

Becker, Johannes; Hopp, Daniel; Süß, Karolin

Working Paper

How Altruistic Is Indirect Reciprocity? - Evidence from Gift-Exchange Games in the Lab

CESifo Working Paper, No. 8423

Provided in Cooperation with:

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Becker, Johannes; Hopp, Daniel; Süß, Karolin (2020) : How Altruistic Is Indirect Reciprocity? - Evidence from Gift-Exchange Games in the Lab, CESifo Working Paper, No. 8423, Center for Economic Studies and Ifo Institute (CESifo), Munich

This Version is available at:

<https://hdl.handle.net/10419/223495>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

How Altruistic Is Indirect
Reciprocity? – Evidence from
Gift-Exchange Games
in the Lab

Johannes Becker, Daniel Hopp, Karolin Süß

Impressum:

CESifo Working Papers

ISSN 2364-1428 (electronic version)

Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH

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

Poschingerstr. 5, 81679 Munich, Germany

Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de

Editor: Clemens Fuest

<https://www.cesifo.org/en/wp>

An electronic version of the paper may be downloaded

- from the SSRN website: www.SSRN.com
- from the RePEc website: www.RePEc.org
- from the CESifo website: <https://www.cesifo.org/en/wp>

How Altruistic is Indirect Reciprocity? – Evidence from Gift-Exchange Games in the Lab

Abstract

Indirect reciprocity is defined as a specific kind of behavior: An agent rewards or penalizes another agent for having behaved kindly or unkindly toward a third party. This paper analyzes the question of what drives indirect reciprocity: Does the agent reward or penalize because she (altruistically) cares for the third party? Or does she take the other agent's behavior as a signal of how the latter would treat her if they met? In order to measure the relative importance of the altruism motive versus the signaling motive, we consider a gift-exchange game with three players: an employer pays wages to a worker and a coworker, before the worker (but not the coworker) may reciprocate by exerting effort. We offer a theoretical framework to analyze both motives for indirect reciprocity and run a series of lab experiments. The treatments manipulate the worker's information on wages. We find that, if only the coworker's wage is observable, the worker's effort increases in the coworker's wage. In contrast, if the worker can observe her own wage, the coworker's wage does not affect worker effort at all. We interpret this as support for the signaling motive: Indirect reciprocity is rather a byproduct of direct reciprocity than an act of altruism.

JEL-Codes: A130, C920, D910, J310.

Keywords: gift-exchange, indirect reciprocity, signaling.

*Johannes Becker**
Institute of Public Economics
University of Münster / Germany
johannes.becker@wiwi.uni-muenster.de

Daniel Hopp
Institute of Public Economics
University of Münster / Germany
daniel.hopp@wiwi.uni-muenster.de

Karolin Süß
Ruhr Graduate School in Economics and Chair
of Quantitative Methods in Economics,
University of Duisburg-Essen / Germany
karolin.suess@uni-due.de

*corresponding author

1 Introduction

Indirect reciprocity is believed to provide a powerful mechanism that induces stable levels of cooperation in large, anonymous groups such as modern societies (Herne et al., 2013). The notion of indirect reciprocity has first been addressed by Alexander (1987) and generally refers to rewards or punishments by an agent in response to interactions between other agents (see e.g. Falk and Fischbacher (2006)).¹ To be precise, in a group of three individuals (A, B and C), individual C’s behavior is characterized by indirect reciprocity if she behaves kindly (unkindly) toward A in response to A’s kind (unkind) behavior toward B. That is, C is willing to invest own resources in order to sanction an action she is not directly affected by.

A large experimental and theoretical literature focuses on the existence², evolution³ and stability⁴ of indirect reciprocity. However, there are only a few contributions on the question what *drives* indirectly reciprocal behavior. Reciprocal behavior is often rationalized through intention-based⁵ preferences (Rabin, 1993).⁶ In the context of indirect reciprocity, intention-based means that individual A’s payoff increases (decreases) C’s utility if C perceives A to be kind (unkind). Observing A’s behavior toward B allows for an update on A’s kindness (as perceived by C). Based on the update, C has an incentive to reward or punish A – e.g. if A’s behavior is interpreted as well-intended, her payoff enters positively into C’s utility function.

This paper focuses on the nature of the update: what exactly does C perceive as well-intended or kind? We consider two polar cases. First, if C were completely selfish, only A’s behavior toward C would matter. In this case, A’s behavior toward B is only interesting if it provides a *signal* how A would treat C if they met. Indirect reciprocity, ‘*in this sense, would just be a side-effect of direct reciprocity*’ as it ‘*serves as a noisy signal of [A’s] kindness towards [C]*’ (Engelmann and Fischbacher, 2009, p.406).⁷ Similarly, Alexander (1987) considers ‘*indirect reciprocity as a consequence of direct reciprocity occurring in the presence of interested audiences [...] who continually evaluate the members*

¹Such behavior is sometimes referred to as *downstream* indirect reciprocity (Nowak and Sigmund, 2005). Throughout the paper, we define reciprocity as a specific kind of behavior based on preferences rather than being a preference itself. For similar approaches, see e.g. Fehr et al. (1998b), Gächter and Falk (2002), Falk and Fischbacher (2006), Charness and Levine (2007) and Daryl and Riedl (2017).

²See Diekmann (2004), Cox and Deck (2005) and Moreno-Okuno and Mosiño (2017).

³See Axelrod and Hamilton (1981), Nowak and Sigmund (2005), Boyd and Richerson (1989), Lotem et al. (1999), Leimar and Hammerstein (2001), Sethi and Somanathan (2001), Bowles and Gintis (2004), Barta et al. (2011), Barta (2016), Enke (2019), Dufwenberg and Kirchsteiger (2019) and Ito et al. (2019).

⁴See Güth (1995), Nowak and Sigmund (1998), Fishmann et al. (2001), Leimar and Hammerstein (2001), Berger and Grüne (2016) and Tanaka et al. (2016).

⁵In contrast, Fehr and Schmidt (1999), Levine (1998) and Bolton and Ockenfels (2000) assume other-regarding preferences that are outcome-based.

⁶Charness and Rabin (2002), Falk and Fischbacher (2006) and Cox et al. (2007) offer models that combine outcome-based and intention-based preferences, see also Dufwenberg and Kirchsteiger (2019) and Dufwenberg and Kirchsteiger (2000) for an application to the labor market. Falk et al. (2008) find that intentions are crucial to explain the experimental data.

⁷Behavior is characterized by direct reciprocity when C reciprocates to A’s behavior toward C.

of their society as possible future interactants’ (pp. 93/94). Second, if C were (partly) altruistic, kindness would be defined more broadly and include A’s behavior toward B and others (see e.g. Nowak and Sigmund (2005)). Then, even if C knew for certain how A would treat C, A’s behavior toward B would still be relevant since C cares about B’s payoff.⁸

In order to rationalize reciprocal behavior, we use the generalized model from Dufwenberg and Kirchsteiger (2004) and augment it by an altruism term. The model is then used to clearly define the *signaling motive* and the *altruism motive* for indirect reciprocity. We design an experiment that allows disentangling the two motives for rewarding or penalizing other agents’ behavior.

We propose a classic gift-exchange game⁹ with three players: an employer, a worker and a coworker. The employer may set different wages for the worker and the coworker. Only the worker may reciprocate by choosing an effort level (that increases the employer’s payoff). The experimental treatments vary the information provided to the worker who is informed about either (i) only her own wage or (ii) only her coworker’s wage or (iii) both wages. These three different information settings allow testing for the impact of the coworker’s wage on effort levels with and without information on the worker’s own wage. If indirect reciprocity were rooted in altruistic preferences, a positive relationship between effort and the coworker’s wage should occur irrespective of the available information. In contrast, if indirect reciprocity is driven by selfishness and the employer’s wage for the coworker is just a signal, the coworker’s wage is not expected to affect the effort once the own wage becomes observable (as in this case, the employer’s kindness toward the worker does not need to be deduced from the coworker’s wage).

We find strong support for the *signaling motive*. The worker’s effort increases in the coworker’s wage only when her own wage is unobservable. The worker (correctly) anticipates that her own wage is positively correlated with the coworker’s wage. If both wages can be observed, the coworker’s wage seems to be irrelevant for the worker’s effort choice.

One concern with this interpretation is that the treatments differ in the complexity level. When both wages are observable, participants need to cope with two pieces of information and may opt to focus on just one of them: their own wage. To address this concern, we consider an additional treatment that allows for comparisons between decisions with similar complexity level. We let the computer randomly determine the worker’s wage (while the employer keeps setting the coworker’s wage). Thus, the worker’s own wage does not allow for conclusions about the employer’s kindness. In this setting, we find that the worker’s effort increases in the coworker’s wage – the worker seems to take the coworker’s wage as

⁸An alternative is to assume that individuals intentionally use indirect reciprocity in order to achieve a certain goal, e.g. to establish or enforce social norms, see e.g. Fehr et al. (1998b) and Reuben and Riedl (2013). In this case, C does not care about A’s specific behavior toward B, but anticipates that letting this behavior unsanctioned may erode the implicit rules of the group.

⁹The theory of gift-exchanges (Akerlof (1982) and Akerlof and Yellen (1990)) suggests that employers may overcome moral hazard problems in labor relations where effort is non-enforceable by motivating workers by paying high(er) wages. In theory, employees reciprocate this gift in bilateral gift-exchange games by exerting higher effort levels. Fehr et al. (1993) design a gift-exchange experiment and find that above-minimum-level wages increase workers’ effort. See also Fehr et al. (1997), Fehr et al. (1998a), Charness and Haruvy (2002), Gächter and Falk (2002) and Dariel and Riedl (2017) for further support and Fehr and Gächter (2000) for a survey of bilateral gift-exchange games. More recent papers use field experiments to analyze the importance of direct reciprocity in labor market environments, see Bellemare and Shearer (2009) and, for mixed results, Gneezy and List (2006) and Kube et al. (2012). Alternative experimental settings to analyze reciprocal behavior are sequential dictator games, see e.g. Herne et al. (2013) and Ito et al. (2019), and helping games, see e.g. Wedekind and Milinski (2000), Greiner and Levati (2005), Seinen and Schram (2006) and Engellmann and Fischbacher (2009).

a signal of how the employer would behave toward her if she were able to choose the worker's wage level. On a side note, these findings also inform about the question whether reciprocity is primarily based on outcomes or on intentions: When the computer (instead of the employer) determines wages, the worker increases her effort in response to a higher wage (although the latter is not due to the employer's kindness) – however, not as strongly as when the employer sets the wage level.

Interestingly, when the employer knows which treatment is played, she increases the coworker's wage even when the worker can observe both wages. The employer seems to (falsely) assume that workers respond to the coworker's wage even when their own wage is known.

We are not the first to consider a gift-exchange game with more than one (gift receiving) worker. Akerlof (1982) himself suggests that due to the '*acquire[d] sentiment for each other*' (p. 543), employees care about their coworker's wage and thus increase effort in their own and their coworker's wage (see also Akerlof and Yellen (1990)).¹⁰ In a setting where the employer pays the same wage to all workers, Maximiano et al. (2007) do not find significant behavioral differences between a bilateral gift-exchange game and a game with one employer and four employees. Charness and Levine (2007) let the employer pay different wages to workers who differ in productivities. In this setting, workers react strongly to their own wage but seem to be indifferent toward their coworker's wage. Rivas (2009) tests the impact of the coworkers' wages on effort when the respective productivity levels are public information. In this experiment, inequality in wages only affects effort negatively if wages are remarkably unfair and the high productivity worker earns 30 times more than the low productivity worker. Gächter and Thöni (2010) explicitly test whether workers evaluate the fairness of their wage relatively to the wages in a reference group that is most likely to be their coworkers. In line with Akerlof and Yellen (1990), they find that on average, workers' effort increases in their own wage as long as their wage falls short of their coworkers' wages and that they do not further increase effort if they earn more than their reference group. However, only 15 percent of the participants increase their effort in response to higher coworkers' wages.¹¹ Gross et al. (2015) find that highly productive workers decrease effort if they do not earn more than their less productive coworkers.¹²

The evidence for indirect reciprocity in real world settings is mixed as well. Heinz et al. (2017) find lower levels of productivity of participants in a field experiment after their colleagues have been treated unfair. In contrast, Cohn et al. (2014) find no effect on performance when the own wage is constant but the coworker's wage is cut.

To summarize, most existing studies do not find strong support for indirect reciprocity in multi-player gift-exchange experiments since many participants seem to be indifferent toward their coworker's well-being. This is different in helping games where indirect reciprocity is observed more frequently.¹³ One reason for this difference could be that, in most gift-exchange games considered so far, the workers know their own wage, i.e. the employers' behavior toward them can be observed. In helping games,

¹⁰One implication may be that a profit maximizing firm has an incentive to compress wages in order to elicit high levels of effort.

¹¹In a series of experiments, Gächter et al. (2012, 2013) let employees choose their effort sequentially. In their setting, instead of equalizing payoffs, subjects try to mimic other participant's behavior, meaning their effort. Pay comparison plays a much weaker role in predicting reciprocal behavior than the comparison of effort levels does.

¹²This result is also supported by Bolton and Werner (2016), who find that highly productive workers demand higher wages than less productive workers to exert the same level of effort. Similarly, employees with high effort levels expect to obtain substantially higher wages than their coworkers who exert lower effort levels (Abeler et al., 2010).

¹³See Wedekind and Milinski (2000), Greiner and Levati (2005), Seinen and Schram (2006) and Engelmann and Fischbacher (2009).

however, there is no direct interaction with the receiver and therefore the decision is solely based on her image score.¹⁴ This paper contributes to the literature by focusing on the interference of direct and indirect reciprocity. It provides an explanation (the signaling motive) for the contradictory results from multi-player gift-exchange games and helping games.

In the next section, we describe the experimental set-up and develop a model to clarify and distinguish altruism and selfishness as potential drivers of indirect reciprocity. Sect. 3 presents the results of the main experiment, Sect. 4 some robustness checks. The last section discusses and summarizes our main findings.

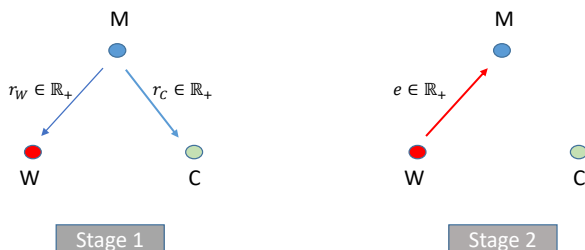
2 Theory and experimental design

In this section, we outline the theory as well as the experimental design.

2.1 Theory

To distinguish the two motives for indirect reciprocity, we start considering a more general decision setting of the following type. There are three agents, labeled M , W and C . At the first stage, agent M chooses an action $r_W \in \mathbb{R}_+$ and an action $r_C \in \mathbb{R}_+$. These two actions imply a cost for M and affect the *material* payoffs, π_W and π_C of agents W and C , respectively. At the second stage, agent W decides on $e \in \mathbb{R}_+$ that is costly for W and affects M 's material payoff, π_M .

Figure 2.1: Gift-exchange game with three players.



Such a setting encompasses a range of typical situations of direct reciprocity (where W responds to M 's choice of r_W) and indirect reciprocity (where W responds to r_C), e.g. gift-exchange games and helping games. These typical situations usually have the feature that, with purely selfish players, there is a subgame perfect Nash equilibrium with $(r_W^*, r_C^*, e^*) = (0, 0, 0)$. However, in lab experiments, the actually chosen levels of r_W , r_C and e are significantly different from zero. This may be due to altruism or, if e responds to r_W and/or r_C , due to preferences that induce direct or indirect reciprocity.

¹⁴This image score reflects how often the current receiver has helped other participants in the experiment before and, therefore, may be interpreted as a signal for the receiver's willingness to act kindly toward the donor if they had a direct interaction.

We consider a specific variant of the above decision setting: a gift-exchange game. We refer to M as the *manager*, to W as the *worker*, to C as the *coworker*, to r_W and r_C as wage rates and to e as effort. Given r_W , r_C and e , the agents' *material* payoffs π_i are $\pi_M = v(e) - r_C - r_W$ and $\pi_W = r_W - k(e)$ and $\pi_C = r_C$. We will assume that $v'(e), k'(e) > 0$, $v''(e) \leq 0$ and $k''(e) \geq 0$.

In order to analyze reciprocal behavior, we focus on the worker W who may respond to M 's choices of r_W and r_C at stage 2. Rabin (1993) argues that reciprocal behavior can be rationalized through a utility function that depends not only on material outcomes but also on intention-dependent fairness considerations.¹⁵ We assume a generalized version of the utility function (for the worker) proposed by Dufwenberg and Kirchsteiger (2004):

$$U_W = \pi_W + Y \cdot \underbrace{\lambda_W(r_W, r_C, e^b)}_{M's \text{ kindness as perceived by } W} \cdot \underbrace{\kappa_W(e, r_W, r_C)}_{W's \text{ kindness toward } M}$$

where $Y > 0$ is a shift parameter. The worker's utility has, thus, two parts. The first part reflects the worker's own material payoff, π_W . The second part depends, apart from Y , on two variables. $\kappa_W(e, r_W, r_C)$ measures how kind W is toward M . The higher W 's choice of e , the kinder is W 's behavior toward M . The notion of kindness is potentially scaled by the levels of r_W and r_C .¹⁶ Dufwenberg and Kirchsteiger (2004) suggest a specific functional form of $\kappa(\cdot)$ that we will adopt here to illustrate our main point:

$$\kappa_W(e, r_W, r_C) = \pi_M(r_W, r_C, e) - \pi_M(r_W, r_C, \bar{e}) = v(e) - v(\bar{e})$$

where \bar{e} is a level of effort deemed to be the kindness threshold (in the sense that $e > \bar{e}$ is kind, and $e < \bar{e}$ is unkind behavior).¹⁷

The variable $\kappa_W(\cdot)$ is multiplied by a variable $\lambda_W(r_W, r_C, e^b)$ that indicates to which degree the worker W *wants to be kind* toward the manager M . For instance, if r_W and (possibly) r_C have been chosen at high levels and are thus considered kind, $\lambda_W(r_W, r_C, e^b)$ is positive and large, implying that W has an incentive to exert high effort to benefit M . In contrast, if r_W and (possibly) r_C are low and regarded as unkind, $\lambda_W(r_W, r_C, e^b)$ may even be low or negative implying that the worker wants to set e to zero (or even harm the manager, if this were possible). Again, the notion of M 's kindness toward W is potentially scaled by the expected effort exerted by W at the second stage. Since this belief is not observable, W needs to form a belief on M 's belief on W 's choice of e (denoted by e^b).

¹⁵It is, thus, a "psychological game" in the sense of Geanakoplos et al. (1989).

¹⁶For instance, if M has chosen high wage rates r_W and r_C , a moderate effort choice e may be regarded as unkind, whereas the same level of effort may be perceived as kind if r_W and r_C are low.

¹⁷In contrast, Rabin (1993) assumes that the difference between M 's actual material payoff and the fair benchmark is normalized by the fair benchmark itself:

$$\kappa_W(e, r_W, r_C) = \frac{\pi_M(r_W, r_C, e) - \pi_M(r_W, r_C, \bar{e})}{\pi_M(r_W, r_C, \bar{e})}$$

In parallel to the above, Dufwenberg and Kirchsteiger (2004) propose¹⁸ $\lambda_W(r_W, r_C, e^b) = \pi_W(r_W, e^b) - \pi_W(\bar{r}_W, e^b) = r_W - \bar{r}_W$ where \bar{r}_W is a wage rate level deemed to be the kindness threshold. To allow for the possibility that W 's attitude toward M also depends on her behavior toward C , we augment the above functional form as follows:

$$\begin{aligned}\lambda_W(r_W, r_C, e^b) &= \pi_W(r_W, e^b) - \pi_W(\bar{r}_W, e^b) + \alpha[\pi_C(r_C) - \pi_C(\bar{r}_C)] \\ &= r_W - \bar{r}_W + \alpha[r_C - \bar{r}_C]\end{aligned}$$

where α is a measure of W 's altruism (or resentment) toward C .

At stage 2, r_W and r_C have been determined and e^b is only relevant at the time the wage rates are set. As a consequence, $\kappa_W(\cdot)$ only varies in e . The utility-maximizing choice of e is therefore implied by

$$\frac{\partial U_W}{\partial e} = -k'(e) + Y\lambda_W(r_W, r_C, e^b) \frac{\partial \kappa_W(e, r_W, r_C)}{\partial e} \leq 0 \quad (1)$$

Provided that $\frac{\partial \kappa_W(\cdot)}{\partial e} > 0$, the optimal effort e^* is zero if $\lambda_W(\cdot)$ is sufficiently small,¹⁹ i.e. if the worker does not care about the manager or, worse, wants to harm the manager because she feels treated badly. If $\lambda_W(r_W, r_C, e^b) > 0$ and there is an inner solution, the worker weighs, by choosing e^* , the cost of exerting effort against the benefit of being kind toward the manager.

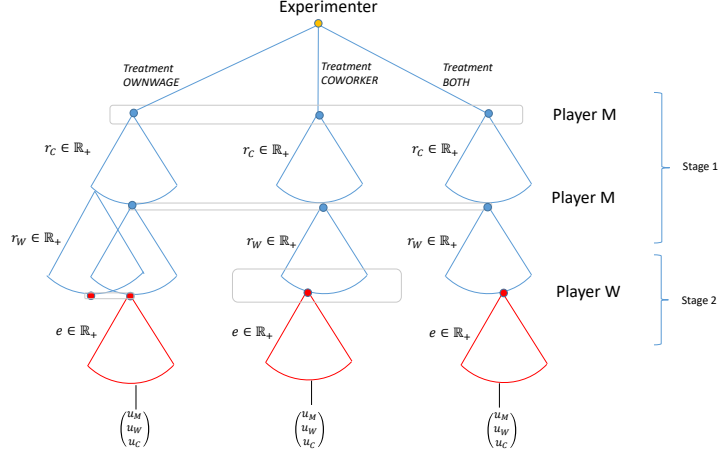
Before we narrow the focus on the analysis of indirect reciprocity, we outline the precise structure of the game under consideration. Figure 2.2 shows its extensive form. The 'experimenter' draws first and determines the 'treatment'. There are three treatments. In the treatment *OWNWAGE*, the worker W observes her own wage r_W but not the coworker's wage r_C . This treatment is mainly used as a benchmark. In treatment *COWORKER*, W observes r_C but not r_W . In this case, r_C may serve as a signal for r_W , i.e. $r_W = r_W^s(r_C)$. If a higher r_C signals higher r_W , the coworker's wage increases M 's kindness as perceived by W . In treatment *BOTH*, W observes both wage rates. Importantly, M does not know whether the treatment is *OWNWAGE*, *COWORKER* or *BOTH*, i.e. whether W can observe r_W or not. That is, the informational content of r_W and r_C does not vary across the treatments.

¹⁸Rabin (1993) proposes, again, to normalize the payoff difference:

$$\lambda_W(r_W, r_C, e^b) = \frac{\pi_W(r_W, e^b) - \pi_W(\bar{r}_W, e^b)}{\pi_W(\bar{r}_W, e^b)}$$

¹⁹Assuming $k'(0) = 0$, we have $e^* = 0$ if $\lambda_W(\cdot) \leq 0$.

Figure 2.2: Game tree of the experiment.



We may now derive the impact of the wage rates r_W and r_C on W 's choice of e :

$$\frac{de}{dr_i} = \frac{\frac{\partial^2 U_W}{\partial e \partial dr_i}}{-\frac{\partial^2 U_W}{\partial e^2}} \quad \text{with } i = W, C$$

With the above first order condition (1) representing a maximum, we have $\frac{\partial^2 U_W}{\partial e^2} = -k''(e) + Y\lambda_W(r_W, r_C, e^b) \frac{\partial^2 \kappa_W(e, r_W, r_C)}{\partial e^2} < 0$, and $\frac{\partial^2 U_W}{\partial e \partial dr_i}$ is given by

$$\frac{\partial^2 U_W}{\partial e \partial dr_i} = k'(e) \left(\frac{\frac{\partial \lambda_W(r_W, r_C, e^b)}{\partial r_i}}{\lambda_W(r_W, r_C, e^b)} + \frac{\frac{\partial^2 \kappa_W(e, r_W, r_C)}{\partial e \partial dr_i}}{\frac{\partial \kappa_W(e, r_W, r_C)}{\partial e}} \right)$$

With $\kappa_W(e, r_W, r_C) = v(e) - v(\bar{e})$ and $\lambda_W(r_W, r_C, e^b) = r_W - \bar{r}_W + \alpha[r_C - \bar{r}_C]$, we can state the following.

Observation 1 A change in r_W affects e as follows:

$$\frac{de}{dr_W} = \frac{1}{-\frac{\partial^2 U_W}{\partial e^2}} k'(e) \frac{1}{r_W - \bar{r}_W + \alpha[r_C - \bar{r}_C]}$$

Observation 2 If r_W is observable, we have

$$\frac{de}{dr_C} = \frac{1}{-\frac{\partial^2 U_W}{\partial e^2}} k'(e) \frac{\alpha}{r_W - \bar{r}_W + \alpha[r_C - \bar{r}_C]}$$

i.e. the coworker's wage rate r_C only increases effort if the altruism parameter is strictly positive, $\alpha > 0$.

Now, assume that r_W is unobservable and, thus, r_C may be used as a signal. We have now $\lambda_W(r_W, r_C, e^b) = r_W^s(r_C) - \bar{r}_W + \alpha[r_C - \bar{r}_C]$.

Observation 3 If r_W is not observable, we have

$$\frac{de}{dr_C} = \frac{1}{-\frac{\partial^2 U_W}{\partial e^2}} k'(e) \frac{\alpha + \frac{\partial r_W^s(r_C)}{\partial r_C}}{r_W - \bar{r}_W + \alpha [r_C - \bar{r}_C]}$$

i.e. the coworker's wage rate r_C increases effort if $\alpha + \frac{\partial r_W^s(r_C)}{\partial r_C} > 0$.

The above three effects are only comparable if the beliefs $r_W^s(r_C)$ are correct, i.e. $r_W^s(r_C) = r_W$. If this is the case, the observations above establish that

$$\frac{\frac{de}{dr_C} \Big|_{r_W \text{ observable}}}{\frac{de}{dr_W}} = \alpha \quad \text{and} \quad \frac{\frac{de}{dr_C} \Big|_{r_W \text{ unobservable}}}{\frac{de}{dr_W}} = \alpha + \frac{\partial r_W^s(r_C)}{\partial r_C}$$

This allows us to identify α and, in a second step, $\frac{\partial r_W^s(r_C)}{\partial r_C}$.

2.2 The experiment

At the beginning of each experimental session, each participant receives an initial endowment of 400 points (25 points = 1 Euro). The instructions²⁰ are handed out and read aloud by the experimenter. Afterward, open questions are answered and participants have to calculate two control tasks (available upon request). Then, participants are randomly assigned to the roles of manager, worker and coworker. These roles are kept throughout the whole experiment.

The core experiment consists of 30 (or 31) *periods*. At the beginning of each period, all participants are randomly and anonymously matched in triples, with each role (manager, worker, coworker) represented once. The manager sets the wages and, then, the worker sets her effort level. At the end of each period, the players are informed about their payoffs.

The manager sets a wage $r_i \in \{0, 1, \dots, 100\}$ for each worker i . The worker sets an effort level $e_i \in \{1, 2, \dots, 10\}$. Effort comes at a cost as depicted in table 2.1.²¹ The coworker is not able to choose effort (and, thus, to respond to wages set by the manager); her effort is fixed at $e_C = 4$.²²

Table 2.1: Cost-Effort-Function.

Effort e_i	1	2	3	4	5	6	7	8	9	10
Cost of effort $k(e_i)$	0	2	4	8	12	16	20	24	30	36

Payoffs are as follows. The worker or coworker i has a payoff of

$$\pi_i = r_i - k(e_i).$$

The manager has a payoff of

$$\pi_M = 15(e_W + e_C) - r_W - r_C.$$

²⁰An English translation is available upon request.

²¹The cost-effort function is similar to those used by e.g. Fehr et al. (1993), Fehr et al. (1998a), Gächter and Falk (2002), Maximiano et al. (2007), Rivas (2009), Siang et al. (2011), Bonein (2018), Brown et al. (2017) and Dariel and Riedl (2017).

²²The coworker, however, has to state an hypothetical effort that does not affect outcomes.

The effort levels $\mathbf{e} = (e_W, e_C)$ determine the size of the pie, while wages $\mathbf{r} = (r_W, w_C)$ distribute the pie between the players. The Pareto optimum is given by $\mathbf{e} = (10, 10)$ with arbitrary wage levels. The subgame perfect Nash equilibrium in a one-shot game with selfish players is $\mathbf{e} = (1, 1)$ and $\mathbf{r} = (0, 0)$. In the first period of each session (except for the first three sessions²³), workers are asked to use the strategy method (Selten, 1967) to indicate their effort for 21 wage levels in multiples of 5 (0,5,...,100) *before being informed about their actual wage*.²⁴

In the following 18 periods, the core experiment (the *baseline*) takes place. As outlined above, there are three treatments. In *OWNWAGE*, the worker only observes her own wage r_W . In *COWORKER*, the worker only observes the coworker’s wage r_C . In *BOTH*, she observes both wages. With the altruism motive, the coworker’s wage affects the worker’s choice in the *COWORKER* and *BOTH* treatments. With the signaling motive, the coworker’s wage is only relevant for the effort choice in the *COWORKER* treatment.

The *BOTH*, *OWNWAGE* and *COWORKER* treatments are played in a random order such that the manager is not able to differentiate treatments. This ensures that wage setting does not systematically vary across treatments. After this, treatments are slightly modified to perform some robustness checks (for another 8 to 12 periods, depending on the session) that will be explained in Sect. 4. Subsequently, all participants play a modified dictator game and an ultimatum game and fill in a questionnaire in order to elicit their other-regarding preferences.²⁵

Finally, each participant receives the payoff (in Euro) of one randomly selected round in addition to their initial endowment.

The experimental sessions took place in January, April and June 2018 at the University of Münster. The experiment and the questionnaire were programmed in z-Tree (Fischbacher, 2007). Participants were recruited among university students using the online recruitment system ORSEE (Greiner, 2015). 234 individuals participated in 13 sessions (42.7% female, 46.6% students of Business Administration or Economics, 23.1 average age). Each session lasted approximately 60 minutes. On average, individuals earned 17.10 Euro, with a minimum of 13.20 and a maximum of 21.40 Euro.

3 Results

In this section, we focus on data from the 18 periods of the *baseline* experiment.

²³Tables C.1 and C.2 in the Appendix give an overview of the exact settings played.

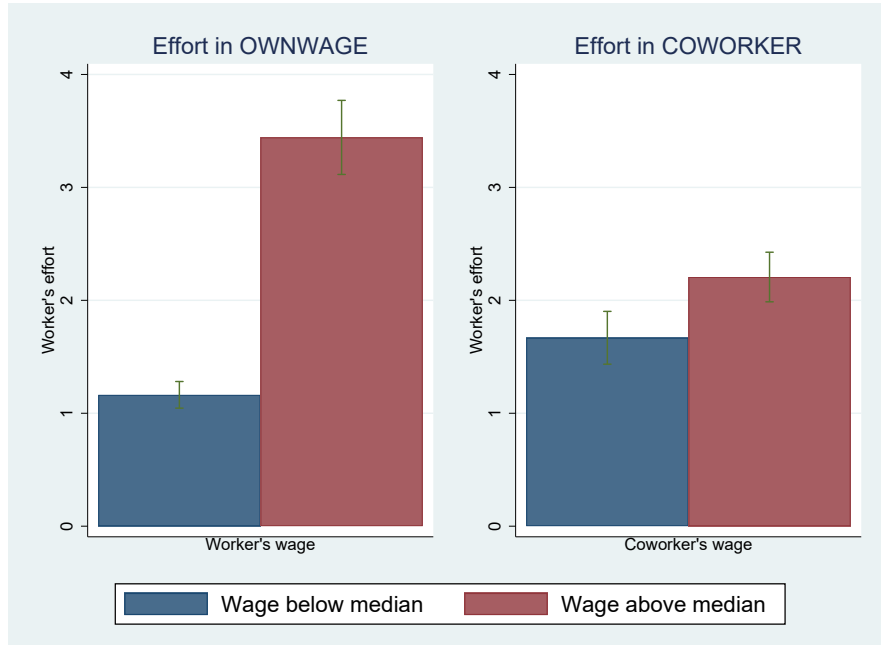
²⁴In the first four periods of session 1-3, we conducted a standard multi-player gift-exchange game to estimate these preferences on cooperative behavior.

²⁵The questionnaire covers personal information such as sex, age, field of study, questions on the complexity of the experiment and the participant’s ability, perceptions of fairness and reciprocity, and political opinions. Questions on perceptions of reciprocity and political opinions are adopted from TNS Infratest Sozialforschung (2015).

3.1 Descriptive evidence

We start by replicating a couple of results well-known from the literature. Figure 3.1 illustrates the evidence for the direct reciprocity pattern usually observed in gift-exchange games. Workers respond to higher wages by increasing their effort. The average effort for above median own wages (in *OWNWAGE*) is 3.44 ($SD = 2.55$) compared to 1.16 ($SD = 0.91$) for below median wages. The average wage for the worker (in *OWNWAGE*) is $r_W = 26.49$ ($SD = 28.65$), with a median of $r_W = 20$. In *COWORKER*, the coworker receives only $r_C = 13.52$ ($SD = 15.71$) with a median of $r_C = 10$.²⁶ This is in line with the manager’s (correct) anticipation that the coworker cannot retaliate. Moreover, there is evidence for indirect reciprocity in the treatment *COWORKER*. The worker rewards the manager for being kind to the coworker. The effort level is generally (and significantly) lower in this treatment, though. The average effort e_W for above median coworker wages is $e_W = 2.21$ ($SD = 1.79$) compared to 1.67 ($SD = 1.72$) for below median wages.²⁷

Figure 3.1: Average worker’s effort in *OWNWAGE* and *COWORKER*.



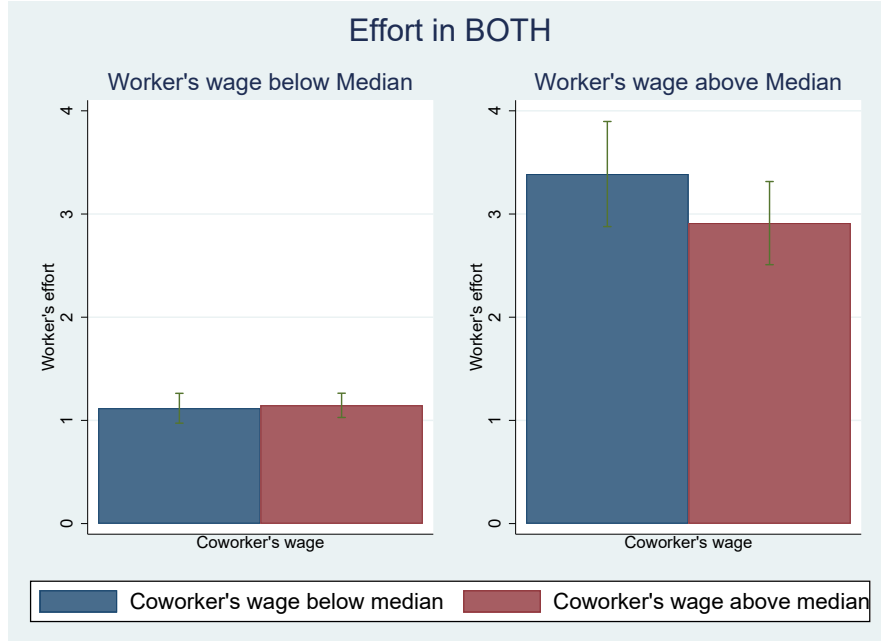
Note: The green lines depict the 95% confidence interval.

We may now analyze our main hypothesis. Is the coworker’s wage r_C merely a signal for the worker’s own wage? Or does it have an impact even when the own wage is observable? Figure 3.2 suggests that r_C is merely a signal. The figure uses observations from *BOTH* and controls for r_W first. Neither for r_W below the median wage nor for r_W above the median wage we observe a positive response toward r_C . If at all, a higher coworker’s wage seems to reduce the worker’s effort.

²⁶As usual, there is substantial heterogeneity in wage setting. 3 out of 78 managers always chose the zero lower bound for the worker, $r_W = 0$. 8 managers always chose the zero lower bound for the coworker, $r_C = 0$.

²⁷15 workers (19.2%) always selected the lowest possible effort of one.

Figure 3.2: Average worker’s effort in *BOTH*.



Note: The green lines depict the 95% confidence interval.

3.2 Regression analysis

In order to fully exploit the richness of our data, we now turn to regression analysis. The results are presented in table 3.1. We use pooled OLS with standard errors clustered at the session level as our preferred specification. Unobserved individual effects may partly explain the dependent variable e_W . However, since wages are set by the manager who neither knows the workers’ characteristics nor which treatment is played, the unobserved individual effect is supposed to be unrelated to the explanatory variables r_W and r_C . Hence, pooled OLS gives consistent estimates.

As columns (1) and (2) show, participants show directly and indirectly reciprocal behavior, depending on which wage is observable. An increase in r_W or r_C by 10 units implies an increase in effort by 0.48 or 0.32, respectively. An increase of one standard deviation of the wage rate r_i increases the effort by 1.39 and 0.50, respectively (which corresponds to 62.13 and 27.88 percent of the standard deviation of effort). The associated elasticities, evaluated at the sample means, are $\epsilon_{r_W} = 0.56$ and $\epsilon_{r_C} = 0.22$. That is, an increase in the wage rate r_W by 10 percent increases effort by 5.6 percent; an increase in r_C by 10 percent increases effort by 2.2 percent.²⁸ The coefficient on r_C in *COWORKER* gives us the estimate for $\frac{de}{dr_C} \Big|_{r_W \text{ unobservable}} = 0.0315$ as derived in the model, see Sect. 2.1.

²⁸As an aside, note that, on average, the manager does not recover the additional wage cost, since $d\pi = 15 \cdot de - dr_W < 0$ for $dr_W = 10$ and $de = \{0.48, 0.32\}$.

Table 3.1: Effect of wages on effort in *baseline*.

	Dependent variable: e_W			
	(1) <i>OWNWAGE</i>	(2) <i>COWORKER</i>	(3) <i>BOTH</i>	(4) <i>OWNWAGE</i> <i>COWORKER</i> <i>BOTH</i>
r_W	0.0484*** (0.0070)		0.0496*** (0.0091)	0.0484*** (0.0070)
r_C		0.0315*** (0.0073)	-0.0115 (0.0087)	0.0315*** (0.0073)
r_W \times <i>BOTH</i>				0.0012 (0.0086)
r_C \times <i>BOTH</i>				-0.0430*** (0.0121)
<i>OWNWAGE</i>				-0.0093 (0.0915)
<i>COWORKER</i>				0.5017*** (0.1270)
Constant	1.0265*** (0.1012)	1.5374*** (0.1369)	1.0357*** (0.0821)	1.0357*** (0.0821)
N	468	468	468	1404
Adj. R^2	0.3847	0.0757	0.4053	0.3164

Standard errors clustered at the session level.

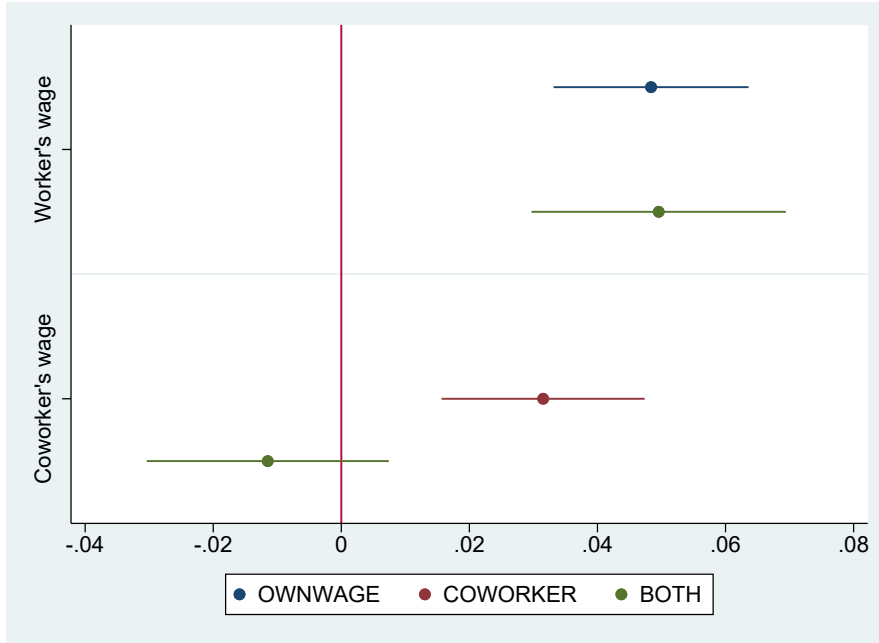
* $p < .1$, ** $p < .05$, *** $p < .01$.

Column (3) demonstrates that, if both wages are observable, only the own wage has an impact. $\frac{de}{dr_W} = 0.0496$ is highly significant, and the coefficient of the coworker's wage is negative and insignificant, i.e., $\frac{de}{dr_C} \Big|_{r_W \text{ observable}}$ is indistinguishable from zero. The worker does no longer reward kindness toward her coworker. This finding is in line with the idea that r_C is just a signal for the unobservable own wage in *COWORKER* – and that indirect reciprocity is *not* rooted in altruism.

In column (4), we present estimates in a sample that pools all three treatments. The coefficient on $r_C \times BOTH$ is negative and significant. Hence, the effect of r_C in *BOTH* is significantly lower than in *COWORKER*. This does not hold for the coefficient on $r_W \times BOTH$. The effect of r_W is not affected by making r_C observable as well.²⁹

As an illustration of our regression results, the confidence intervals of the point estimates shown in table 3.1 are presented in figure 3.3. The magnitude of the effect of r_W on e_W changes only marginally between *OWNWAGE* and *BOTH*. Standard errors and hence the confidence intervals remain roughly stable. For r_C the effect on e_W is smaller, even in *COWORKER*. The positive effect becomes negative and insignificant in *BOTH*. The new coefficient also does not coincide with the previous confidence interval.

Figure 3.3: Coefficients of table 3.1 with 95% confidence intervals.



The estimated regression models are further plotted in figure 3.4. The effect of r_W on e_W almost coincides for *OWNWAGE* and *BOTH*. For r_C , there is a significantly positive correlation in *COWORKER* and an insignificantly negative correlation in *BOTH*.

²⁹As a robustness check, we further run fixed- and random effects models to obtain more efficient estimates (see table A.2 in the Appendix). Coefficients and significance levels remain stable for any of these models.

The figure also suggests that for wages below 30 the worker exerts a higher effort in *COWORKER* than in *OWNWAGE*. For $r_W \leq 30$ in *OWNWAGE*, the average effort amounts to $e_W = 1.44$ ($SD = 1.30$), while we observe a significantly higher effort of $e_W = 1.81$ ($SD = 1.67$) for $r_C \leq 30$ in *COWORKER*. This might reflect the worker’s correct anticipation that her own (unobservable) wage is higher than the observable r_C in the *COWORKER* treatment. Regressing r_W on r_C , see table 3.2, shows that r_C serves as a reasonable predictor for r_W . It is on average 15.6 units larger than r_C .

Figure 3.4: Effect of wages on worker’s effort in *baseline*.

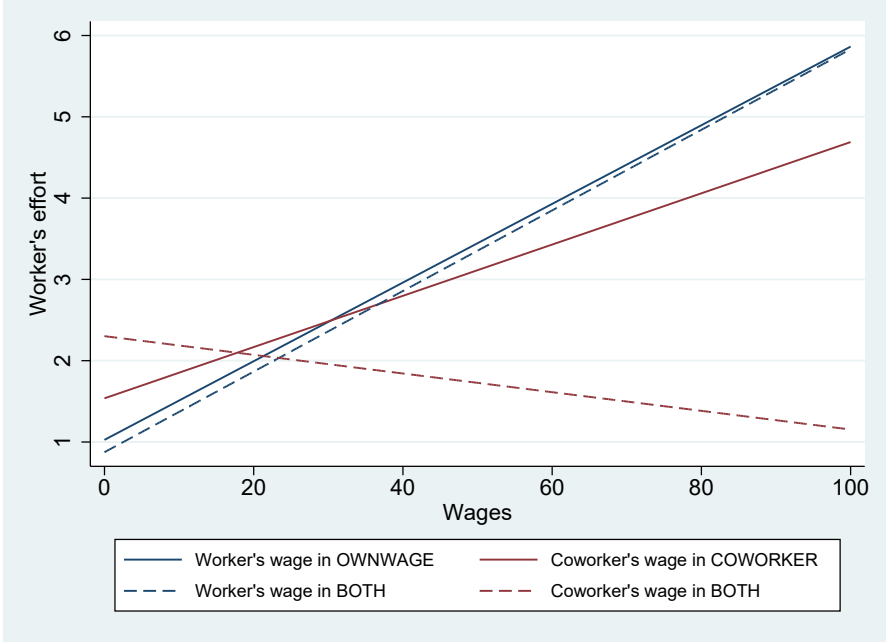


Table 3.2: Regression of r_W on r_C in *COWORKER*.

Dependent variable: r_W	
r_C	0.6199*** (0.1170)
Constant	15.5999*** (3.3569)
N	468
Adj. R^2	0.1204

Standard errors clustered at the session level.
 * $p < .1$, ** $p < .05$, *** $p < .01$.

We may now use the results to derive the structural parameters in the model in Sect. 2.1. As already pointed out, the estimate for α cannot be distinguished from zero. With $\alpha = 0$, the point estimate for the signal is then given by

$$\frac{\frac{de}{dr_C} \Big|_{r_W \text{ unobservable}}}{\frac{de}{dr_W}} = \alpha + \frac{\partial r_W^s(r_C)}{\partial r_C} \approx 0.8670$$

This point estimate suggests that, on average, workers interpret the coworker’s wage as a strong signal for their own wage. However, they expect a closer relationship between the coworker’s wage and their own wage than there is (see table 3.2).

The missing response of e_W to r_C in *BOTH* may be due to heterogeneity across participants, which is typical for studies on reciprocity (see e.g. Gächter and Thöni, 2010; Toussaert, 2017). We therefore look at subsamples of participants with certain characteristics.

As a first step, we focus on those worker type individuals who respond to r_C in *COWORKER* to see if they also show a non-response toward r_C in *BOTH*. We therefore consider all workers whose correlation coefficient between r_C and e_W is positive in the *COWORKER* treatment – which is true for 38 of 78 workers.³⁰ In table A.3 in the Appendix, we report the results of a separated analysis of these 38 participants and the non-indirect-reciprocal individuals. For none of the two groups we observe a positive effect of r_C on e_W in *BOTH* (see table A.3 in the Appendix).

In a second approach, we use the data from the strategy method applied in period 1 of sessions 4 through 13 in order to calculate the reciprocity index (RI) as suggested by Dariel and Riedl (2017).³¹ According to this measure, 78.3% of worker individuals behave directly reciprocally. Workers with above median *RI* show a stronger response toward increases in their own wage than those with below median *RI* (see table A.4 in the Appendix). However, workers with above median *RI* do not show a larger response to the coworker’s wage in the *COWORKER* treatment.³² For both subgroups (below and above median *RI*), r_C has no effect on e_W in *BOTH* where both wages are observable.

As a third approach, we use the data from the modified dictator game (MDG) and the ultimatum game (UG) played at the end of the experiment. These observations allow us to calculate measures of inequality aversion (Fehr and Schmidt, 1999), as proposed by Blanco et al. (2011). Basically, the model of inequality aversion states that, apart from their own material payoff, individuals also care about the distribution of payoffs among players. More specifically, individuals derive disutility from having a lower payoff (*envy*) or a higher payoff (*guilt*) than others. The Appendix B.2 provides more information how we use our data to calculate the indicators for envy and guilt. We consider the impact of wages on effort e_W for below and above median guilt-parameters and for below and above median envy-parameters. The results are presented in table A.5 and A.6 in the Appendix. For all of the four subgroups we observe a positive effect of r_W in *OWNWAGE* and a positive effect of r_C in *COWORKER*. Coefficients on r_W are higher for individuals above the respective median.

³⁰Following the classification of Gächter and Thöni (2010), 29 workers in our setting can be identified as unconditional defectors, i.e. they always choose $e_W = 1$. 11 individuals show a negative correlation of e_W and r_C , i.e. they may be called ‘envious’.

³¹The strategy method provides us with exactly one effort response to all possible wages. This allows calculating a Spearman’s rank correlation coefficient between effort and wages for each individual. Reciprocity is defined as a significant positive correlation at the 1%-level. In a second step, the index itself is calculated for reciprocal individuals to capture the magnitude of reciprocity. The calculation is explained in Appendix B.1. The index lies on the interval $RI \in [0, 1]$. $RI = 0$ stands for a completely selfish individual, who always chooses an effort of 1 irrespective of the own wage. $RI = 1$ is a total altruist, who always chooses the maximum possible effort ($e_W = 10$).

³²The coefficient of r_C becomes even smaller for more direct reciprocal workers, with the difference being insignificant, though.

The coefficient on r_C is close to zero in *BOTH* for all subgroups, except for workers with below median envy. These participants even seem to punish their managers for paying higher wages to the coworker (pointing to $\alpha < 0$).³³ Overall, however, we may conclude that these approaches do not yield convincing evidence for indirect reciprocity if own wages are observable.

4 Robustness checks and extensions

4.1 Equalizing complexity levels

One concern about the above results is that the treatments differ in complexity levels. Specifically, we compare the *COWORKER* treatment where one wage rate is displayed with the *BOTH* treatment with two wage rates. Due to e.g. attention constraints, participants may find it difficult to cope with two pieces of information and, therefore, opt to focus on just one wage rate: their own.

In order to rule out that attention constraints and other complexity related issues drive our findings, we consider the following robustness check (applied in sessions 4 through 13). We let the computer randomly determine $r_W \in \{10, 40, 60\}$.³⁴ The manager is informed about the computer's choice r_W and the treatment (*OWNWAGE* or *BOTH*) before she chooses r_C . *OWNWAGE* and *BOTH* are played with certainty for 6 periods each. We label this treatment '*exo*'.³⁵

Column (1) in table 4.1 shows that, with exogenous r_W , the worker still reacts positively to her wage when only her own (computer-set) wage can be observed. This finding may be interpreted as reflecting outcome-based reciprocity, since the worker cannot deduce the manager's *intentional* kindness from the computer-set wage rate. The coefficient is smaller than in *OWNWAGE* with endogenous wages (see column (1) of table 3.1) and of similar magnitude as the coefficient on r_C in *COWORKER* (see column(2) of table 3.1).

In the *BOTH* treatment with exogenous wages, the effect of r_W drops further compared to the regular *OWNWAGE* setting. At the same time, r_C positively affects effort (and significantly so when using Random or Individual Fixed Effects, see table A.7 in the Appendix). Hence, when r_C becomes observable, the worker seems to shift her attention from r_W toward r_C to assess the manager's kindness.³⁶ Column (3) reports the effects of both wages on effort between *exo* and *baseline*.³⁷ r_W has a significantly larger effect on effort in *baseline*; the effect of r_C is significantly smaller in *baseline*. The treatment suggests that workers do not neglect r_C in *BOTH* in *baseline* solely due to attention constraints, as both wages become significant if r_W is set exogenously (see column (2) and (4) in table A.7 in the Appendix).

³³This seems to be counterintuitive. It is in line, though, with Blanco et al. (2011) who also note that the envy-guilt model has limited predictive power at the individual level.

³⁴These values are the 25th-, 50th- and 75th percentiles of the wage rates set by the manager in sessions 1-3.

³⁵The changes in the experimental setting were explained in the instructions at the beginning of the experiment with a reminder just before the treatment under consideration.

³⁶Again, the correlation between r_W and r_C in this setting is neglectable ($\text{corr}(r_W, r_C) = 0.0658$), which is also stressed by figure A.1 in the Appendix. Even though the manager knows r_W before setting r_C , this does not significantly affect her decision about r_C . Hence, our standard errors in this setting do not become inflated due to multicollinearity.

³⁷In table A.9 in the Appendix, we also compare this setting to the setting with endogenous wages described in Sect. 4.2.

Table 4.1: Effect of setting r_W exogenously in *OWNWAGE* and *BOTH*.

	Dependent variable: e_W		
	(1) <i>OWNWAGE</i> - <i>exo</i>	(2) <i>BOTH</i> - <i>exo</i>	(3) <i>BOTH</i> - <i>baseline</i> \mathcal{E} - <i>exo</i>
r_W	0.0360*** (0.0050)	0.0259*** (0.0041)	0.0259*** (0.0040)
r_C		0.0144 (0.0085)	0.0144 (0.0084)
r_W \times <i>baseline</i>			0.0237** (0.0101)
r_C \times <i>baseline</i>			-0.0259** (0.0105)
<i>baseline</i>			0.3179 (0.1848)
Constant	0.9261*** (0.1167)	0.7179*** (0.1892)	0.7179*** (0.1869)
N	360	360	828
Adj. R^2	0.1323	0.1247	0.3123

Standard errors clustered at the session level.

* $p < .1$, ** $p < .05$, *** $p < .01$.

Note: Column (1) and (2) use observations from the treatments where *OWNWAGE* and *BOTH* are played with certainty and r_W is randomly chosen by a computer. Column (3) compares the results of column (2) with *BOTH* in *baseline*.

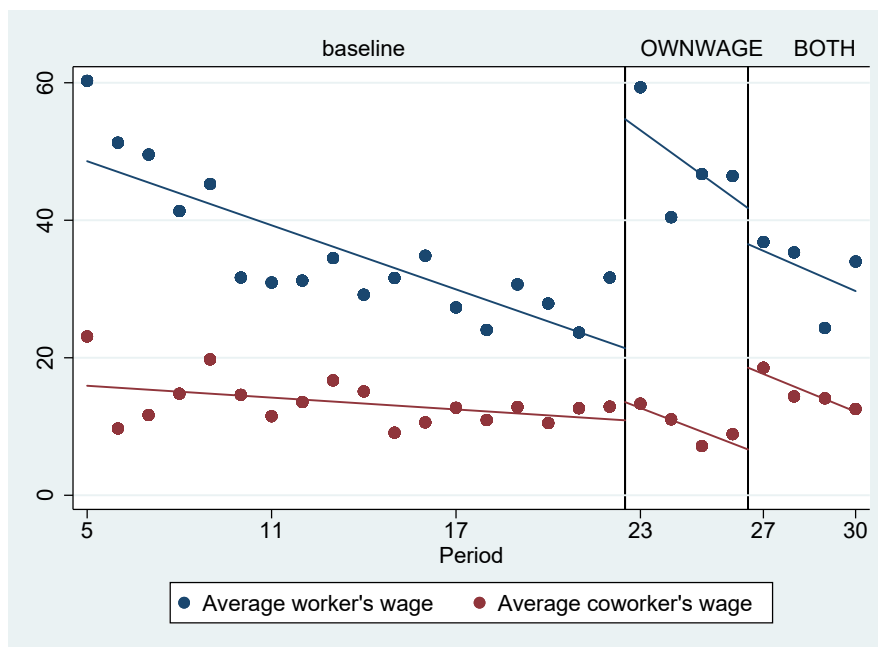
4.2 Wage setting

In the last eight periods of session 1-3, we test for the managers' expectations on the workers' reciprocal behavior. We let the participants play the regular *OWNWAGE* and *BOTH* treatment with certainty. When *OWNWAGE* is played with certainty, there is no incentive to set r_C above a minimum threshold that the manager deems appropriate, since r_C is not revealed to the worker. Hence, if r_C is higher in the *COWORKER* treatment, the manager must believe that r_C somehow affects e_W – as a signal for r_W or because the worker cares for the coworker's payoff. When *BOTH* is played with certainty, the worker knows her own wage and, thus, r_C is irrelevant if its only role is to serve as a signal.

The average wages of session 1-3 are depicted in figure 4.1. In periods 5 through 22, *OWNWAGE*, *COWORKER* and *BOTH* are played in random order (*baseline*). In periods 23 through 26, *OWNWAGE* is played with certainty as well as *BOTH* in periods 27 through 30. The increase in r_W from period 22 to 23 may indicate that managers adjust their wage setting to the increased probability that the worker observes her own wage (the probability increases from 2/3 to 1).

Similarly, r_C is significantly higher in *BOTH* ($r_C = 14.90$, $SD = 15.14$) than in *OWNWAGE* ($r_C = 10.11$, $SD = 12.28$). The increase indicates that managers expect the worker to react positively to their coworker's wage in *BOTH*. Thus, it seems that managers do not expect r_C to serve solely as a signal of r_W – although the above discussed data on the worker's behavior suggests it does. Managers seem to expect that the workers' indirect reciprocity is (partly) motivated by altruistic preferences.

Figure 4.1: Average wages in session 1-3.



5 Conclusion

This paper offers a theoretical framework that defines two potential motivations of indirect reciprocity: the altruism motive and the signaling motive. It presents evidence from a lab experiment that is designed to disentangle these two kinds of motivation. We find strong support for the signaling motive in the experiment. Participants exhibit indirect reciprocity only when they have no information about the manager's kindness toward themselves. If, however, direct interaction takes place, the manager's behavior toward the coworker becomes irrelevant. These findings suggest that participants take their coworker's wage as a signal of their own wage.

Our paper, therefore, helps to explain why there is robust evidence for indirect reciprocity in e.g. helping games, but mixed to little evidence for it in gift-exchange games. Indirect reciprocity vanishes as soon as the interaction allows for direct reciprocity.

References

- Abeler, J., Altmann, S., Kube, S., and Wibral, M. (2010). Gift Exchange and Workers' Fairness Concerns: When Equality Is Unfair. *Journal of the European Economic Association*, 8(6):1299–1324.
- Akerlof, G. A. (1982). Labor Contracts as Partial Gift Exchange. *The Quarterly Journal of Economics*, 97(4):543–569.
- Akerlof, G. A. and Yellen, J. L. (1990). The Fair Wage-Effort Hypothesis and Unemployment. *The Quarterly Journal of Economics*, 105(2):255–283.
- Alexander, R. D. (1987). *The biology of moral systems*. Foundations of human behavior. Aldine de Gruyter, New York.
- Axelrod, R. and Hamilton, W. D. (1981). The Evolution of Cooperation. *Science*, 211:1390–1396.
- Barta, Z. (2016). Individual variation behind the evolution of cooperation. *Philosophical Transactions Royal Society B*, 371(20150087):1–8.
- Barta, Z., McNamara, J. M., Huszár, D. B., and Taborsky, M. (2011). Cooperation among non-relatives evolves by state-dependent generalized reciprocity. *Proceedings. Biological sciences*, 278(1707):843–848.
- Bellemare, C. and Shearer, B. (2009). Gift giving and worker productivity: Evidence from a firm-level experiment. *Games and Economic Behavior*, 67(1):233–244.
- Berger, U. and Grüne, A. (2016). On the stability of cooperation under indirect reciprocity with first order information. *Games and Economic Behavior*, 98:19–33.
- Blanco, M., Engelmann, D., and Normann, H. T. (2011). A within-subject analysis of other-regarding preferences. *Games and Economic Behavior*, 72(2):321–338.
- Bolton, G. and Werner, P. (2016). The influence of potential on wages and effort. *Experimental Economics*, 19(3):535–561.
- Bolton, G. E. and Ockenfels, A. (2000). ERC: A Theory of Equity, Reciprocity, and Competition. *The American Economic Review*, 90(1):166–193.
- Bonein, A. (2018). Peer pressure and social comparisons with heterogeneous ability. *Managerial and Decision Economics*, 39:142–157.
- Bowles, S. and Gintis, H. (2004). The evolution of strong reciprocity: Cooperation in heterogeneous populations. *Theoretical Population Biology*, 65(1):17–28.
- Boyd, R. and Richerson, P. J. (1989). The evolution of indirect reciprocity. *Social Networks*, 11(3):213–236.

- Brown, C., Evans, J., Moser, D. V., and Presslee, A. (2017). How Does Reducing Pay Dispersion Affect Employee Behavior? *AAA 2017 Management Accounting Section (MAS) Meeting; 2017 Canadian Academic Accounting Association (CAAA) Annual Conference*.
- Charness, G. and Haruvy, E. (2002). Altruism, equity, and reciprocity in a gift-exchange experiment: an encompassing approach. *Games and Economic Behavior*, 40(2):203–231.
- Charness, G. and Levine, D. I. (2007). Intention and Stochastic Outcomes: An Experimental study. *The Economic Journal*, 117(522):1051–1072.
- Charness, G. and Rabin, M. (2002). Understanding Social Preferences with Simple Tests. *The Quarterly Journal of Economics*, 117(3):817–869.
- Cohn, A., Fehr, E., Herrmann, B., and Schneider, F. (2014). Social Comparison and Effort Provision: Evidence from a Field Experiment. *Journal of the European Economic Association*, 12(4):877–898.
- Cox, J. C. and Deck, C. A. (2005). On the Nature of Reciprocal Motives. *Economic Inquiry*, 43(3):623–635.
- Cox, J. C., Friedman, D., and Gjerstad, S. (2007). A tractable model of reciprocity and fairness. *Games and Economic Behavior*, 59(1):17–45.
- Dariel, A. and Riedl, A. (2017). Reciprocal Preferences and Gift-Exchange. *Working Paper*.
- Diekmann, A. (2004). The Power of Reciprocity: Fairness, Reciprocity, and Stakes in Variants of the Dictator Game. *The Journal of Conflict Resolution*, 48(4):487–505.
- Dufwenberg, M. and Kirchsteiger, G. (2000). Reciprocity and wage undercutting. *European Economic Review*, 44(4):1069–1078.
- Dufwenberg, M. and Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and Economic Behavior*, 47(2):268–298.
- Dufwenberg, M. and Kirchsteiger, G. (2019). Modelling kindness. *Journal of Economic Behavior & Organization*, 167:228–234.
- Engelmann, D. and Fischbacher, U. (2009). Indirect reciprocity and strategic reputation building in an experimental helping game. *Games and Economic Behavior*, 67(2):399–407.
- Enke, B. (2019). Kinship, Cooperation, and the Evolution of Moral Systems. *The Quarterly Journal of Economics*, 134(2):953–1019.
- Falk, A., Fehr, E., and Fischbacher, U. (2008). Testing theories of fairness—Intentions matter. *Games and Economic Behavior*, 62(1):287–303.
- Falk, A. and Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior*, 54(2):293–315.
- Fehr, E. and Gächter, S. (2000). Fairness and Retaliation: The Economics of Reciprocity. *Journal of Economic Perspectives*, 14(3):159–182.

- Fehr, E., Gächter, S., and Kirchsteiger, G. (1997). Reciprocity as a Contract Enforcement Device: Experimental Evidence. *Econometrica*, 65(4):833–860.
- Fehr, E., Kirchler, E., Weichbold, A., and Gächter, S. (1998a). When Social Norms Overpower Competition: Gift Exchange in Experimental Labor Markets. *Journal of Labor Economics*, 16(2):324–351.
- Fehr, E., Kirchsteiger, G., and Riedl, A. (1993). Does Fairness Prevent Market Clearing? An Experimental Investigation. *The Quarterly Journal of Economics*, 108(2):437–459.
- Fehr, E., Kirchsteiger, G., and Riedl, A. (1998b). Gift exchange and reciprocity in competitive experimental markets. *European Economic Review*, 42(1):1–34.
- Fehr, E. and Schmidt, K. M. (1999). A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics*, 114(3):817–868.
- Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.
- Fishmann, M. A., Lotem, A., and Stone, L. (2001). Heterogeneity Stabilizes Reciprocal Altruism Interactions. *Journal of Theoretical Biology*, 209(1):87–95.
- Gächter, S. and Falk, A. (2002). Reputation and Reciprocity: Consequences for the Labour Relation. *Scandinavian Journal of Economics*, 104(1):1–26.
- Gächter, S., Nosenzo, D., and Sefton, M. (2012). The Impact of Social Comparisons on Reciprocity. *Scandinavian Journal of Economics*, 114(4):1346–1367.
- Gächter, S., Nosenzo, D., and Sefton, M. (2013). Peer Effects in Pro-Social Behavior: Social Norms or Social Preferences? *Journal of the European Economic Association*, 11(3):548–573.
- Gächter, S. and Thöni, C. (2010). Social comparison and performance: Experimental evidence on the fair wage–effort hypothesis. *Journal of Economic Behavior & Organization*, 76(3):531–543.
- Geanakoplos, J., Pearce, D., and Stacchetti, E. (1989). Psychological games and sequential rationality. *Games and Economic Behavior*, 1(1):60–79.
- Gneezy, U. and List, J. A. (2006). Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments. *Econometrica*, 74(5):1365–1384.
- Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association*, 1(1):114–125.
- Greiner, B. and Levati, M. V. (2005). Indirect reciprocity in cyclical networks: An experimental study. *Journal of Economic Psychology*, 26(5):711–731.
- Gross, T., Guo, C., and Charness, G. (2015). Merit pay and wage compression with productivity differences and uncertainty. *Journal of Economic Behavior & Organization*, 117:233–247.
- Güth, W. (1995). An evolutionary approach to explaining cooperative behavior by reciprocal incentives. *International Journal of Game Theory*, 24(4):323–344.

- Heinz, M., Jeworrek, S., Mertins, V., Schumacher, H., and Sutter, M. (2017). Measuring Indirect Effects of Unfair Employer Behavior on Worker Productivity A Field Experiment. *MPI Collective Goods Preprint*, 22.
- Herne, K., Lappalainen, O., and Kestilä-Kekkonen, E. (2013). Experimental comparison of direct, general, and indirect reciprocity. *The Journal of Socio-Economics*, 45:38–46.
- Ito, T., Suzuki, R., and Arita, T. (2019). Evolution of four forms of reciprocity in the prisoner’s dilemma game. *Artificial Life and Robotics*, 24(2):140–146.
- Kube, S., Maréchal, M. A., and Puppe, C. (2012). The Currency of Reciprocity: Gift Exchange in the Workplace. *The American Economic Review*, 102(4):1644–1662.
- Leimar, O. and Hammerstein, P. (2001). Evolution of cooperation through indirect reciprocity. *Proceedures of the Royal Society London*, 268(1468):745–753.
- Levine, D. K. (1998). Modeling Altruism and Spitefulness in Experiments. *Review of Economic Dynamics*, 1(3):593–622.
- Lotem, A., Fishmann, M. A., and Stone, L. (1999). Evolution of cooperation between individuals. *Nature*, 400(6741):226–227.
- Maximiano, S., Sloof, R., and Sonnemans, J. (2007). Gift Exchange in a Multi-Worker Firm. *The Economic Journal*, 117(522):1025–1050.
- Moreno-Okuno, A. T. and Mosiño, A. (2017). A theory of sequential group reciprocity. *Latin American Economic Review*, 26(1).
- Nowak, M. A. and Sigmund, K. (1998). The dynamics of indirect reciprocity. *Journal of Theoretical Biology*, 194(4):561–574.
- Nowak, M. A. and Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, 437(7063):1291–1298.
- Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review*, 83(5):1281–1302.
- Reuben, E. and Riedl, A. (2013). Enforcement of contribution norms in public good games with heterogeneous populations. *Games and Economic Behavior*, 77(1):122–137.
- Rivas, M. F. (2009). Wage dispersion and workers’ effort. *Economics Bulletin*, 29(2):788–794.
- Seinen, I. and Schram, A. (2006). Social status and group norms: Indirect reciprocity in a repeated helping experiment. *European Economic Review*, 50:581–602.
- Selten, R. (1967). Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperimentes. *Beiträge zur experimentellen Wirtschaftsforschung. - Tübingen : Mohr. - (1967), p. 136-168.*
- Sethi, R. and Somanathan, E. (2001). Preference Evolution and Reciprocity. *Journal of Economic Theory*, 97(2):273–297.

- Siang, C. K., Requate, T., and Waichman, I. (2011). On the role of social wage comparisons in gift-exchange experiments. *Economics Letters*, 112(1):75–78.
- Tanaka, H., Ohtsuki, H., and Ohtsubo, Y. (2016). The price of being seen to be just: An intention signalling strategy for indirect reciprocity. *Proceedings of the Royal Society B: Biological Sciences*, 283(1835).
- TNS Infratest Sozialforschung (2015). SOEP 2015 – Erhebungsinstrumente 2015 (Welle 32) des Sozio-oekonomischen Panels: Personenfragebogen, Altstichproben.
- Toussaert, S. (2017). Intention-based reciprocity and signaling of intentions. *Journal of Economic Behavior & Organization*, 137:132–144.
- Wedekind, C. and Milinski, M. (2000). Cooperation Through Image Scoring in Humans. *Science*, 288(5467):850–852.

Appendix

A Results

A.1 Statistics

Table A.1: Median wages in *baseline*.

<i>OWNWAGE</i>	<i>COWORKER</i>	<i>BOTH</i>			
$r_W = 20$	$r_C = 10$	$r_W = 20$		$r_C = 10$	
		For $r_W < 20$:	For $r_W \geq 20$:	For $r_C < 10$:	For $r_C \geq 10$:
		$r_C = 0$	$r_C = 20$	$r_W = 0$	$r_W = 30$

A.2 Robustness checks *baseline*

Figure A.1: Distribution of r_C in *exo* for different values of r_W .



Table A.2: Effect of wages on effort in *baseline* – Random and Fixed Effects.

		Dependent variable: <i>ew</i>											
		Random Effects				Individual Fixed Effects				Session Fixed Effects			
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
		<i>OWNWAGE</i>	<i>COWORKER</i>	<i>BOTH</i>	<i>OWNWAGE</i>	<i>OWNWAGE</i>	<i>COWORKER</i>	<i>BOTH</i>	<i>OWNWAGE</i>	<i>OWNWAGE</i>	<i>COWORKER</i>	<i>BOTH</i>	<i>OWNWAGE</i>
					<i>COWORKER</i>				<i>COWORKER</i>				<i>COWORKER</i>
					<i>BOTH</i>				<i>BOTH</i>				<i>BOTH</i>
<i>r_W</i>		0.0448*** (0.0065)		0.0450*** (0.0067)	0.0429*** (0.0057)	0.0427*** (0.0064)		0.0426*** (0.0055)	0.0418*** (0.0054)	0.0423*** (0.0068)		0.0415*** (0.0060)	0.0407*** (0.0059)
<i>r_C</i>			0.0282*** (0.0063)	-0.0080 (0.0065)	0.0288*** (0.0067)		0.0271*** (0.0063)	-0.0065 (0.0056)	0.0282*** (0.0065)		0.0239*** (0.0062)	-0.0103 (0.0072)	0.0256*** (0.0062)
<i>BOTH</i>					-0.0018 (0.0073)				-0.0023 (0.0071)				-0.0002 (0.0078)
<i>BOTH</i>	$\times r_W$				-0.0339*** (0.0099)				-0.0322*** (0.0095)				-0.0347*** (0.0096)
<i>OWNWAGE</i>					0.0092 (0.0640)				0.0127 (0.0629)				-0.0046 (0.0942)
<i>COWORKER</i>					0.4126*** (0.1152)				0.3947*** (0.1176)				0.3830*** (0.0990)
Constant		1.1204*** (0.0990)	1.5820*** (0.1276)	1.1034*** (0.0719)	1.1610*** (0.0861)	1.1775*** (0.1689)	1.5975*** (0.0852)	1.1427*** (0.1225)	1.1876*** (0.1058)	1.1881*** (0.1796)	1.6403*** (0.0841)	1.2266*** (0.1384)	1.2344*** (0.1043)
<i>N</i>		468	468	468	1404	468	468	468	1404	468	468	468	1404
Adj. <i>R</i> ²						0.3595	0.0817	0.3772	0.2841	0.2831	0.0448	0.2888	0.2267

Standard errors clustered at the session level.

* $p < .1$, ** $p < .05$, *** $p < .01$.

A.3 Tests for heterogeneity

Table A.3: Comparing effect of wages on e_W for $corr(e_W, r_C)$.

		Dependent variable: e_W							
		$corr(e_W, r_C) \leq 0$				$corr(e_W, r_C) > 0$			
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		OWNWAGE	COWORKER	BOTH	BOTH	OWNWAGE	COWORKER	BOTH	BOTH
					COWORKER				COWORKER
r_W		0.0231*** (0.0075)		0.0308*** (0.0063)		0.0683*** (0.0055)		0.0610*** (0.0124)	
r_C			0.0005 (0.0073)	-0.0163 (0.0099)	0.0005 (0.0073)		0.0476*** (0.0109)	0.0012 (0.0118)	0.0476*** (0.0109)
BOTH	$\times r_C$				-0.0167** (0.0075)				-0.0464*** (0.0121)
BOTH	$\times r_W$				0.0308*** (0.0063)				0.0610*** (0.0124)
BOTH					-0.3721* (0.1825)				-0.8155*** (0.2471)
Constant		1.1700*** (0.1460)	1.5654*** (0.2177)	1.1933*** (0.1235)	1.5654*** (0.2181)	0.9006*** (0.1063)	1.6561*** (0.2285)	0.8406*** (0.0938)	1.6561*** (0.2290)
N		240	240	240	480	228	228	228	456
Adj. R^2		0.1312	-0.0042	0.2004	0.0927	0.6401	0.2150	0.5670	0.4430

Standard errors clustered at the session level.

* $p < .1$, ** $p < .05$, *** $p < .01$.

Table A.4: Comparing effect of wages on e_W for RI .

		Dependent variable: e_W							
		$RI \leq median(RI)$				$RI > median(RI)$			
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		<i>OWNWAGE</i>	<i>COWORKER</i>	<i>BOTH</i>	<i>BOTH</i> <i>COWORKER</i>	<i>OWNWAGE</i>	<i>COWORKER</i>	<i>BOTH</i>	<i>BOTH</i> <i>COWORKER</i>
r_W		0.0393*** (0.0104)		0.0392*** (0.0089)		0.0507*** (0.0077)		0.0449*** (0.0066)	
r_C			0.0388*** (0.0101)	0.0048 (0.0121)	0.0388*** (0.0101)		0.0218*** (0.0043)	-0.0038 (0.0126)	0.0218*** (0.0043)
<i>BOTH</i> $\times r_C$					-0.0340*** (0.0094)				-0.0256* (0.0119)
<i>BOTH</i> $\times r_W$				0.0392***				0.0449***	
<i>BOTH</i>									
<i>BOTH</i>									
<i>BOTH</i>									
Constant		1.1474*** (0.1488)	1.1291*** (0.1360)	0.8683*** (0.1156)	1.1291*** (0.1365)	1.0384*** (0.1434)	1.4799*** (0.1992)	1.0915*** (0.1652)	1.4799*** (0.1999)
N		144	144	144	288	138	138	138	276
Adj. R^2		0.2704	0.1697	0.3170	0.2496	0.3721	0.0585	0.3886	0.2788

Standard errors clustered at the session level.

* $p < .1$, ** $p < .05$, *** $p < .01$.

Table A.5: Comparing effect of wages on e_W for disadvantageous inequality.

		Dependent variable: e_W							
		$\alpha \geq median(\alpha)$				$\alpha < median(\alpha)$			
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)		
OWNWAGE	COWORKER	BOTH	BOTH	OWNWAGE	COWORKER	BOTH	BOTH	BOTH	COWORKER
r_W	0.0529*** (0.0074)	0.0485*** (0.0146)	0.0321*** (0.0101)	0.0456*** (0.0111)	0.0357** (0.0157)	0.0499*** (0.0068)	0.0357** (0.0158)	0.0357** (0.0158)	0.0357** (0.0158)
r_C		-0.0002 (0.0194)	0.0321*** (0.0101)						
BOTH $\times r_C$			-0.0323* (0.0159)						-0.0566*** (0.0127)
BOTH $\times r_W$			0.0485*** (0.0146)						0.0499*** (0.0068)
BOTH			-0.4952* (0.2501)						-0.4571** (0.1676)
Constant	0.9190*** (0.0853)	1.4725*** (0.2747)	0.9773*** (0.0768)	1.4725*** (0.2753)	1.0846*** (0.1694)	1.5155*** (0.2075)	1.0584*** (0.1291)	1.5155*** (0.2080)	1.5155*** (0.2080)
N	228	228	228	222	222	222	222	222	444
Adj. R^2	0.4645	0.1032	0.3883	0.2977	0.3473	0.0746	0.4432	0.4432	0.2614

Standard errors clustered at the session level.

* $p < .1$, ** $p < .05$, *** $p < .01$.

Table A.6: Comparing effect of wages on e_W for advantageous inequality.

		Dependent variable: e_W							
		$\beta > median(\beta)$				$\beta \leq median(\beta)$			
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		OWNWAGE	COWORKER	BOTH	BOTH	OWNWAGE	COWORKER	BOTH	BOTH
					COWORKER				COWORKER
r_W		0.0564*** (0.0066)		0.0484*** (0.0128)		0.0415*** (0.0092)		0.0518*** (0.0082)	
r_C			0.0494*** (0.0128)	-0.0140 (0.0128)	0.0494*** (0.0128)		0.0232*** (0.0047)	-0.0077 (0.0090)	0.0232*** (0.0048)
BOTH	$\times r_C$				-0.0634*** (0.0140)				-0.0309** (0.0108)
BOTH	$\times r_W$				0.0484*** (0.0128)				0.0518*** (0.0082)
BOTH					-0.2848 (0.1923)				-0.5999*** (0.1392)
Constant		0.9114*** (0.0742)	1.3589*** (0.2115)	1.0741*** (0.1221)	1.3589*** (0.2121)	1.1580*** (0.1800)	1.5164*** (0.1704)	0.9165*** (0.0734)	1.5164*** (0.1707)
N		198	198	198	396	234	234	234	468
Adj. R^2		0.5528	0.1802	0.4157	0.3212	0.2670	0.0518	0.4382	0.2882

Standard errors clustered at the session level.

* $p < .1$, ** $p < .05$, *** $p < .01$.

A.4 Setting r_W exogenously

Table A.7: Effect of setting r_W exogenously in *OWNWAGE* and *BOTH* - Random and Fixed Effects.

	Dependent variable: e_W											
	Random Effects				Individual Fixed Effects				Session Fixed Effects			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)			
	<i>OWNWAGE</i>	<i>BOTH</i>	<i>BOTH</i>	<i>OWNWAGE</i>	<i>BOTH</i>	<i>BOTH</i>	<i>OWNWAGE</i>	<i>BOTH</i>	<i>BOTH</i>			
r_W	0.0360*** (0.0050)	0.0258*** (0.0040)	0.0259*** (0.0040)	0.0360*** (0.0050)	0.0258*** (0.0040)	0.0259*** (0.0040)	0.0360*** (0.0050)	0.0260*** (0.0041)	0.0260*** (0.0041)			
r_C	0.0153** (0.0075)	0.0145* (0.0077)	0.0145* (0.0077)	0.0157* (0.0073)	0.0145* (0.0076)	0.0145* (0.0076)	0.0132 (0.0091)	0.0130 (0.0087)	0.0130 (0.0087)			
r_W \times <i>endo</i>			0.0179** (0.0076)			0.0157** (0.0064)			0.0156** (0.0067)			
r_C \times <i>endo</i>			-0.0220*** (0.0085)			-0.0206** (0.0080)			-0.0229** (0.0090)			
<i>endo</i>			0.3167* (0.1885)			0.3128 (0.1955)			0.3437* (0.1907)			
Constant	0.9261*** (0.1167)	0.7038*** (0.1701)	0.8125*** (0.2089)	0.9261*** (0.1844)	0.6990*** (0.2148)	0.7924*** (0.2296)	0.9261*** (0.1844)	0.7354** (0.2352)	0.8150*** (0.2451)			
N	360	360	828	360	360	828	360	360	828			
Adj. R^2				0.2581	0.2157	0.3029	0.1394	0.1235	0.2304			

Standard errors clustered at the session level.

* $p < .1$, ** $p < .05$, *** $p < .01$.

Note: Column (1), (2), (4), (5), (7) and (8) use observations from the treatments where *OWNWAGE* and *BOTH* are played with certainty and r_W is randomly chosen by a computer. Column (3), (6) and (9) compare the results of column *BOTH- exo* with *BOTH* in *baseline*.

Table A.8: Effect of setting r_W exogenously in OWNWAGE.

		Dependent variable: ew							
POLS		Random Effects		Individual Fixed Effects		Session Fixed Effects			
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)		
<i>OWNWAGE</i> (<i>exo, baseline</i>)	<i>OWNWAGE</i> <i>BOTH</i> (<i>exo</i>)	<i>OWNWAGE</i> (<i>exo, baseline</i>)	<i>OWNWAGE</i> <i>BOTH</i> (<i>exo</i>)	<i>OWNWAGE</i> (<i>exo, baseline</i>)	<i>OWNWAGE</i> <i>BOTH</i> (<i>exo</i>)	<i>OWNWAGE</i> (<i>exo, baseline</i>)	<i>OWNWAGE</i> <i>BOTH</i> (<i>exo</i>)	<i>OWNWAGE</i> (<i>exo, baseline</i>)	<i>OWNWAGE</i> <i>BOTH</i> (<i>exo</i>)
r_W	0.0360*** (0.0050)	0.0360*** (0.0050)	0.0360*** (0.0050)	0.0360*** (0.0050)	0.0360*** (0.0050)	0.0360*** (0.0050)	0.0360*** (0.0050)	0.0360*** (0.0050)	0.0360*** (0.0050)
r_W \times <i>baseline</i>	0.0123 (0.0078)	0.0090 (0.0066)	0.0077 (0.0063)	0.0073 (0.0071)	0.0073 (0.0071)	0.0073 (0.0071)	0.0073 (0.0071)	0.0073 (0.0071)	0.0073 (0.0071)
<i>baseline</i>	0.1004 (0.1348)	0.1378 (0.1321)	0.1504 (0.1390)	0.1600 (0.1443)	0.1600 (0.1443)	0.1600 (0.1443)	0.1600 (0.1443)	0.1600 (0.1443)	0.1600 (0.1443)
r_W \times <i>BOTH</i>	-0.0102* (0.0046)	-0.0102** (0.0045)	-0.0102** (0.0045)	-0.0102** (0.0045)	-0.0102* (0.0045)	-0.0102* (0.0045)	-0.0102* (0.0047)	-0.0102* (0.0047)	-0.0102* (0.0047)
r_C \times <i>BOTH</i>	0.0144 (0.0085)	0.0151** (0.0061)	0.0151** (0.0061)	0.0152** (0.0059)	0.0152** (0.0059)	0.0152** (0.0059)	0.0132 (0.0086)	0.0132 (0.0086)	0.0132 (0.0086)
<i>BOTH</i>	-0.2082 (0.1539)	-0.2192* (0.1227)	-0.2192* (0.1227)	-0.2192* (0.1227)	-0.2207* (0.1199)	-0.2207* (0.1199)	-0.1897 (0.1603)	-0.1897 (0.1603)	-0.1897 (0.1603)
Constant	0.9261*** (0.1153)	0.9261*** (0.1169)	0.9760*** (0.1403)	0.9261*** (0.1169)	0.9678*** (0.1698)	0.9261*** (0.1584)	0.9685*** (0.1708)	0.9261*** (0.1580)	0.9261*** (0.1580)
N	828	720	828	720	720	828	720	828	720
Adj. R^2	0.2864	0.1346	0.3104	0.2189	0.2367	0.2367	0.2367	0.2367	0.1364

Standard errors clustered at the session level.

* $p < .1$, ** $p < .05$, *** $p < .01$.

Note: The table compares *OWNWAGE-exo* with *OWNWAGE* in *baseline* (odd column numbers) and *BOTH-exo* (even column numbers).

Table A.9: Effect of wages on effort if *BOTH* is played with certainty.

Dependent variable: e_W						
	POLS		Random Effects		Indiv. Fixed Effects	Session Fixed Effects
	(1)	(2)	(3)	(4)	(5)	(6)
	<i>endo</i>	<i>exo, endo</i>	<i>endo</i>	<i>exo, endo</i>	<i>endo</i>	<i>endo</i>
r_W	0.0135 (0.0086)	0.0135* (0.0073)	0.0163 (0.0114)	0.0168* (0.0100)	0.0186 (0.0140)	0.0046 (0.0070)
r_C	0.0574*** (0.0056)	0.0574*** (0.0047)	0.0416*** (0.0053)	0.0386*** (0.0055)	0.0276 (0.0123)	0.0566** (0.0059)
r_W $\times exo$		0.0124 (0.0083)		0.0090 (0.0108)		
r_C $\times exo$		-0.0430*** (0.0096)		-0.0232** (0.0093)		
<i>exo</i>		-0.3183 (0.2901)		-0.5055* (0.2932)		
Constant	1.0362* (0.2628)	1.0362*** (0.2214)	1.1817*** (0.2808)	1.2096*** (0.2398)	1.3139* (0.3424)	1.3412** (0.3066)
N	72	432	72	432	72	72
Adj. R^2	0.2844	0.1665			0.2125	0.1893

Standard errors clustered at the session level.

* $p < .1$, ** $p < .05$, *** $p < .01$.

Note: The table uses treatments where *BOTH* is played with certainty and r_W is chosen by the manager. In column (2) and (4) this is compared to treatments where *BOTH* is played with certainty and r_W is randomly determined by the computer. Note that this comparison cannot be run with Fixed Effects models as the dummy *exo* does not vary within individuals or within sessions.

Surprisingly, r_W becomes insignificant if *BOTH* is played with certainty and both wages are set by the manager endogenously in column (1). Instead, r_C has a significant positive effect on e_W . This contradicts previous findings. There is no plausible explanation why workers should react stronger toward their coworker's wage if *BOTH* is played with certainty. Coefficients of this specification are also highly sensitive to introducing fixed and random effects (column (3), (5) and (6)). We cannot rule out that this result is rooted in the low number of observations ($N = 72$) or the fact that this setting is played only in the last four periods of session 1-3 where we observe a severe downwards trend that might bias our results (see figure 4.1).

Table A.10: Effect of r_W on effort if *OWNWAGE* is played with certainty.

Dependent variable: e_W						
	POLS		Random Effects		Indiv. Fixed Effects	Session Fixed Effects
	(1)	(2)	(3)	(4)	(5)	(6)
	<i>endo</i>	<i>exo, endo</i>	<i>endo</i>	<i>exo, endo</i>	<i>endo</i>	<i>endo</i>
r_W	0.0538*	0.0538***	0.0425**	0.0433***	0.0379	0.0436
	(0.0165)	(0.0139)	(0.0192)	(0.0162)	(0.0196)	(0.0173)
r_W \times <i>exo</i>		-0.0178		-0.0073		
		(0.0148)		(0.0169)		
<i>exo</i>		-0.0339		-0.5391		
		(0.3661)		(0.4862)		
Constant	0.9600	0.9600**	1.5043***	1.4652***	1.7294	1.4521
	(0.4102)	(0.3474)	(0.5691)	(0.4723)	(0.9430)	(0.8343)
N	72	432	72	432	72	72
Adj. R^2	0.3342	0.2289			0.2537	0.2144

Standard errors clustered at the session level.

* $p < .1$, ** $p < .05$, *** $p < .01$.

Note: The table uses treatments where *OWNWAGE* is played with certainty and r_W is chosen by the manager. In column (2) and (4) this is compared to treatments where *OWNWAGE* is played with certainty and r_W is randomly determined by the computer. Note that this comparison cannot be run with Fixed Effects models as the dummy *exo* does not vary within individuals or within sessions.

B Indices

B.1 Reciprocity Index

The reciprocity index by Dariel and Riedl (2017) measures the extent of the employees reciprocation. It is defined as

$$RI_i = \frac{\sum_{k=0}^M (e_{ki} - \underline{e})}{M(\bar{e} - \underline{e})}. \quad (2)$$

e_{ki} is the effort individual i chooses for wage $w = k$, $\underline{e} = 1$ is the minimum effort, $\bar{e} = 10$ is the maximum effort. M is the number of possible wages - thus, the number of decisions an employee has to make ($M = 21$ in our setting).

B.2 Inequality aversion

To control for the individual degree of inequality aversion, participants were solicited to indicate their preferences in a modified dictator game (MDG) and an ultimatum game (UG) at the end of the experiment. Results are used to calculate parameters of guilt and envy as suggested by Blanco et al. (2011). They enter the following utility function U_i of individual i :

$$U_i(\mathbf{x}) = x_i - \alpha_i \max\{x_j - x_i, 0\} - \beta_i \max\{x_i - x_j, 0\}, \quad i \neq j. \quad (3)$$

The utility function was introduced by Fehr and Schmidt (1999). Individuals derive positive utility from their own payoff x_i . The difference between their own and their opponent's payoff x_j enters the utility function negatively. The second term of the equation $\alpha_i \max\{x_j - x_i, 0\}$ captures disutility from disadvantageous inequality ($x_j > x_i$), the third term $\beta_i \max\{x_i - x_j, 0\}$ captures disutility from advantageous inequality ($x_i > x_j$).

B.2.1 Disadvantageous inequality

To estimate the envy parameter α for disadvantageous inequality, answers from the UG (figure B.1) are used. In the UG, individuals are offered eleven payoff allocations between player A and B. For each of them they had to decide whether to accept or reject the offer. Allocations ranged from (10, 0) over (5, 5) to (0, 10). In case of rejection both participants get a zero payoff.

Figure B.1: Screenshot of the Ultimatum Game.

Geben Sie für jede Verteilung, die Spieler A vorschlagen kann, an, ob Sie diese als Spieler B annehmen oder ablehnen.
Wir erinnern Sie daran, dass, sofern Sie annehmen, sowohl Person A als auch Person B den entsprechenden Betrag erhalten. Sofern sie ablehnen, erhalten beide 0.

Entscheidung	Aufteilung (Gewinn A - Gewinn B)	Ihre Wahl
Entscheidung 1	10 - 0	akzeptieren <input type="radio"/> ablehnen <input type="radio"/>
Entscheidung 2	9 - 1	akzeptieren <input type="radio"/> ablehnen <input type="radio"/>
Entscheidung 3	8 - 2	akzeptieren <input type="radio"/> ablehnen <input type="radio"/>
Entscheidung 4	7 - 3	akzeptieren <input type="radio"/> ablehnen <input type="radio"/>
Entscheidung 5	6 - 4	akzeptieren <input type="radio"/> ablehnen <input type="radio"/>
Entscheidung 6	5 - 5	akzeptieren <input type="radio"/> ablehnen <input type="radio"/>
Entscheidung 7	4 - 6	akzeptieren <input type="radio"/> ablehnen <input type="radio"/>
Entscheidung 8	3 - 7	akzeptieren <input type="radio"/> ablehnen <input type="radio"/>
Entscheidung 9	2 - 8	akzeptieren <input type="radio"/> ablehnen <input type="radio"/>
Entscheidung 10	1 - 9	akzeptieren <input type="radio"/> ablehnen <input type="radio"/>
Entscheidung 11	0 - 10	akzeptieren <input type="radio"/> ablehnen <input type="radio"/>

OK

According to the model of inequality aversion, subjects are expected to reject highly unequal offers, and thus prefer a zero payoff for both players over an unequal payoff distribution where they are worse off than their counterpart. To define the degree of inequality aversion, the point, where individuals switch from rejection to acceptance in the ultimatum game, is sought. For the first 5 decisions player B is behind, hence according to equation (3) B's utility reads $U_B = \pi_B - \alpha_B(10 - \pi_B - \pi_B)$. Player A's payoff is denoted by $\pi_A = 10 - \pi_B$. At the switching point $\pi_B = s_B$, the utility of acceptance is equal to the utility of rejection and receiving a zero payoff: $U_B = s_B - \alpha_B(10 - s_B - s_B) = 0$. Thus, the envy parameter can be defined as $\alpha_B = \frac{s_B}{10 - 2s_B}$. The switching point is approximated as the average between the highest offer rejected and the lowest offer accepted, which equals the lowest acceptable offer minus 0.5.

Note that in order to calculate α , offers above 5 do not need to be taken into account since this constitutes advantageous inequality.

Individuals who did not accept all offers below (5, 5) are assumed to have a maximal aversion against disadvantageous inequality which is $\alpha = 4.5$. On the other hand, subjects who accepted any offer, do not seem to have a preference for equality and hence $\alpha = 0$.

Any individual with inconsistent preferences, is ruled out from the analysis. This applies to individuals who accept an unequal distribution but reject another distribution with less inequality and higher payoff for themselves.

B.2.2 Advantageous inequality

Similarly, β is estimated to measure the refusal of advantageous inequality. We use the switching point of the MDG (figure B.2). In the MDG, subjects have to make eleven decisions between two payoff distributions. In each of the eleven choices, they have to decide between an unequal distribution where they receive 10 and their partner receives 0 and an equal split ranging from (0, 0) in the first decision to (10, 10) in the last decision.

Figure B.2: Screenshot of the modified Dictator Game.

Stellen Sie sich vor, dass Sie (Spieler A) zusammen mit einem anderen Spieler (Spieler B) in einer Gruppe sind.
Tun Sie für jede der 11 folgenden Entscheidungen so, als ob Sie Person A wären und wählen Sie die linke oder die rechte Verteilung.

Entscheidung	Gewinn A	Gewinn B	Ihre Wahl	Gewinn A	Gewinn B
Entscheidung 1	10	0	links <input type="radio"/> rechts <input type="radio"/>	0	0
Entscheidung 2	10	0	links <input type="radio"/> rechts <input type="radio"/>	1	1
Entscheidung 3	10	0	links <input type="radio"/> rechts <input type="radio"/>	2	2
Entscheidung 4	10	0	links <input type="radio"/> rechts <input type="radio"/>	3	3
Entscheidung 5	10	0	links <input type="radio"/> rechts <input type="radio"/>	4	4
Entscheidung 6	10	0	links <input type="radio"/> rechts <input type="radio"/>	5	5
Entscheidung 7	10	0	links <input type="radio"/> rechts <input type="radio"/>	6	6
Entscheidung 8	10	0	links <input type="radio"/> rechts <input type="radio"/>	7	7
Entscheidung 9	10	0	links <input type="radio"/> rechts <input type="radio"/>	8	8
Entscheidung 10	10	0	links <input type="radio"/> rechts <input type="radio"/>	9	9
Entscheidung 11	10	0	links <input type="radio"/> rechts <input type="radio"/>	10	10

OK

In this setting, participants are assigned to the role of player A. The switching point s_A , where they are indifferent between an equal split e.g., (6, 6) and the payoff allocation (10, 0) yields the guilt parameter β . At this point utility from equation (3) reads as $U_A = 10 - \beta(10 - 0) = s_A$ and thus $\beta = 1 - \frac{s_A}{10}$. Again, the switching point is the lowest egalitarian allocation accepted minus 0.5.

Individuals who always chose the equal split are assigned the maximum guilt parameter $\beta = 1$, while individuals who always prefer (10, 0) have $\beta = 0$ since they do not show any concern for their partner's outcome.

Furthermore, any individual with inconsistent preferences was excluded from the analysis. In this case this corresponds to individuals that prefer an egalitarian outcome $\pi_A = \pi_B$ over (10, 0) but prefer (10, 0) over another equal offer with higher payoffs.

C Overview experiment

Table C.1: Overview on experimental setting in session 1-3.

	<i>Intro</i>	<i>baseline</i>	<i>certain</i>	<i>certain</i>
Rounds	4	18	4	4
Wages	$r_W \in [0, 100]$ $r_C \in [0, 100]$	$r_W \in [0, 100]$ $r_C \in [0, 100]$	$r_W \in [0, 100]$ $r_C \in [0, 100]$	$r_W \in [0, 100]$ $r_C \in [0, 100]$
Effort	$e_W \in [1, 10]$ $e_C \in [1, 10]$	$e_W \in [1, 10]$ $e_C = 4$	$e_W \in [1, 10]$ $e_C = 4$	$e_W \in [1, 10]$ $e_C = 4$
Treatment	<i>BOTH</i>	<i>BOTH</i> <i>OWNWAGE</i> <i>COWORKER</i>	<i>OWNWAGE</i>	<i>BOTH</i>
Treatments played with certainty	yes	no	yes	yes

BOTH: employees observe both wages

OWNWAGE: employees observe their own wages

COWORKER: employees observe coworker's wages

Table C.2: Overview on experimental setting in session 4-13.

	<i>Intro</i>	<i>baseline</i>	<i>certain</i>	<i>certain</i>
Rounds	1	18	6	6
Wages	$r_W \in [0, 100]$ $r_C \in [0, 100]$	$r_W \in [0, 100]$ $r_C \in [0, 100]$	$r_W \in [10, 40, 60]$ (exogenous) $r_C \in [0, 100]$	$r_W \in [10, 40, 60]$ (exogenous) $r_C \in [0, 100]$
Effort	Strategy method $e_W \in [1, 10]$ $e_C \in [1, 10]$	$e_W \in [1, 10]$ $e_C = 4$	$e_W \in [1, 10]$ $e_C = 4$	$e_W \in [1, 10]$ $e_C = 4$
Treatment	<i>OWNWAGE</i>	<i>BOTH</i> <i>OWNWAGE</i> <i>COWORKER</i>	<i>OWNWAGE</i>	<i>BOTH</i>
Treatments played with certainty	yes	no	yes	yes

BOTH: employees observe both wages

OWNWAGE: employees observe their own wages

COWORKER: employees observe coworker's wages

Note: The order of the treatments *baseline* and *certain* varied between sessions.