(6) "I am ready to undergo personal costs to help somebody who helped me before"</li> </ol> <p>as measured by the 15-item questionnaire, implemented, among others, in the German Socio-Economic Panel and which covers the factors on openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism. "I'm somebody who..."</p>
Big Five (?) (5-point Likert Scale)	<ol style="list-style-type: none"> <li>(1) "has an active imagination" (openness)</li> <li>(2) "values artistic, aesthetic experiences" (openness)</li> <li>(3) "is original, comes up with new ideas" (openness)</li> <li>(4) "works thoroughly" (conscientiousness)</li> <li>(5) "does things efficiently" (conscientiousness)</li> <li>(6) "tends to be lazy" (conscientiousness, negative)</li> <li>(7) "is reserved, quiet" (extraversion, negative)</li> <li>(8) "is talkative" (extraversion)</li> <li>(9) "is outgoing, sociable" (extraversion)</li> <li>(10) "is sometimes rude to others" (agreeableness, negative)</li> <li>(11) "is considerate and kind to almost everyone" (agreeableness)</li> <li>(12) "has a forgiving nature" (agreeableness)</li> <li>(13) "worries a lot" (neuroticism)</li> <li>(14) "is relaxed, handles stress well" (neuroticism, negative)</li> <li>(15) "gets nervous easily" (neuroticism)</li> </ol> <p>the three questions on locus of control are part of a 15-item question that involve the concepts of locus of control, anxiety, self-image, self-confidence, and imagination</p>
Locus of control Borghans et al. (2006) (5-point Likert Scale)	<ol style="list-style-type: none"> <li>(1) "Setbacks are usually due to mistakes people make"</li> <li>(2) "Most people do not realize well to which extent their live is determined by chance"</li> <li>(3) "Whether I reach the goals I have in my life is not a matter of luck"</li> </ol>
Cognitive test score (CentERdata, 2007) (correct / incorrect)	<p>The cognitive test score is computed as the average of correct answers</p> <ol style="list-style-type: none"> <li>(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?"</li> <li>(2) "A rubber ball jumps up half the distance that 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?"</li> <li>(3) "Four girls are 100 cm, 150 cm, 125 cm, and 75 cm, respectively. Debbie is the tallest of the four girls. Karin is the shortest. Emmy is taller than Sara. How tall is Sara?"</li> <li>(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?"</li> <li>(5) "If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?"</li> <li>(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?"</li> </ol>
Loyalty (Mowday et al., 1979) (5-point Likert Scale)	<ol style="list-style-type: none"> <li>(1) "In general, I'm telling positive things about [the firm], even if other people criticize it"</li> <li>(2) "Management of [the firm] does not sufficiently motivate me to increase my performance" (negative)</li> <li>(3) "I'm happy that I chose [the firm] instead of a job at a different firm"</li> <li>(4) "I have no interest in the development of [the firm]" (negative)</li> <li>(5) "Management of [the firm] know what they do"</li> </ol>

Table A.3: Returns to training and sample definition

	(1)	(2)	(3)	(4)
Sample	Initial randomization		Estimation sample	
Treatment dummy	0.0828*** (0.0220)	0.0898*** (0.0205)	0.0861*** (0.0229)	0.0877*** (0.0207)
Share peak hours	-0.2291** (0.0753)	-0.3499*** (0.0760)	-0.2501** (0.0949)	-0.3621*** (0.0808)
Calls per FTE	0.0001 (0.0000)	0.0001** (0.0001)	0.0001 (0.0000)	0.0001* (0.0001)
Time trend	0.0025* (0.0013)	0.0022** (0.0009)	0.0021 (0.0013)	0.0022** (0.0008)
Constant	-1.0965*** (0.0805)	-1.0350*** (0.0867)	-1.0473*** (0.0938)	-1.0210*** (0.0904)
Observations	1,858	1,859	1,672	1,673
Number of agents	74	74	63	63
Adjusted R-squared	0.0715	0.5984	0.0706	0.6102
Individual FE	No	Yes	No	Yes

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . 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.4: Effect of reciprocity and worker-specific characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treatment dummy	0.1070*** (0.0169)	0.0844*** (0.0112)	0.1085*** (0.0156)	0.1117*** (0.0194)	0.1066*** (0.0139)	0.0989*** (0.0168)	0.0377 (0.0289)	0.1165*** (0.0195)	0.1097*** (0.0178)	0.1097*** (0.0197)
Reciprocity (std.)	0.0046 (0.0241)	0.0015 (0.0258)	0.0006 (0.0284)	0.0067 (0.0271)	-0.0066 (0.0263)	-0.0028 (0.0284)	0.0061 (0.0243)	0.0019 (0.0265)	0.0021 (0.0272)	0.0055 (0.0217)
Treatment × reciprocity	0.0475** (0.0174)	0.0528** (0.0215)	0.0470* (0.0222)	0.0516* (0.0242)	0.0403 (0.0260)	0.0549 (0.0330)	0.0444* (0.0221)	0.0550** (0.0232)	0.0513** (0.0225)	0.0414* (0.0187)
Conscientiousness	-0.0090 (0.0287)					0.0003 (0.0321)				
Treatment × conscientiousness	0.0132 (0.0251)					-0.0191 (0.0304)				
Extraversion		-0.0038 (0.0105)				-0.0214** (0.0083)				
Treatment × extraversion		0.0561** (0.0222)				0.0498*** (0.0133)				
Agreeableness			0.0051 (0.0146)			0.0021 (0.0180)				
Treatment × agreeableness			0.0177 (0.0139)			0.0129 (0.0130)				
Openness to experience				-0.0400 (0.0291)		-0.0339 (0.0284)				
Treatment × openness				-0.0212 (0.0515)		-0.0177 (0.0535)				
Neuroticism					-0.0251 (0.0145)	-0.0327* (0.0161)				
Treatment × neuroticism					-0.0188 (0.0318)	0.0106 (0.0334)				
Cognitive test score							-0.1002* (0.0517)			
Treatment × cog. test							0.1634*** (0.0350)			
Negative reciprocity								-0.0128 (0.0092)		
Treatment × neg. reciprocity								-0.0211 (0.0316)		
Locus of control									0.0067 (0.0180)	
Treatment × Locus of control									-0.0050 (0.0205)	
Loyalty										0.0478** (0.0185)
Treatment × loyalty										0.0396* (0.0210)
Constant	-0.9581*** (0.0740)	-0.9563*** (0.0707)	-0.9591*** (0.0681)	-0.9827*** (0.0781)	-0.9618*** (0.0698)	-0.9549*** (0.0658)	-0.9013*** (0.0670)	-0.9516*** (0.0706)	-0.9581*** (0.0807)	-0.9317*** (0.0682)
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
Adjusted R-squared	0.2787	0.2843	0.2800	0.2990	0.2923	0.3122	0.2271	0.2837	0.2786	0.3269

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Dependent variable:  $\log(y_{it})$ . All regressions include a constant, share in peak hours, calls per full-time 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. Due to partial nonresponse, the sample used in Columns (7) is based on 57 agents instead of 63 agents.

Table A.5: Reciprocity and the effect of tenure on returns to training

	(1)	(2)
Treatment dummy	0.0877*** (0.0150)	0.0706*** (0.0107)
Treatment $\times$ reciprocity	0.0151** (0.0059)	0.0145 (0.0085)
Treatment $\times$ tenure		0.0037 (0.0035)
Constant	-1.0162*** (0.0541)	-1.0163*** (0.0541)
Observations	1,672	1,672
Number of agents	63	63
Adjusted R-squared	0.6109	0.6111
Individual FE	Yes	Yes

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Dependent variable:  $\log(y_{it})$ . All regressions include the share in peak hours, calls per full-time 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 A.6: Contemporaneous effect of peers in training on own performance

	(1)
Share of peers in training	-0.0381 (0.0782)
Constant	-1.1323*** (0.0653)
Observations	1,672
Number of agents	63
Adjusted R-squared	0.6000
Individual FE	Yes

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Dependent variable:  $\log(y_{it})$ . Share of peers in training is defined as the number of co-workers on an agent's team, divided by the overall number of agents on a team. All regressions include share, in peak hours, calls per full-time 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.