‘Kinked Conformism’ in Voluntary Cooperation*

VERY PRELIMINARY VERSION
2 June 2008

Christian Thöni, University of St. Gallen[a]

Simon Gächter, University of Nottingham[b], CESifo & IZA

ABSTRACT:

Field evidence suggests that people’s work morale depends on those of their co-workers. We are interested in understanding the behavioural logic of such “social interaction effects”. We design a three-person gift-exchange experiment where workers can observe each others’ effort levels and adjust their own effort if they so wish. Our design avoids the “reflection problem” (Manski (1993)) which plagues identification of social interaction effects. The one-shot nature of our experiment also excludes any strategic incentives for changing efforts. We nevertheless observe many adjustments, most of them down to the lower effort level. We show that efforts are strategic complements rather than strategic substitutes as predicted by various models of social preferences. Conformism seems to be the dominant motive behind the social interaction effects we observe.

* Acknowledgments: This paper is part of a research project “Soziale Interaktionen, Unternehmenskultur and Anreizgestaltung”, financed by the Grundlagenforschungsfonds of the University of St. Gallen. We are grateful for very helpful comments by Martin Kolmar, Tatiana Kornienko, Bentley MacLeod, Rupert Sausgruber, Karl Schlag, Jean-Robert Tyran, as well as by the participants of the Verein für Socialpolitik conference in Bonn, the meeting of the Swiss Society of Economics and Statistics, the IZA Bonn Workshop on Behavioural Labour Economics, the ESA meeting in Rome, and research seminar participants in Dallas, Edinburgh, Innsbruck, Jena, Lyon, Nottingham, Canberra, Melbourne, and Stockholm.

[a] FEW-HSG, Varnbüelstrasse 14, CH-9000 St. Gallen; christian.thoeni@unisg.ch.

[b] School of Economics, Sir Clive Granger Building, University Park, Nottingham, NG7 2RD, United Kingdom; simon.gaechter@nottingham.ac.uk.

1. Introduction

Field (experimental) evidence suggests that an employee's work morale depends strongly on the work morale of his or her co-workers (Ichino and Maggi (2000); Falk and Ichino (2006); Mas and Moretti (2007); Bellemare and Shearer (2007). Identifying such "social interaction effects" with field data requires overcoming the "reflection problem" (Manski (1993); Manski (2000)) which results from the fact that workers mutually influence one another in their work morale. Approaches working with field data try to avoid the reflection problem by controlling for as many co-variates as possible and by finding appropriate instruments (see for instance Ichino and Maggi (2000) 2000; Mas and Moretti (2007), Bandiera, Barankay and Rasul (2007)). Field experimental approaches (e.g., Falk and Ichino (2006) 2006; Bellemare and Shearer (2007)) can avoid many of the identification problems of field data, but, like the field data, are too coarse for a detailed understanding of the behavioural logic behind social interaction effects.

This paper complements the field (experimental) literature by shedding light on the behavioral logic behind social interaction effects in a situation where we can fully eliminate the reflection problem by experimental design and therefore give a *causal* interpretation to social interaction effects. Part of our behavioral insights will come from a confrontation of our data with detailed theoretical predictions as derived from recent theories of social preferences.

Our goals strongly suggest a laboratory experimental test since field (experimental) data are usually either not detailed or not controlled enough for our purposes.[1] The framework for our analysis will be a new three-player version of the original two-player gift-exchange game (Fehr, Kirchsteiger and Riedl (1993)).[2] Our version consists of a principal and two workers, which is the minimal setup if one wants to allow for effort comparisons. There are two reasons for why we chose to base our design on the gift exchange game. First, this game has been developed for understanding labor relationships as we have them in mind here (Fehr, Gächter and Kirchsteiger (1997); Fehr, Kirchler, Weichbold and Gächter (1998); Charness (2004) Fehr and Gächter (2000)). Second, the theories of social preferences we will look at later in the paper can all explain the experimental evidence from the two-player gift exchange game. It is therefore a natural question whether these theories can also rationalize behavior in the three-person gift exchange game or whether one has to go beyond current theories of social preferences to understand social interaction effects in our context.

We design our one-shot three-person gift exchange game such that we are able to identify the *marginal effect of social interaction*. The main ideas of our design are as follows. An

[1] There is a literature addressing social interaction effects in public goods games. Falk, Fischbacher and Gächter (2004) study the behaviour of subjects who play in two public goods games in separate groups. Bardsley and Sausgruber (2005) propose a design able to distinguish between reciprocity and conformity as motives for determining contributions.

[2] The original gift exchange game is a two-player game in which a principal pays his worker a wage and the worker responds with a costly effort choice. Incentives are such that efforts should always be minimal, irrespective of the wage the principal pays. The results of numerous gift exchange experiments refuted this prediction and show that effort and wages are positively correlated (Fehr and Gächter (2000)).

employer is paired with two workers. The workers are identical, i.e. they have the same productivity and effort-cost function. The employer pays an identical wage to both of his or her workers. The workers learn their wage and know that the other worker receives the same wage. The workers then choose their efforts simultaneously. After that, both workers learn the effort of their co-worker and are given the opportunity to *reconsider* their effort choice. The workers are told that only one randomly determined effort will effectively be changed, if the respective worker so wishes. We call the first effort choice the *initial effort* which by construction is independent of the co-worker's effort choice. The second effort choice is called the *revised effort.* It identifies the marginal influence of "social interaction", that is, the change in effort as *caused* by being informed about the co-player's effort and holding the co-player's effort constant.

Our main hypothesis, the "Social Interaction Hypothesis", is inspired by the field (experimental) evidence cited above. It predicts that the effort revisions can be explained by the difference between the two workers' efforts: seeing that the other worker chose a higher effort than herself will induce a worker to increase her own effort and vice versa.

Our results support the Social Interaction Hypothesis unambiguously, as the workers' effort choices are highly significantly positively correlated. The effort adjustments are asymmetric, however: An agent who observes that his co-worker has chosen a lower effort induces workers to reduce their effort significantly, while they increase their efforts much more reluctantly if they learn that the co-worker has provided a higher effort. Thus, the first main contribution of this paper is to demonstrate a social interaction effect in a one-shot experimental design that eliminates the reflection problem. We can therefore interpret effort revisions as being *caused* by the mere observation of the co-worker's effort, in the absence of any strategic reason to adjust efforts.

Our second contribution is to confront our results with predictions from theories of social preferences. The theories we look at are models of inequality aversion (Fehr and Schmidt (1999); Bolton and Ockenfels (2000)), reciprocity (Dufwenberg and Kirchsteiger (2004)), type-based reciprocity (Levine (1998)) and three models that combine interpersonal payoff comparisons and intentionality (Falk and Fischbacher (2006); Charness and Rabin (2002); Cox, Friedman and Gjerstad (2007)). We apply these theories because our setup is a well-defined game and these theories have been developed to explain behaviour in games. Moreover, all these theories can explain the experimental results of the two-player gift exchange game.

As we will show, theories of social preferences mostly predict that worker *i*'s and worker *j*'s efforts should be negatively related, that is, they are *strategic substitutes*. By contrast, our data – and in particular those from a specially-designed experiment – show that efforts are *strategic complements*. Thus, in our sort of environment the behavioural logic behind social interaction effects is more likely *conformism*, not some form of reciprocity or distributional concerns.

2. Design & Procedures

We aim to measure social interaction effects among experimental workers. To identify such effects we use a three-person gift-exchange game where one employer is paired with two workers. In this setup a social interaction effect occurs if a worker's effort decision depends on the effort decision of a co-worker. In order to avoid confounds with strategic incentives we apply a *one-shot* design. Observing social interaction effects among workers in a one-shot situation could be done by letting the workers choose their effort sequentially. We could then test whether the effort decision of the second mover depends on the effort decision of the first moving worker. However, it is difficult to disentangle social interaction effects from the worker's disposition to reciprocate towards the principal.

In order to avoid these confounds we design a variant of the three-person gift-exchange game where the workers first choose their efforts simultaneously and then, after having learned the effort decision of their co-worker, are given the opportunity to *revise* their effort decision, holding their co-worker's effort constant. In this section we describe the game and provide the details about the experimental sessions.

A. The Three-Person Gift-Exchange Game with a Revision Stage

The three-person gift exchange game we used in this study is a sequential game. In a first step the employer chooses the wages $w \in \{50, 100, 200\}$ for the two workers. The employer has to choose the same wage for both workers and the workers know this. Both workers then learn their wage. The two workers then decide simultaneously about their effort, that is, they choose $e_i \in \{1, 2, ..., 20\}$. In some of the sessions we elicit the workers' beliefs about their co-worker's effort choice e'_j (we will provide our rationale for eliciting beliefs in section 5B).[3]

In a next step, the workers enter the *revision stage* where they learn e_j, the effort decision of their co-worker. In the light of this new information they can revise their effort. Both workers choose a revised effort $\hat{e}_i \in \{1, 2, ..., 20\}$. However, the revised effort is only effective for one of the two workers. A random device generates $r =$ ‘ with equal probability. In case of $r = 1$ worker 1 can actually revise his effort and the revised effort of worker 2 has no effect on any of the payoffs. In case of $r = 0$, worker 2 is the reviser and worker 1's effort remains unchanged. Subjects know this procedure when choosing the revised effort. The expected payoff of the principal is

$$\pi_P\left(w, e_1, e_2\right) = v\left[r(\hat{e}_1 + e_2) + (1-r)(e_1 + \hat{e}_2)\right] - 2w, \quad (1)$$

where v is the constant marginal product of the workers' efforts. The earnings of the two workers are calculated as

$$\pi_1\left(w, e_1\right) = w - rc(\hat{e}_1) - (1-r)c(e_1) \qquad \text{and} \qquad \pi_2\left(w, e_2\right) = w - (1-r)c(\hat{e}_2) - rc(e_2), \quad (2)$$

[3] The instructions for the three-person gift exchange game can be found in Appendix A.

where the cost of effort is equal to $c(e_i) = 7(e_i - 1)$ for both workers. Note that we do not allow the principal to differentiate the wages between the two workers. The reason for doing so is that we want to observe the two workers in the simplest possible situation. Allowing for different wages would have introduced an additional source of distortion. For the same reason the two workers have the identical marginal productivity (v). From a principal's point of view the two workers' efforts are perfect substitutes, as the two workers are technologically independent and equally productive.

The most important design feature is the *revision stage*. This stage allows measuring the influence of the co-worker's effort decision. We interpret e_i as the isolated, interaction-free *initial* effort decision. We use the revised effort to identify social interaction effects. The only change between the initial effort decision and the revise stage is the additional information about the co-worker's effort. We therefore interpret $\Delta e_i = \hat{e}_i - e_i$ as reaction to the effort information, i.e., as an indication for a social interaction effect.

It is important to note that we measure the social interaction effect in a situation where the other agent's effort remains unchanged. It is this feature that avoids the central identification problem inherent in field data: the interdependence of the efforts (Manski (1993), Manski (2000)). Revised efforts in our experiment are not mutually dependent. The random device ensures this and allows us to collect revision decisions from all workers. When choosing the revised effort $\hat{e}_i$, worker i knows that either the decision has no effect ($r = 0$) or the effort of the other worker (e_j) remains unchanged ($r = 1$).

A caveat is in order, however. We cannot rule out the possibility that subjects might want to change their effort decision in the *revision stage* for reasons unrelated to social interaction effects. For instance, one might be concerned that the mere existence of the *revision stage* induces an "experimenter demand effect" (Orne (1962)). If subjects are asked to decide again about their effort they might feel urged to change their decision. A second reason might be "virtual learning" (Weber (2003)): the revision stage provides subjects with an additional opportunity to think through the problem. Third, effort revisions might simply occur due to change of mind or errors. Thus, in order to isolate social interaction effects from other sources of effort revisions we need a control treatment. Our control treatment is identical to the game explained above but for one small detail. When reaching the *revision stage* subjects are *not* informed about the effort choice of the co-worker. However, they still have to enter a revised effort.[4]

When workers decided on their initial effort (e_i) they did not yet know about the possibility to revise their effort. The workers received the additional instructions about the revision decision only after everyone had decided on his or her initial effort and the game therefore had proceeded to the revision stage. This precludes an anticipation effect in the initial effort decisions. The information about the revision possibility and its description appeared on a separate screen (for the exact wordings see Appendix A).

[4] We expected the participants to perceive this treatment as rather peculiar. However, to our surprise not a single participant had a question when asked to revise the effort lacking any new information.

In order to check for the robustness of our results we varied several contextual parameters across sessions. First, we varied the level of the workers' productivity ($v = 18$ or $v = 35$ for both workers) and therefore the gains from cooperation. Second, in some sessions we elicited beliefs about the co-worker's initial effort choice, to be able to test theoretical predictions (section 5). Finally, in some of the sessions subjects played six periods of a repeated three-person gift-exchange game (without being informed about the effort choice of their co-workers) before they played this experiment.[5] This contextual variation allows us to check whether increased experience with the game influences social interaction effects.

B. Further Design Features and Procedural Details

The experiment was conducted in computerized laboratories where subjects were separated by partitions and thus could take their decisions in isolation and without possibilities for communication. Moreover, all decisions were anonymous.

We used the software z-Tree (Fischbacher (2007)) to conduct our experiments. To avoid contextual framing effects, we used a 'buyer-seller' terminology.

An important feature of our experiment is that it is a one-shot experiment. We therefore took great care to ensure that subjects understand the rules for effort choices and revisions, as well as the payoff consequences of their decisions. Therefore, subjects had to answer a set of control questions that ensured their understanding of payoff consequences. To help them in calculating payoffs, the software provided a 'What-if calculator', where they could calculate the payoff consequences for all players and all possible combinations of efforts and wages. The 'What-if calculator' was available at all stages of the experiment. Figure 1 illustrates the decision screen subjects saw at the revision stage after having been informed about the revision stage.

[5] We discuss the results of these experiments in Gächter and Thöni (2008).

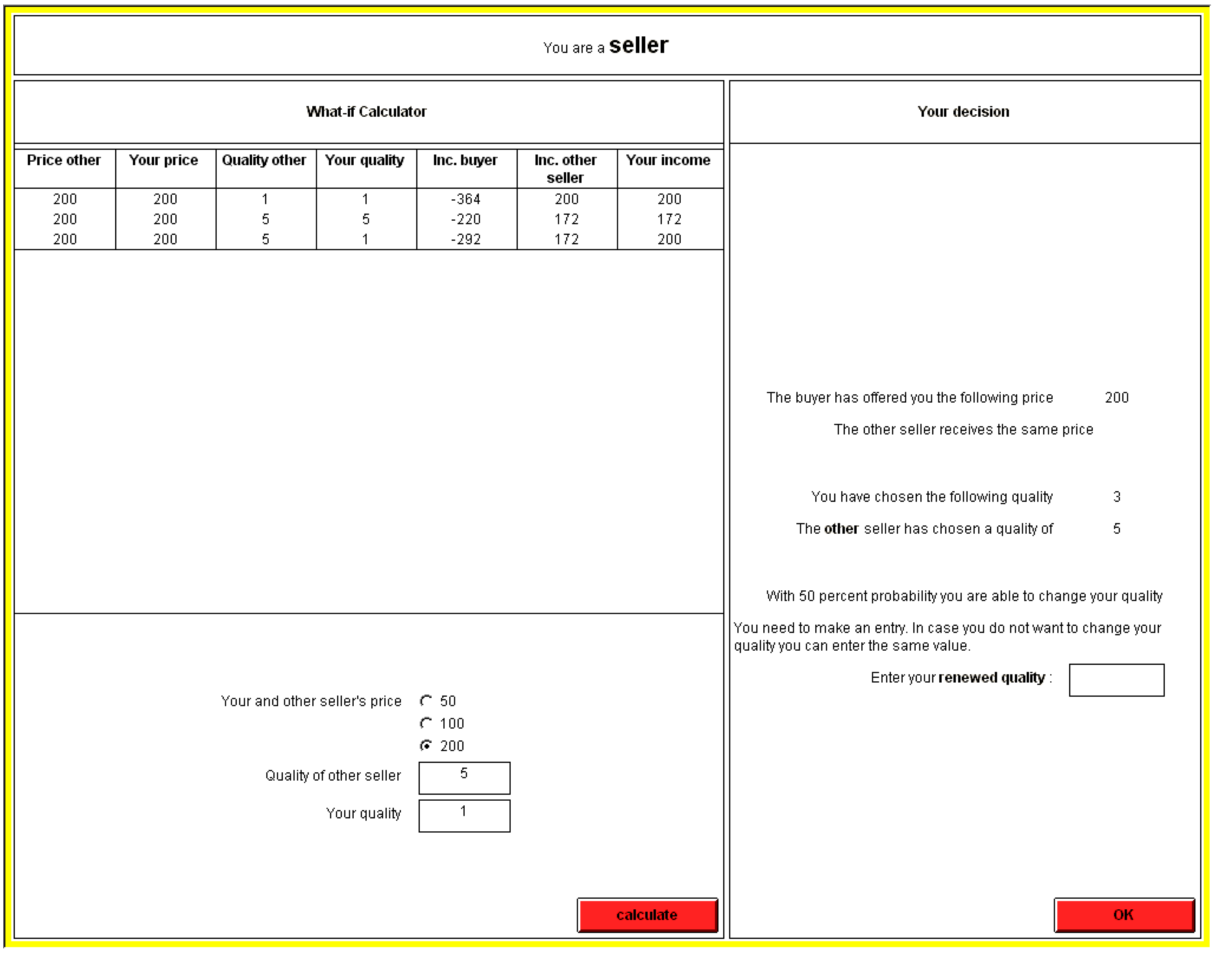

Figure 1: Screen shot of the decision screen at the revision stage.

We have observations from 17 sessions with a total of 465 participants, 310 workers and 155 principals. The majority of these participants (306) played the main treatment, where the workers are informed about the effort choice of their co-worker. The remaining 159 subjects played the control treatment. Table B1 in Appendix B provides an overview of the number of observations by treatment and contextual variation. Five sessions took place in the laboratory of the University of St. Gallen (114 observations). The remaining 12 sessions were conducted at the University of Zurich (351 observations). We imposed no explicit time limit for the decisions. The experiment lasted about 30 minutes and the average earnings were CHF 13.8 (€ 8.8).

3. Hypotheses

Why and how should an agent's effort choice depend on the effort of the other agent? In the following we consider a series of theoretical arguments and their predictions with regard to social interaction effects. We limit our analysis to the subgame starting when the two workers choose their effort. We derive worker i's reaction function to worker j's effort decision, that is, $e_i = R(e_j)$ and focus on the derivative with respect to e_j. Under standard assumptions agents have a dominant strategy to provide minimal effort. There is no reason to reconsider this choice in the *revision stage*, thus the prediction is the absence of social interaction effects, i.e., $de_i/de_j = 0$. However, a large number of laboratory gift-exchange experiments showed that standard assumptions cannot predict the behavior of a majority of the subjects. Alternatively we consider models of (i) learning, (ii) social preferences, and (iii) conformity to predict social interaction effects.

A. Effort Revisions due to Learning

In the revise stage of the main treatment subjects get the opportunity to observe the decision of another subject who is in the exact same position as they are themselves. If subjects are in some way boundedly rational in the sense that they do not fully capture the incentive structure of the game then the other worker's effort choice might be informative. The revise stage provides an opportunity to imitate the behavior of the other worker or to reconsider the own decision in the light of the lesson learnt by observing the other worker's effort choice (Schlag (1998)). An essential ingredient of learning and imitation models is that only superior behavior is imitated. Learning and imitation thus predicts asymmetric effort revisions. In the revise stage the agent with the higher effort observes the other agent playing a superior strategy and adapts. The agent with the lower effort has no reason to change her strategy.

B. Effort Revisions in Theories of Social Preferences (preliminary!)

Next we consider the predictions of theories of social preferences. There are two main reasons for such an analysis. First, theories of social preferences aim at explaining (or predicting) behavior also in novel games like ours, not just existing ones. Second, and related, among many other games, these theories can explain the two-player version of the gift-exchange game. It is thus obvious to explore the explanatory power of these theories in the three-player gift-exchange game. Furthermore, we not only explore the implications of one particular theory (like the frequently used Fehr-Schmidt 1999 model) but compare predictions for all major economic theories of social preferences we are aware of. We consider models of reciprocity (Dufwenberg and Kirchsteiger (2004)), type-based reciprocity (Levine (1998)), distributional social preferences (Bolton and Ockenfels (2000); Fehr and Schmidt (1999)), and hybrid models that combine interpersonal payoff comparisons and intentionality (Charness and Rabin (2002); Falk and Fischbacher (2006); Cox, Friedman and Gjerstad (2007)). All theories of social preferences considered contain the money-maximizing player as a special case. Thus, $de_i/de_j = 0$ is compatible with all theories, as long as minimal effort is chosen. Table 4 summarizes the results and Figure 4 illustrates them (for $w = 200$, $v = 35$). In the following we briefly discuss the basic results and the underlying intuition. For details see the Appendix.

Reciprocity as formalized in the model of Dufwenberg and Kirchsteiger (2004) does not predict any dependency among the two efforts. The reason for this is simple: Worker *j*'s effort has no influence on worker *i*'s profit. Thus worker *j* is neither kind nor unkind to worker *i*. The only reason for choosing a non-minimal effort is to reward the principal for a high wage, irrespective of the other worker's actions. In case of *type-based reciprocity* formulated by Levine (1998) the results are similar. In this model players gain (dis)utility from other workers' income if they are altruistic (spiteful) types. However, since the workers cannot influence their co-worker's income they cannot act altruistically (or spitefully) towards them and thus, do not take their actions into account. Hence, both models predict $de_i/de_j = 0$.

Reaction functions become more interesting if we consider *distributional models of social preferences*. Consider for example the model of inequity aversion proposed by Bolton and Ockenfels (2000). Players are assumed to care for their *share of the total income*. In our game with groups of three the preferred income share is, ceteris paribus, one third. Now consider the case of a very strong inequity averse player. In the role of worker *i*, such a player would choose her effort in a way that her share of the total income equals one third:

$$\frac{\pi_i(w,e_i)}{\pi_i(w,e_i)+\pi_j(w,e_j)+\pi_P(w,e_i,e_j)} = \frac{1}{3}. \qquad (3)$$

It is easy to see that for such a player the two efforts are strategic substitutes. Consider an *increase* of player *j*'s effort. This decreases π_j and increases π_F. Since providing more effort is efficient the sum of these two increases by $v - 7 > 0$ and the left-hand expression in (3) drops below one third. To reestablish equality worker *i* must *decrease* her effort in order to increase π_i. The reaction function of such a worker is depicted in the left panel of Figure 4. Using the payoff functions (1) and (2) and solving (3) for e_i one can show that the slope of the reaction function is $(7 - v)/(14 + v) < 0$. Players with weaker inequity aversion face a trade off between the benefit of their own payoff and the discomfort of earning a relative income above one third. Lower concerns for inequity aversion lead to lower efforts ceteris paribus (see also left panel of Figure 4). As in all models there is a lower limit of inequity aversion under which behavior is identical to money maximizing players (thin line). However, irrespective of the exact specification of the utility function it can be shown that, for interior solutions ($1 < e_i < 20$), the slope of the reaction function is always in $(-1, 0)$, i.e., the two efforts are always strategic substitutes. The intuition behind this result is straightforward. An inequity-averse worker providing low effort suffers from earning more than her equal share. To relieve this adverse feeling there are two possibilities: (i) she increases her own effort and thereby lowers her income, or (ii) the co-worker increases his effort and thereby increases the total payoff, which in turn brings the (unchanged) income of the worker at hand closer to the equal share.

The model of Fehr and Schmidt (1999) is also built on the notion of inequity aversion. However, unlike in the model by Bolton and Ockenfels (2000) subjects make *bilateral comparisons* with all group members. Subjects get utility from their own monetary payoff and disutility from any payoff difference with the comparison partner. Specifically, the disutility of earning less than another group member is equal to α times the payoff difference. Earning more than another group member leads to a similar disutility, weighed by β. If a worker's disutility from earning more than others is sufficiently small ($\beta < 14/(v+14)$) then she will provide minimal effort and earn a higher income than the principal and a (weakly) higher income than the other worker. On the other hand, if the worker's discomfort from earning more than others is sufficiently strong, then she is ready to increase effort. Due to the linearity of the model this holds until some other player's income is matched. The middle panel in Figure 4 shows where this is the case. In region A worker *i* earns the highest payoff. The 45-degree line is the locus where worker *i* and worker *j* earn the same income and the negatively sloped graph shows the locus where worker *i*'s income is identical to the principal's income.[6] Solving $\pi_F = \pi_i$ for e_i gives the slope of this graph as $-v/(v+7)$, which is in $(-1,\ 0)$. Consequently, the intersection of the two is the effort combination where all three players earn the same income, which is the case at $\bar{e} = (3w+7)/(2v+7)$. Region B shows the are where worker *i* earns more than the principal but less than the co-worker. Consider the case where worker *j* chose an effort below $\bar{e}$. If worker *i*'s β is high enough to induce non-minimal efforts she will do so until her effort matches the other worker's effort and hence they both earn the same payoff. Still, inequality towards the principal remains and might induce worker *i* to choose an even higher effort, trading inequality towards the principal with inequality towards the other worker. Here we have to distinguish two types. The first type we call 'Weak egalitarian' (*WE*). A *WE* worker chooses the effort such that her income matches the higher of the other two incomes. For low co-worker's efforts ($e_j \leq \bar{e}$) the worker matches the other worker's effort, otherwise effort is chosen such that the principals income is matched (as long as possible). The second type we call 'Strong egalitarian' (*SE*). Such a worker always chooses an effort which leads to identical incomes of the principal and herself.[7] Thus, for high co-worker's efforts ($e_j > \bar{e}$) the two types are identical. The difference lies in the region for low co-worker's efforts, where the *SE* prefers to put herself in a situation where she earns less than the co-worker in order to prevent earning more than the principal. Thus, an SE worker's concern for inequality towards other players who earn less (β) must be relatively strong compared to the concern for inequality towards player who earn more (α), or, more precisely, $\beta > (14+7\alpha)/(v+7)$ must hold. For *SE* workers the two efforts are strategic substitutes. For *WE* workers the two efforts are strategic complements as long as $e_j \leq \bar{e}$ and strategic substitutes after that. Note that, if the two efforts are strategic complements, then they must be one-to-one complements, i.e., the co-worker's effort is matched exactly.

[6] Solving $\pi_F = \pi_i$ for e_i gives the slope of this graph as $-v/(v+7)$, which is in $(-1,\ 0)$.

[7] Note that this does not mean that a *SE* worker is especially loyal towards the employer. It is rather the fact that a change in effort of one reduces or increases the inequality towards the other worker by 7 and towards the principal by v+7. In this sense, reducing inequality is more effective when adjusting to the principal's income.

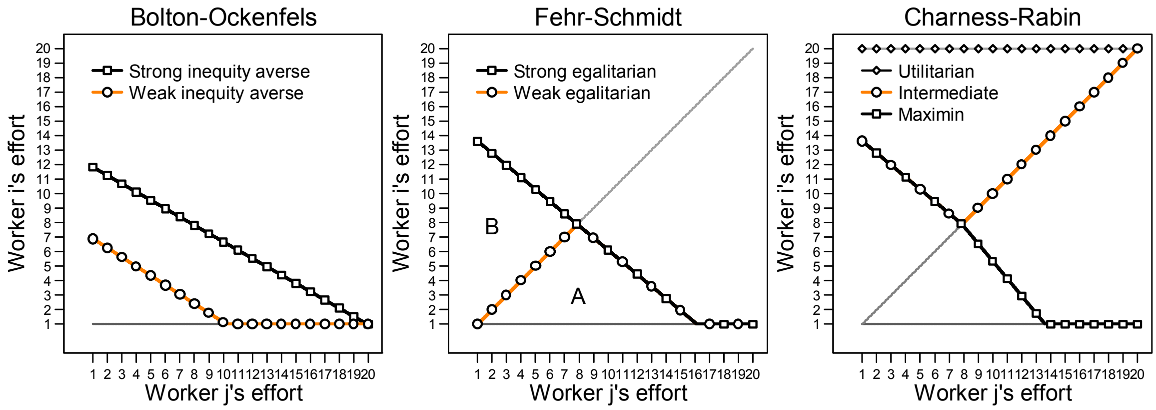

Figure 4: Reaction functions predicted by theories of social preferences. The reaction functions are drawn for $w=200$ and $v=35$.

In contrast to the distributional models previously discussed, Charness and Rabin (2002) propose preferences for efficiency (utilitarian) and/or care for the least fortunate (maximin). A sufficiently concerned maximin type will maximize the payoff of the player least well off (which might be herself). The resulting reaction function is shown in the right panel of Figure 4. For low co-worker's efforts ($e_j \leq \bar{e}$) this worker chooses her effort such that her payoff is equal to the principal's payoff, acting like a *SE* worker in the Fehr-Schmidt model. Otherwise ($e_j > \bar{e}$) worker *i* chooses an effort such that the principal earns as much as the co-worker. Thus, the reaction function has a kink at $e_j = \bar{e}$ and the right part of the graph is mirrored at the 45 degree line. A worker who is sufficiently utilitarian will provide full effort irrespective of what the other worker does. In addition to that some parameter constellations give rise to an intermediate type, whose reaction function is u-shaped and follows the reaction function of a maximin type in $e_j \leq \bar{e}$. For higher co-worker's efforts this worker chooses her effort such that her income is equal to the co-worker's income, so the reaction function is equal to the 45 degree line. The reason for this lies in the mix between maximin and utilitarian motives. The worker chooses the highest possible effort that does not put herself in the strict minimal-payoff position.

Finally, there is a series of hybrid models that combine distributional concerns and intentional motivation. Charness and Rabin (2002) introduce a parameter for 'demerits', which reduces a player's concern for another player's well-being if the other player does not deserve it. If worker *i* thinks that the principal does not deserve to be treated well then negatively sloped parts of the reaction function of both maximin and intermediate players move downwards, leaving the slopes unchanged and leaving the kink on the 45 degree line. Thus, the effort level of the kink depends negatively on the degree of demerit towards the principal (d_P), i.e., the kink is at $e_j = e(d_P) \leq \bar{e}$, with $e' < 0$. As a special case the kink can be at the minimal effort and then the intermediate type has a reaction function that is identical to the entire 45 degree line in the right panel of Figure 4.

The model of Falk and Fischbacher (2006) combines interpersonal payoff comparisons with reciprocity. Like in the case of Dufwenberg and Kirchsteiger (2004), reciprocity does not predict a direct link between the two efforts. However, since worker *i* wants to reciprocate to the principal and cares for payoff differences the predictions of the Falk-Fischbacher model are very similar to the predictions of the *SE* type in the Fehr-Schmidt model. For very strong social preferences the reaction function is identical to the *SE* type, weaker social preferences result in a parallel downwards shift.

The model by Cox, Friedman and Gjerstad (2007) allows for different marginal rates of substitution between a player's own payoff and other players' payoffs. In addition, a parameter measures the emotional state towards the other players. If worker *i*'s has negative emotions towards the employer then she will always choose minimal effort. In case of a positive emotional state her effort choice depends on whether the two payoffs are complements or substitute in her utility function. In case of complements the reaction function is identical to the Fehr-Schmidt *SE* worker. If the two payoffs are substitutes, then the reaction function is shifted upwards or downwards, depending on the strength of the positive emotion towards the principal.

Table 4 provides an overview of the models and their predictions with regard to the slope of the reaction function. With the exception Fehr-Schmidt *WE* type and the Charness-Rabin intermediate type all models incorporating various notions of social preferences predict either that there is no dependency between the two efforts or that efforts are strategic substitutes.[8] The two exceptions allow for strategic complementarity among the two efforts. However, the complementarity has to be one-to-one, i.e., the two workers choose identical efforts.

	Model	Slope of $R(e_j)$ for interior solutions ($1 < e_i < 20$)
	Money maximizing	$\frac{de_i}{de_j} = 0$ (no interior solutions)
(Type based) reciprocity	Dufwenberg and Kirchsteiger (2004)	$\frac{de_i}{de_j} = 0$ (no interior solutions)
	Levine (1998)	$\frac{de_i}{de_j} = 0$ (no interior solutions)
Distributional	Bolton and Ockenfels (2000)	$-1 < \frac{de_i}{de_j} < 0$
	Fehr and Schmidt (1999)	
	Strong Egalitarian (*SE*)	$-1 < \frac{de_i}{de_j} = -\frac{v}{v+7} < 0$
	Weak Egalitarian (*WE*)	$\frac{de_i}{de_j} = \begin{cases} 1 & \text{for } e_j \le \bar{e} \\ -v/(v+7) & \text{else} \end{cases}$
r b ,	Charness and Rabin (2002)	

[8] The fact that in all but one cases the predicted slope of the reaction functions lies in (–1, 0) ensures that there is a unique equilibrium in the subgame where the two workers choose their effort. The Charness-Rabin maximin type is the only exception among the negatively sloped reaction functions. If two workers of this type are paired then the two reaction functions overlap entirely.

Maximin	$\frac{de_i}{de_j} = \begin{cases} -v/(v+7) & \text{for } e_j \le e(d_P) \\ -(v+7)/v & \text{else} \end{cases}$
Intermediate	$\frac{de_i}{de_j} = \begin{cases} -v/(v+7) & \text{for } e_j \le e(d_P) \\ 1 & \text{else} \end{cases}$
Utilitarian	$\frac{de_i}{de_j} = 0$ (no interior solutions)
Falk and Fischbacher (2006)	$-1 < \frac{de_i}{de_j} = -\frac{v}{v+7} < 0$
Cox, Friedman and Gjerstad (2007)	$-1 < \frac{de_i}{de_j} < 0$

Table 4: Predictions for the slope of the worker *i*'s reaction function to worker *j*'s effort, $e_i = R(e_j)$. Details are provided in the Appendix.

C. Effort Revisions due to Conformity

We have seen that the overwhelming majority of models of social preferences discussed in the previous section predict a negative connection between the two efforts. This might come as a surprise because it seems to contradict the empirical findings reported in several recent papers on social interaction effects (Ichino and Maggi (2000), Falk and Ichino (2006)), which suggest strategic complementarity between the two efforts. What could explain strategic complementarity?

One of the models able to account for strategic complementarity is the concept of social proof (Cialdini (2001)). Experimental evidence from social psychology (cite? Asch …) suggests that observing other subjects' actions influences individuals' behavior. There seems to be an urge to conform to other peoples' actions. Bernheim (1994) proposes a theory of conformity, where agents are assumed to gain utility from conforming to a norm in the form of esteem. In a similar way, Ellingsen and Johannesson (Forth) propose a model where they assume that subjects care about their social esteem and have, depending on the esteem of other players, prefer to act pro-socially towards them. Finally Benabou and Tirole (2006) make an even finer distinction of the motivational structure and separate concerns for social reputation (external approval) and for self respect (internal approval). The latter distinction seems to be important when we apply the models to laboratory settings. In our experiment subjects interact anonymously and therefore there is no way of communicating discontent to other subjects. Ellingsen and Johannesson argue that even in anonymous laboratory settings subjects care about how other subjects think of them.

If subjects in our experiment care in some way about what others think of them they might have an incentive to adjust their effort to the other agent's effort. If the other chose a lower effort than they did, they might reconsider their degree of kindness towards the principal and reduce their effort in the revise stage. If they observe that the other agent had chosen a higher effort, then they expect the other players to disapprove their effort choice and increase their effort in the revise stage. A desire to conform to a norm is compatible with strategic complementarity among the efforts.

Are the reactions to higher other agents' efforts and lower agents' efforts symmetric? There is an empirically motivated argument that which suggests an asymmetry. Experimental evindence from public goods games (FGF) shows that many subjects are conditionally cooperative with a self serving bias. This means that they are ready to contribute if others do so but prefer to contribute less than the others. In our game, self interest always demands for lower efforts. Thus, if agent *j* chose a lower effort then the agent *i* has two reasons to lower the effort in the revise stage, whereas in the other case, agent *i* must trade off her preference to conform against her self interest. It is thus likely that we will observe more negative effort revisions than positive effort revisions.

D. Hypotheses

What are the hypotheses we can extract from the theoretical considerations? We first address the question whether the information about the other agent's effort should have an effect at all on effort revisions. Money maximizing and pure reciprocity models predict that the effort information in the revise stage is irrelevant. The null hypothesis is that effort revisions in the effort information treatment (*EIT*) are equal to effort revisions in the control treatment (*CT*), i.e., $\Delta e_i^{EIT} = \Delta e_i^{CT}$. On the other hand, social preferences and conformity arguments provide reasons for revising the effort in the light of the new information. The alternative hypothesis is therefore that effort revisions are stronger in the main treatment, i.e., $\left|\Delta e_i^{EIT}\right| > \left|\Delta e_i^{CT}\right|$.

Theories of social preferences and theories of conformity do however differ in their predictions with regard to the direction of the effort revision in the *EIT*. Conformity arguments lead to a quite simple prediction. If agent *j* chose a lower effort than agent *i*, then *i* should decrease her effort and, presumably to a lesser extent, vice versa. Thus the hypothesis can be written as $\Delta e_i = f(e_j - e_i)$ with $f' > 0$ and $f(0) = 0$. On the other hand, most models of social preferences do predict the efforts to be strategic substitutes. It is not immediately clear what this means for the slope of $f(\bullet)$, because the models predict negatively sloped reaction functions $e_i = R(e_j)$, which are independent of whether the effort of *i* is higher or lower than the effort of *j*. In order to test for the slope of the reaction function we will use data about agents' beliefs.

4. Results

A. Wage Setting and the Initial Effort Choice

The first step of our analysis is to look at the workers' reaction to wages. Recall that both treatments are identical up to the *revision stage*. Therefore, the initial effort decision cannot be influenced by our treatment variation. For the following analysis we therefore pool the data of the main and control treatment. The average effort chosen at the lowest wage was 1.54; the intermediate wage triggered an average effort of 3.01 and the highest wage an average effort of 5.52. Minimal efforts occurred in 67.8 percent, 48.9 percent and 37.1 percent of the cases in which principals paid the low, intermediate and high wage, respectively.[9]

Thus, as expected from previous gift-exchange experiments (Fehr and Gächter 2000), efforts clearly increased in wages. In Table 1 we check whether the increase of the effort in the wage is significant and whether the initial effort decision is systematically influenced by the contextual variations. We use a Tobit estimation with the initial effort decision as dependent variable. The independent variables are dummies for the high and low wage level and dummies that identify the contextual variations. The baseline case is the intermediate wage and the observations stemming from the low productivity experiments in St. Gallen.

[9] Among the 155 principals in our sample 47.1 percent paid the lowest possible wage of 50. Another 30.3 percent paid the intermediate wage of 100 and the remaining 22.6 percent offered the highest wage of 200.

	Dependent variable: Initial effort	
	Coefficient	SE
Low wage	-2.941**	0.733
High wage	3.156**	0.899
Experienced	0.137	0.763
Belief	-2.307**	0.824
High productivity	2.078	1.503
Zurich	0.386	1.188
Constant	0.071	1.120
Sigma	4.630	
N	310	
ll	-510.462	
Prob > chi2	0.000	

Table 1: Tobit estimation for the initial effort decision. All independent variables are dummies. The first two variables measure the different wages; the intermediate wage is the reference wage. The remaining variables are dummies for the contextual variations. We estimate robust standard errors with clustering within matching group for the data in the *Experienced* treatments.

The coefficients for the two wage dummies are highly significant, reflecting the fact that workers, on average, react positively to wage increases. Surprisingly, after controlling for the wage, the contextual variation *Experienced* has no influence on efforts. The same is true for the increase in the productivity parameter (*High productivity*) and the change in location (*Zurich*). Eliciting beliefs has a highly significantly negative impact on the initial effort decision. Asking the workers about their beliefs with regard to their co-worker's effort reduces the average effort from 3.43 (without the belief question) to 2.25.[10]

B. Existence of Social Interaction Effects

We now turn to the results from the revision stage where we test our two hypotheses. In the main treatment we observe 204 revision decisions and in the control treatment (without effort information) we observe 106 revision decisions. Panel A of Figure 2 shows the main treatment effect. In 69 (33.8 percent) of the cases in the main treatment workers do actually revise their effort, in the remaining cases they leave their effort unchanged. Effort revisions also occur rather frequently in the control treatment; 24 out of the 106 subjects (22.6 percent) revise their effort in a situation where no information about the co-worker's effort is provided.

In some cases social interaction effects could only result in upward revisions because the initial effort level was minimal. Moreover, people with no or weak other-regarding preferences might also be less influenced by social interaction effects, compared to people who showed a willingness to deliver non-minimal effort levels initially. In order to investigate effort revisions of these people we study a reduced sample we only look at cases where the workers chose a non-minimal initial effort ($n = 120$). Panel B of Figure 2 shows the results

[10] This result is consistent with Croson 2000 who found that eliciting beliefs in the public good game reduces voluntary contributions, relative to a treatment without belief elicitation.

for the restricted sample. In the main treatment 50 out of 71 (70 percent) of the workers revised their effort while 19 out of 41 (46 percent) did so in the control treatment.

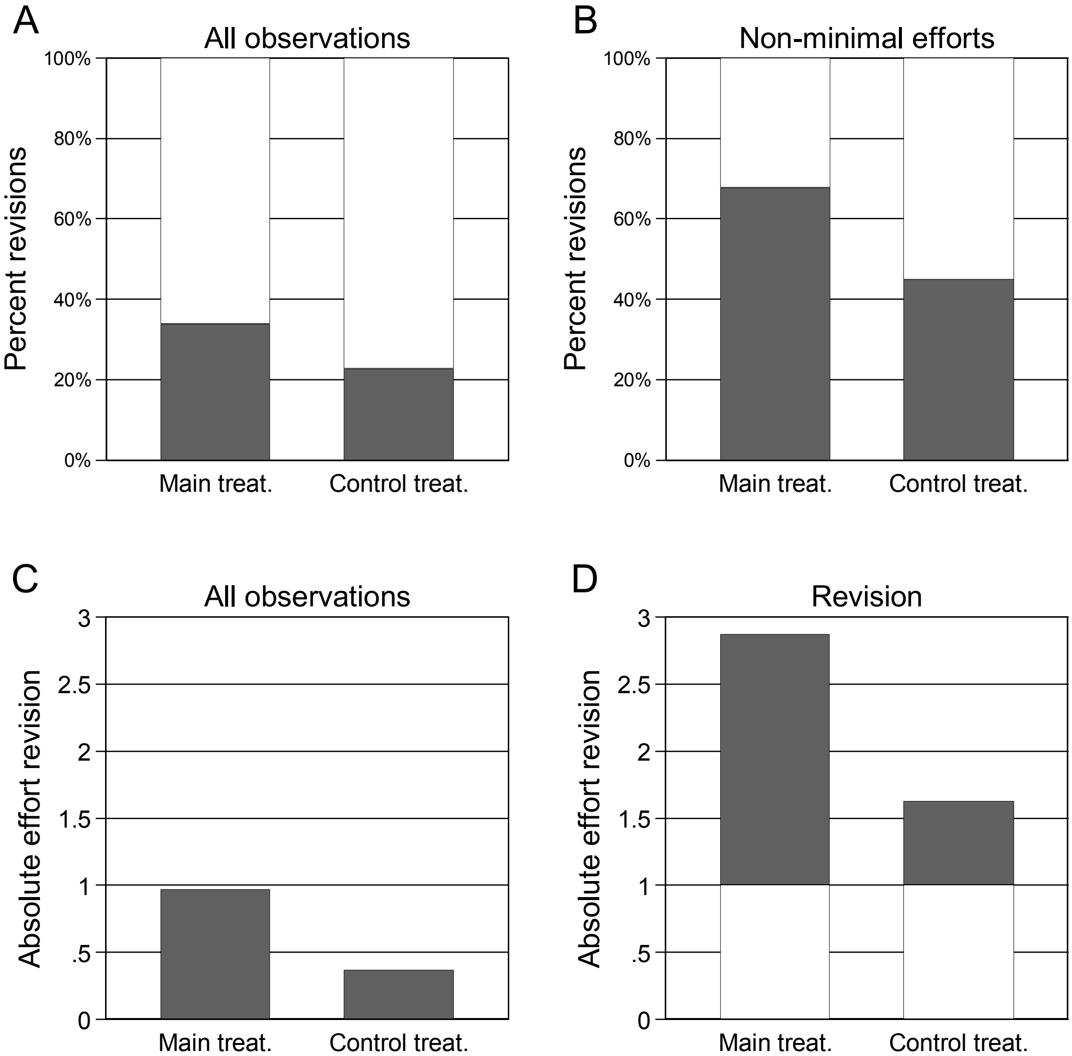

Figure 2: Treatment effects. Frequency of effort revisions in the whole data set (Panel A) and in the restricted data set of the workers with non-minimal efforts (Panel B), and the absolute magnitude of the effort revision for all observations (Panel C), and for the subset of workers who do change their effort in the revise stage (Panel D).

When drawing statistical inference we have to consider that observations within a principal are not independent. Both workers receive the same wage and, in the main treatment, information about the co-worker's effort choice. We control for this by clustering the data on the principal level.[11] Table 2 reports the results of Probit estimations. The dependent variable is *Revision*, a dummy for the decision to revise the effort, which equals one if $\Delta e_i \neq 0$ and zero otherwise.

Model 1 shows that the treatment variable *Effort transparent* increases the probability of an effort revision significantly. In Table 2 we also report marginal effects. Thus, the

[11] As before the observations from the *Experienced* sessions are clustered within matching group.

probability of an effort revision is 15 percent higher in the main treatment than in the control treatment. None of the other design parameters has a significant effect on the probability to revise the effort. We introduce the initial effort decision by two variables in order to allow for changes in the behavior of workers with minimal effort and non-minimal effort. *Initial effort* is simply the effort chosen (e_i) and *Minimal initial effort* is a dummy for $e_i = 1$. Both variables are highly significant, indicating that higher initial efforts increase the probability of an effort revision.[12]

In Model 2 we repeat the estimation for the restricted sample of workers with non-minimal initial efforts. The marginal effect of the treatment variation increases to 24 percent. Among the other explanatory variables the belief question now has a significant positive and substantial effect on the probability to revise effort. Asking the subjects about their belief increases the probability of an effort revision by 23 percent in this subsample. A closer look at the influence of the belief question reveals that this positive effect stems exclusively from the main treatment, where the workers learn whether their belief was correct or not. If we estimate Model 2 for the main and control treatment separately we obtain a highly significant marginal effect of 37 percent in the former case and an insignificant marginal effect of –13 percent in the latter case.

	Dependent variable: $\Delta e_i \neq 0$ (Dummy)					
	Model 1 (all observations)			Model 2 (only non-minimal initial efforts)		
	Coef	SE	ME	Coef	SE	ME
Effort information treatment (D)	0.556**	0.199	0.153	0.655**	0.238	0.251
Initial effort	0.138**	0.039	0.041	0.142**	0.039	0.054
Minimal initial effort (D)	-1.367**	0.265	-0.411			
Experienced (D)	-0.395	0.227	-0.108	-0.612	0.340	-0.238
Belief elicited (D)	0.337	0.234	0.101	0.605*	0.305	0.221
High productivity (D)	0.251	0.360	0.069	0.293	0.510	0.114
Zurich (D)	0.045	0.317	0.013	0.020	0.380	0.008
Constant	-1.095**	0.375		-1.246*	0.502	
N	310			139		
ll	-119.595			-81.327		
Prob>χ^2	0.000			0.001		

Table 2: Probit estimations for the decision to revise the effort. Apart from *Initial effort* all independent variables are dummies (D). We report coefficients, standard errors (SE), and marginal effects (ME). We apply a robust estimation of the standard errors clustered within principal, * denotes significance at 5 percent, ** at 1 percent.

Panel C in Figure 2 shows that the absolute magnitude of the effort revisions is considerably larger in the main treatment than in the control treatment. The difference is .75 effort units. Thus, the average absolute effort revision differs by a factor of 2.7 in the main treatment compared to the control treatment. Panel D of Figure 2 shows that this effect is not only driven by the fact that workers revise their effort more frequently when information about their co-worker's effort is provided. In the subsample of workers who actually do

[12] The wage is not used as explanatory variable since it is highly correlated with the initial effort. Wage dummies added to the Models in Table 2 are insignificant.

revise their effort ($\Delta e_i \neq 0$) the difference between the average absolute effort revision increases to 1.2 effort units. In both cases the treatment variable is significant.[13] We summarize these findings as follows:

Result 1: *Our evidence supports the Social Interaction Existence Hypothesis. Information about the other worker's effort produces significantly more and larger effort revisions compared to the Control treatment.*

In a next step we test our second hypothesis, the Social Interaction Direction Hypothesis.

C. Direction of Social Interaction Effect

Our second hypothesis posits that effort revisions are a function of the observed difference in the initial efforts of the two workers $\Delta e_i = f(e_j - e_i)$, with $f' > 0$ and $f(0) = 0$. Figure 3 provides a scatter plot of the connection between the observed effort and the effort revision in the main treatment. The size of the dots is proportional to the number of underlying observations. Observations on the thin horizontal line stem from workers who left their effort unchanged. The second thin line is the 45° line. Observations on this line mean that a worker exactly matched the other worker's effort. The numbers in the scatter plot indicate the numbers of observation within a region. Numbers at the end of a thin line count the observations on the line for negative or positive effort differences, respectively. Numbers in areas between lines count the observations within the regions between the thin lines. The number in the middle of the graph indicates the number of observations with zero effort difference and no effort revision. A vast majority of these observations (89 percent) stem from workers choosing minimal initial effort.

Most observations lie between the horizontal and the 45-degree line (89 percent, observations on the lines included). There is very little 'overshooting' in the sense that effort revisions surpass the observed effort differential. In case of negative effort differentials (when the other worker chose a lower effort) many observations lie between the two thin lines. These workers revise their effort towards the other worker's effort but do not match it. In case of positive effort differentials there are only very few observations in this range. Workers either match the other worker's effort, or, in most cases, do not revise their effort at all. This suggests asymmetric reactions to positive and negative effort differentials. When fitting the data with a regression line (bold line) we therefore allow for different slopes.

---

[13] We apply a Tobit regression with the absolute effort revision as dependent variable and the identical controls as used in Table 2. The treatment dummy is significant in both cases ($p = 0.001$ for all observations and $p = 0.042$ for the observations with effort revision).

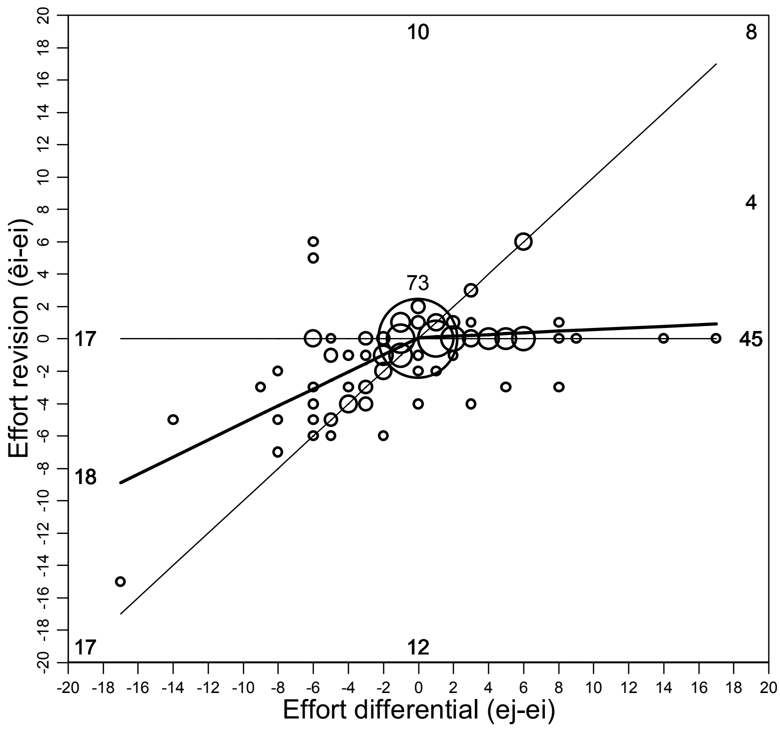

Figure 3: Scatter plot of the effort revision dependent on the difference between the other worker's and the own initial effort. The thin lines show the limit cases of no social interactions (horizontal line) and 'perfect' social interaction (45 degree line). The bold line shows the result of the OLS regression.

We apply OLS estimations to test our second hypothesis. We regress the effort revision on the initial effort differential, the initial effort, and the contextual parameters. Model 1 in Table 3 shows the results of this estimation. The effort differential is positive and highly significant. An increase of the co-worker's effort by one unit induces a worker to increase her effort in the revise stage by .17 units, *ceteris paribus*.[14]

However, as Figure 3 suggests, there are substantial differences between positive and negative effort differentials. In Model 2 we allow for different slopes by adding a second variable for the initial effort differential. The variable *Initial effort differential if positive* is calculated as $\max[e_j - e_i, 0]$. The results of Model 2 confirm the impression gained from Figure 3. The coefficient of *Initial effort differential* is highly significant and positive. Workers who learn that their co-worker had chosen a lower effort reduce their effort on average by .35 effort units per unit of the differential. The interaction variable *Initial effort differential if positive* has a negative coefficient, indicating that the reaction to the effort differential is lower in the positive domain. The reduction is weakly significant. The net effect in the domain of positive effort differentials is the sum of the first and second coefficient. The effect is still positive (.07) but not significantly different from zero ($p = .217$). Thus, the interaction between the two efforts in mainly driven by effort reductions

[14] Surprisingly, this result is almost identical to the magnitude of peer effects found in previous studies in various field contexts. For their respective measures of social interaction effects Ichino and Maggi (2000) find values between .14 and .18; the Falk and Ichino (2006) estimates result in .14, Mas and Moretti (2007) report .17, and Bandiera, Barankay and Rasul (2007) report .13.

of the high-effort workers. This stands in contrast to the results reported by Falk and Ichino (2006) and Mas and Moretti (2007) who find exactly the opposite, namely that mutual observability increases the productivity of the low effort workers.

Among the remaining variables only *Initial effort* has a significant impact on the effort revision. Unlike in the estimates shown in Table 2 the coefficient is negative. Thus, higher initial efforts are more likely to be revised downwards than lower efforts. All dummies for the contextual variations are insignificant. However, this does not necessarily mean that the contextual variations have no influence on the decision to revise the effort since the dummies in Model 1 and 2 control only for level effects. In order to investigate whether the contextual variables influence the reaction to the effort information we add interaction variables. The most interesting contextual variable is the dummy *Experienced*. Here we can test whether subjects who are experienced with the game (because they played a gift-exchange game before this experiment) react differently to the information about the co-worker's effort. In case of negative effort differentials the point estimate for the experienced subjects in .56, which is substantially higher than the coefficient in the whole data set. However, the interaction variable is far from being significant. In case of positive effort differentials the resulting coefficient for experienced subjects is -.01 (the sum of the first four coefficients). The interaction variable is insignificant. Thus, the average reaction to positive effort differentials is virtually zero in the experienced subject pool. The reaction to negative effort differentials seems to be even stronger among experienced subjects.[15]

	Dependent variable Δe_i					
	Model 1		Model 2		Model 3	
	Coef	SE	Coef	SE	Coef	SE
Initial effort differential	0.168**	0.048	0.347**	0.117	0.340**	0.119
Experienced×Effort difference					0.218	0.289
Initial effort differential if > 0			-0.277	0.143	-0.266	0.148
Experienced×Initial effort diff. if >0					-0.303	0.347
Initial effort	-0.305**	0.075	-0.224**	0.045	-0.231**	0.047
Minimal initial effort	-0.506	0.371	-0.488	0.340	-0.549	0.366
Experienced (D)	-0.035	0.399	-0.205	0.383	-0.084	0.359
Belief elicited (D)	0.110	0.468	0.191	0.466	0.196	0.470
High productivity (v=35) (D)	-0.875	0.672	-0.670	0.558	-0.687	0.566
Zurich	0.951	0.606	0.673	0.464	0.680	0.471
Constant	0.697	0.535	0.797	0.494	0.846	0.507
N	204		204		204	
Ll	-387.833		-383.305		-383.045	
P	0.000		0.000		0.000	
r2	0.407		0.433		0.434	

Table 3: OLS regression of the effort revision Δe_i dependent on the difference in the initial efforts (*Initial effort differential* $e_j - e_i$), the effort differential in the positive range, i.e., $\max[e_j - e_i, 0]$. In Model 2 we add interaction variables between *Experienced* and *Initial effort differential.* Robust standard errors in parentheses, two workers in a group are clustered; * means significance at 5%, ** means significance at 1%.

We summarize our findings in Result 2.

[15] We also tested interaction effects with the other contextual variables. All interaction effects remain insignificant.

Result 2: *The difference between the co-worker's and the worker's effort is a significant predictor for effort revisions. Workers who learn that their co-worker has provided less effort than they did, reduce their effort significantly, whereas workers who chose a lower initial effort than their co-worker increase their effort only insignificantly.*

The asymmetry in the reactions to the observed other worker's effort suggests that effort transparency reduces average efforts. This is clearly the case. Average efforts drop by .4 effort units from 2.9 to 2.5 in the revise stage. In the control treatment we observe a considerably smaller drop from 2.8 to 2.6. Thus the effort information tends to strengthen the tendency to decrease the effort in the revise stage. However, the difference between the main and the control treatment are not significant. If we estimate Model 1 in Table 3 for the whole data set then the dummy for the treatment variation has a coefficient of -.20 and a *p*-value of .248.

5. Understanding the Driving Forces Behind Social Interaction Effects

We believe that learning and imitation cannot account for our results, however. The fact that we observe an asymmetric pattern of effort revisions could be seen as supporting evidence. However, a closer look at our data reveals that it is very unlikely that the downward revision of the efforts is driven by erroneously high effort choices in the first place. First, we have shown in Model 3 of Table 3 that subjects who are experienced with the game from previous play do not show weaker reactions to the information about the co-worker's effort (in fact, they seem to react even stronger). Second, recall that the subjects had access to a profit calculator when taking their decision. The calculations they made were recorded. The data shows that among our 465 subjects all but two calculated the payoffs for the Nash equilibrium efforts. Thus, virtually all subjects have actively explored the payoff consequences of minimal effort choices and should not be surprised by the fact a co-worker with a lower effort earns a higher payoff.

C. What Flavour of Social Preferences?

We have derived theoretical predictions about worker's effort choice in the three person gift exchange game and have seen that, with the exception of a few special cases, the slope of the reaction function $e_i = R(e_j)$ lies between –1 and 0 for interior effort decisions. Efforts are therefore strategic substitutes. In Figure 3 we have shown that workers paired with a low-effort co-worker tend to reduce their effort and (to a lesser extent) vice versa. One might be tempted to interpret this as evidence for the efforts to be strategic complements. However, there is one problem with this argument: Consider a worker *i* choosing a low initial effort. Assume this worker learns that the co-worker had chosen an intermediate effort and chooses to increase the own effort to the intermediate level in the revise stage. This observation is compatible with a negatively (positively) sloped reaction function if worker *i* expected worker *j* to choose a high (low) effort at the time she chose her initial effort.

In order to measure the slope of the reaction function we therefore need information about worker *i*'s *belief* about worker *j*'s initial effort. In a subset of our data ($n = 94$) we elicited workers' beliefs about their co-worker's initial effort decision. In case the belief was wrong we observe two points on a subject's reaction function, which provides a direct measure for the *slope* of the reaction function. We denote worker *i*'s belief about *j*'s initial effort decision by e'_j. The two points on the reaction function are (e_i, e'_j) and ($\hat{e}_i, e_j$). If worker *j* chooses a higher effort than worker *i* thought ('positive surprise', $e_j - e'_j > 0$), and worker *i* increases her effort ($\Delta e_i > 0$), then the reaction function has a positive slope and vice versa. We apply an OLS estimate for the slope of the reaction function, that is we estimate $\Delta e_i = \alpha + \beta(e_j - e'_j) + \varepsilon$. According to the theories discussed above, the difference between the belief and the actual effort of the other worker is the *only* reason to revise the effort. Thus, all theories predict $\alpha = 0$. The predicted slope of the reaction function depends on the productivity parameter *v*. However, we do not have to control for this because all our observations with the belief question stem from experiments with $v = 35$. In case of Fehr-Schmidt and Falk-Fischbacher the predicted slope is $\beta = -.833$.

The resulting coefficient is positive and highly significant ($\beta = .202$, $p = .005$). The constant is small and insignificant ($\alpha = -.190$, $p = .325$). Figure 5 shows a scatter plot of the observations and the OLS regression line. For comparison we also show the theoretical reaction function of a Fehr-Schmidt SE and Falk-Fischbacher worker.

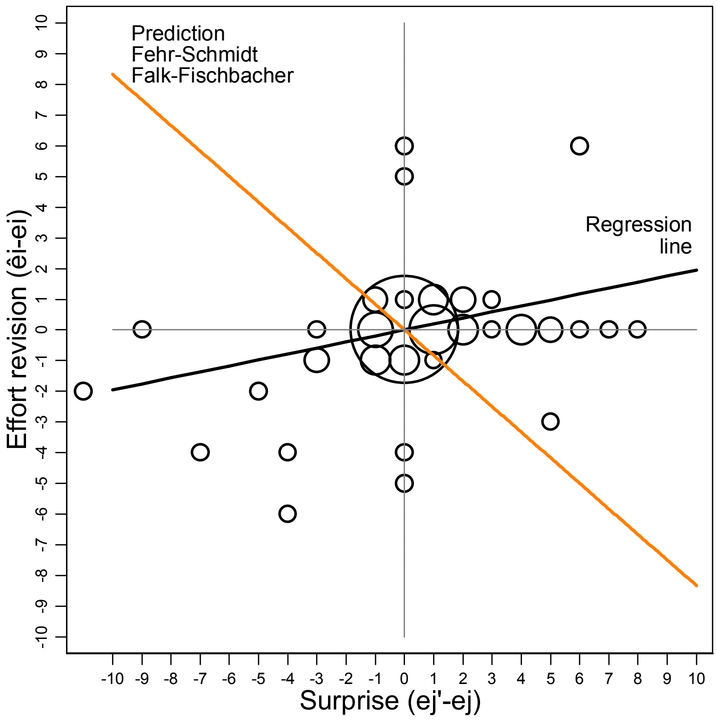

Figure 5: Effort revisions dependent on the difference between the actual e_j and worker *i*'s belief e_j'.

In case of the Bolton-Ockenfels model the slope of the reaction function also depends on the wage. If we add two dummies for the intermediate and the high wage to the regression equation the coefficient for $(e_j - e_j^b)$ remains highly significant ($\beta = .201$, $p = .003$). The coefficients for the wage dummies are insignificant. Thus, a negatively sloped reaction function is clearly refuted by the data. On average, efforts are thus strategic complements rather than strategic substitutes.

A theory that can account for a positively sloped reaction function is the Fehr-Schmidt model. A *WE* Fehr-Schmidt worker chooses to match the other worker's effort up to a certain threshold. According to the parameter estimate provided by Fehr and Schmidt (1999) we should observe only 10 percent *WE* type workers in the experiments with the high productivity parameter. This seems to be too small a number to account for an overall positive slope of the reaction function. Could it be that the *WE* type worker is much more frequent among our subjects?

We check this by a case-by-case evaluation of compatibility with the prediction. An observation is called *WE compatible* if (i) the initial effort is chosen according to the best reply function given the belief about e_j, and (ii) the revised effort is chosen according to the best reply function given the observed e_j.[16]

Among the 204 effort revisions observed in the main treatment 87 (43 percent) are compatible with the *WE* prediction. However, a lot of these observations are workers who choose the minimal effort and are therefore also compatible with the standard prediction. If

[16] Condition (i) can only be checked in the subsample where we elicited beliefs. For the remaining observations we check whether the initial effort is within the range allowed be the reaction function as shown in Figure 1.

we restrict our sample to the workers with non-minimal initial efforts then only 15 out of 90 (17 percent) choose their efforts in accordance to the *WE* type. Another way to assess the predictive power is to look at the fraction of effort revisions ($\Delta e_i \neq 0$) that are explained by the *WE* type behavior. Among the 69 workers who do revise their effort 13 (19 percent) do so according to the prediction. Thus, the majority of effort revisions in our experiment cannot be explained by any of the discussed models of social preferences.

The second theory that predicted positively sloped reaction functions was the utilitarian type in the Charness-Rabin model. In an extreme case, this model is compatible with perfect matching of the other worker's effort. Among the workers who do revise their effort 36 percent match the other worker's effort. The majority of workers revises effort in the direction of the other worker's effort but do not fully match it. Furthermore, although the estimates for the slope of the reaction function shown above are clearly positive, they are nowhere near unity, as the two theoretical cases would predict. An F-test refutes the hypothesis $\beta = 1$ at any conventional significance level.

Result 3: *Efforts are predominately strategic complements rather than substitutes as predicted by most theories of social preferences.*

The explanations offered by current theories of social preferences do not capture the kind of social interaction effects we observe in our experiment. One possibility to account for our result would be to alter the definition of the reference group. In line with the theories we applied we assumed that the reference group comprises all three players that make up a group. This assumption is quite natural, given that in our experiments payoff functions were known by all players. In the context of our game it could be the case that workers feel more attached to the co-worker than to the principal. If we would assume that the reference group of a worker contains only the co-worker, then also models of inequity aversion would predict efforts to be strategic complements.

6. Conclusions

In this paper we were interested to shed light on the behavioural logic behind social interaction effects. We chose a setting of employment relation as social interaction effects might be particularly present in work place interactions (Bewley (1999); Ichino and Maggi (2000); Falk and Ichino (2006); Mas and Moretti (2007); Bandiera, Barankay and Rasul (2007)). To analyse social interaction effects we presented a laboratory experimental design based on a new version of a three-person gift exchange game. Our design eliminates all strategic incentives for social interaction effects as well as the "reflection problem" (Manski (1993)) inherent in the identification of social interaction effects. Thus, we were able to show that a worker's effort is *causally* influenced by his co-worker's effort in the absence of any strategic reason for such an effect. Our laboratory results therefore complement the empirical findings of Ichino and Maggi (2000) and the field experimental findings of Falk and Ichino (2006).

One goal of our analysis was to contrast the data with predictions from theories of social preferences. One prediction of various versions of inequality aversion is that the workers' effort should be strategic *substitutes*, that is, if worker *i* increases his effort to reduce inequality, worker *j* does not have to do so, and vice versa. Our results show that this prediction is not met by the data. The workers in our experiment mostly change their effort such that the gap between the own effort and the other worker's effort becomes smaller – efforts are strategic *complements*, not substitutes. This suggests that the driving force behind social interaction effects in our environment is conformity, rather than inequity aversion or reciprocity.

7. Appendix

A. Summary of the Observations

	Contextual variation				Total
v	18	35	35	35	
Belief	no	no	yes	Yes	
Experience	no	no	no	Yes	
Main treatment	45	120	69	72	306
Control treatment	15	72	36	36	159
Total	60	192	105	108	465

Table A1: Summary of the observations by treatment variation and contextual parameters. Numbers indicate subjects as workers and principals. The number of worker observations is two thirds of the numbers shown.

B. Bolton and Ockenfels (2000)

In the 'equity-reciprocity-competition' (ERC) model of Bolton and Ockenfels (2000) agents maximize an objective function $u_i(\pi_i, \sigma_i)$, where π_i is player i's monetary payoff and $\sigma_i(w, e_i, e_j) = \pi_i(w, e_i) / \Pi(e_i, e_j)$ is player i's share of the total payoff (with $\Pi(e_i, e_j) = \pi_i(w, e_i) + \pi_j(w, e_j) + \pi_P(w, e_i, e_j)$, payoff functions defined in the design section).[17] In the following we treat e_i and e_j as continuous choice variables. Regarding the derivatives with respect to the first and second argument the ERC model assumes $u_{i1} \geq 0$, $u_{i11} \leq 0$, $u_{i2} = 0$ for $\sigma_i = 1/n$, and $u_{i22} < 0$.[18] As a result, the motivation function is strictly concave in the income share, and, for a given payoff π_i, it is maximised when i earns exactly the equal share, i.e., $\sigma_i = 1/n$. An ERC worker maximises his objective function $u_i[\pi_i(w, e_i), \sigma_i(w, e_i, e_j)]$. Worker i's first order condition is:

$$\frac{\partial u_i}{\partial e_i} = \frac{\partial u_i}{\partial \pi_i}\frac{\partial \pi_i}{\partial e_i} + \frac{\partial u_i}{\partial \sigma_i}\frac{\partial \sigma_i}{\partial e_i} \equiv u_{i1}\pi_{i2} + u_{i2}\sigma_{i2} = 0 \tag{4}$$

How does the worker react to the co-worker's effort choice? In order to derive the slope of the reaction function calculate the total differential of equation (4):

$$\left[u_{i11}\left(\pi_{i2}\right)^2 + u_{i1}\,\pi_{i22} + u_{i22}\left(\sigma_{i2}\right)^2 + u_{i2}\,\sigma_{i22}\right]de_i + \left[u_{i22}\,\sigma_{i3}\,\sigma_{i2} + u_{i2}\,\sigma_{i23}\right]de_j = 0\,. \tag{5}$$

From the definitions given in the main text we easily can show that $\pi_{i2} = \text{-}7$, $\Pi_1 = \Pi_2 = (v - 7) > 0$, and all higher order derivatives are zero. Furthermore, interior solutions are only possible when worker i earns at least the equal share, which implies $u_{i2} < 0$. The derivatives of the relative payoff function are:

[17] The ERC model requires nonnegative payoffs. In order to avoid negative outcomes we calculate all payoffs including the endowment of 400 ECU. Overall losses are not possible in this case.

[18] u_{i1} denotes the derivative of u_i with respect to the first argument (π_i). u_{i11} denotes the second order derivative and u_{i12} denotes cross partial derivatives.

$$\sigma_{i2} = \frac{\overset{-}{\pi_{i2}}\,\overset{+}{\Pi} - \overset{+}{\pi_i}\,\overset{+}{\Pi_1}}{\Pi^2} < 0 \; ; \qquad \sigma_{i3} = -\frac{\overset{+}{\pi_i}\,\overset{+}{\Pi_2}}{\Pi^2} < 0$$

$$\sigma_{i22} = \frac{\overset{0}{\pi_{i22}}\,\overset{+}{\Pi^2} - 2\,\overset{-}{\pi_{i2}}\,\overset{+}{\Pi_1}\,\overset{+}{\Pi} + 2\pi_i\left(\Pi_1\right)^2 - \pi_i\,\overset{0}{\Pi_{11}}\,\Pi}{\Pi^3} > 0 \tag{6}$$

$$\sigma_{i23} = -\frac{\overset{-}{\pi_{i2}}\,\overset{+}{\Pi_1}\,\overset{+}{\Pi} - 2\,\overset{+}{\pi_i}\,\overset{+}{\Pi_1}\,\overset{+}{\Pi_2} + \pi_i\,\overset{0}{\Pi_{11}}\,\Pi}{\Pi^3} > 0$$

Hence, we can derive the slope of the reaction function:

$$\frac{de_i}{de_j} = -\frac{\overset{-}{u_{i22}}\,\overset{-}{\sigma_{i3}}\,\overset{-}{\sigma_{i2}} + \overset{-}{u_{i2}}\,\overset{+}{\sigma_{i23}}}{\underset{-}{u_{i11}}\underset{+}{\left(\pi_{i2}\right)^2} + \underset{-}{u_{i22}}\underset{+}{\left(\sigma_{i2}\right)^2} + \underset{-}{u_{i2}}\,\underset{+}{\sigma_{i22}}} < 0\,. \tag{7}$$

Furthermore, the expressions in (6) show that the effects of e_i on i's payoff share are stronger than the effects of e_j, i.e., $|\sigma_{i2}| > |\sigma_{i3}|$ and $\sigma_{i22} > \sigma_{i23}$. Thus, we can conclude that the slope of the reaction function must lie between 0 and –1 for interior solutions:

$$-1 < \frac{de_i}{de_j} < 0 \quad \text{for } 1 < e_i < 20\,. \tag{8}$$

Consider the extreme case of a perfectly inequality averse ERC agent who exclusively cares for the relative payoff argument. This worker would choose an effort such that, whenever possible, his income share equals exactly one third, i.e., $\sigma_i(w, e_i, e_j) = 1/3$. Solving for e_i gives the worker's reaction function: [19]

$$e_i = R^{ERC}\left(e_j\right) = \left[\frac{7 + 3w - (v-7)e_j}{14 + v}\right]_{[1}^{20]}. \tag{9}$$

C. Fehr-Schmidt (*1999*)

A Fehr-Schmidt worker maximizes the following utility function:

$$u_i = \pi_i - \frac{\alpha_i}{2}\Big[\max\left(\pi_j - \pi_i, 0\right) + \max\left(\pi_P - \pi_i, 0\right)\Big] - \frac{\beta_i}{2}\Big[\max\left(\pi_i - \pi_j, 0\right) + \max\left(\pi_i - \pi_P, 0\right)\Big], \tag{10}$$

where π_i and π_F are functions of e_i. The derivative with respect to e_i is:

$$\frac{\partial u_i}{\partial e_i} = \begin{cases} -7 - 0.5\beta_i\left(-7 - \left(v+7\right)\right) & \text{if } \pi_i > \pi_j \text{ and } \pi_i > \pi_P \quad \text{(i)} \\ -7 - 0.5\alpha_i\left(v+7\right) - 0.5\beta_i\left(-7\right) & \text{if } \pi_i > \pi_j \text{ and } \pi_i < \pi_P \quad \text{(ii)} \\ -7 - 0.5\alpha_i\left(7\right) - 0.5\beta_i\left(-\left(v+7\right)\right) & \text{if } \pi_i < \pi_j \text{ and } \pi_i > \pi_P \quad \text{(iii)} \\ -7 - 0.5\alpha_i\left(7 + \left(v+7\right)\right) & \text{if } \pi_i < \pi_j \text{ and } \pi_i < \pi_P \quad \text{(iv)} \end{cases} \tag{11}$$

---

[19] The notation $\left[x\right]_{[1}^{20]}$ is equivalent to max[min(x, 20), 1].

The expressions in (11) are independent of e_i but the case differentiation is not (the arguments of the payoff functions are omitted). Thus, in general a Fehr-Schmidt worker will choose his effort such that his payoff is either maximal or equals one of the other player's payoff. The two thick graphs in Figure A1 show the locations in the (e_j, e_i) space where this is the case (for $w = 200$ and $v = 35$).

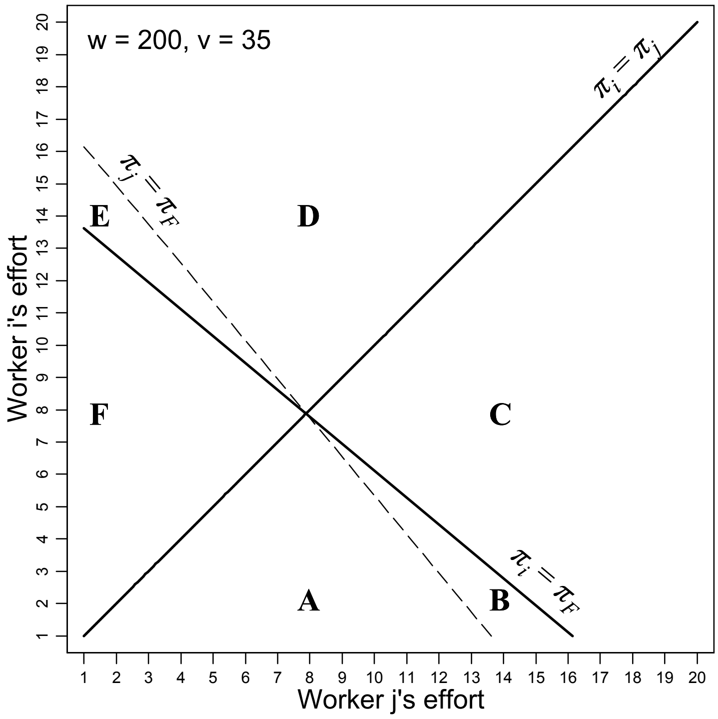

Figure A1: Locations of equal payoffs dependent on e_i and e_j.

Starting with case (i) where worker *i* earns the highest income (area A and B in Figure A1). The derivative is positive if $\beta_i > 14/(v+14)$. A worker with a lower β_i will always choose minimal effort. A worker with a sufficiently high β_i will increase his effort as long as his income is highest.[20] Case (ii) corresponds to the area C in Figure A1. In this situation the derivative is unambiguously negative due to the parameter restriction $\alpha_i \geq \beta_i$ given by Fehr and Schmidt. Thus, the reaction function cannot lie in C. The most interesting situation is case (iii) which corresponds to the area F. If the preference parameters satisfy $(v+7)\beta_i > 14 + 7\alpha_i$ then worker *i* is called *Strong Egalitarian* (*SE*).[21] Such a worker increases his effort to adjust his income to the principal's income. In the opposite case worker *i* is called *Weak Egalitarian* (*WE*), and he decreases his effort to adjust his income to the other

[20] For notational convenience we assume that the preference parameters α_i and β_i are drawn from a continuous density function and therefore rule out the cases where the parameters equal any of the critical values. We thereby get rid of the (not very interesting but notationally tedious) cases where the players are exactly indifferent between several effort levels within a range.

[21] The interpretation of this condition is straightforward: A Strong Egalitarian worker's concern for 'advantageous inequality' (β_i) must be relatively strong compared to his concern for 'disadvantageous inequality' (α_i)

worker's income. Finally, in case (iv) or area D and E the derivative is unambiguously negative.

To conclude, a *Strong Egalitarian* worker chooses his effort such that his income equals the principal's income, i.e. $\pi_i = \pi_P$.[22] Solving for e_i gives the reaction function of a *SE* worker:

$$e_i = R^{FS,SE}\left(e_j\right) = \left[\frac{3w+7-ve_j}{v+7}\right]_{[1}^{20]}. \tag{12}$$

Thus, the slope of the reaction function is within $-1 < \partial e_i / \partial e_j = -v/\left(v+7\right) < 0$ for interior solutions, independent of *w*. A *WE* type wants to match the effort of worker *j* up to a certain threshold. The threshold is reached when the principal's income is equal to the income of the two workers (which must be equal in all multiple equilibria). From there on, worker *i*'s reaction function is identical to the reaction function of a *SE* worker. The reaction function of a *Weak Egalitarian* can be written as

$$R^{FS,WE}(e_j) = \begin{cases} e_j & \text{if } 1 \le e_j \le \left(3w+7\right)/\left(2v+7\right) \\ R^{FS,SE}(e_j) & \text{else.} \end{cases} \tag{13}$$

D. Falk and Fischbacher (*2006*)

The utility function proposed by Falk and Fischbacher has the following (simplified) form:

$$U_i = \pi_i + \rho_i \sum_j \vartheta_j \Delta_j \sigma_i \,. \tag{14}$$

Thereby, π_i is the monetary outcome, $\rho_i > 0$ is a preference parameter measuring the importance of reciprocal motives. For every player *j* the reciprocal motivation is captured by the following terms: ϑ_j is the intention factor, which is unity if player *j* acted intentionally and $0 \le \varepsilon_i \le 1$ otherwise.[23] The outcome term Δ_j measures the kindness of *j* towards *i* and the reciprocation term σ_i, measures the kindness of *i*'s reaction. In the context of the three person gift-exchange game the utility function for worker *i* is:

$$\begin{aligned} u_i = \pi_i\,(w,e_i) \quad &+ \rho_i \vartheta_P \left[\pi_i(w,e_i'') - \pi_P(w,e_i'',e_j'')\right]\left[\pi_P(w,e_i,e_j') - \pi_P(w,e_i'',e_j')\right] \\ &+ \rho_i \vartheta_j \left[\pi_i\,(w,e_i'') - \pi_j\,(w,e_j'')\right]\left[\pi_j(w,e_j') - \pi_j(w,e_j')\right]. \end{aligned} \tag{15}$$

The first expression is worker *i*'s monetary payoff. The second addend is the reciprocity term towards the principal. The outcome term (first squared bracket) depends on worker *i*'s belief about the principal's belief about the two efforts. The first expression in the squared

[22] Note that this does not mean that a *SE* worker is especially loyal towards the employer. It is rather the fact that a change in effort of one reduces or increases the inequality towards the other worker by 7 and towards the principal by *v*+7. In this sense, reducing inequality is more effective when adjusting to the principal's income.

[23] The parameter ε_i is a second preference parameter that measures the importance of payoff comparisons relative to the importance of reciprocal motives. Thereby, $\varepsilon_i = 1$ means that intentions do not matter while $\varepsilon_i = 0$ describes the case where the other workers payoff is only taken into account if his actions are intentional.

brackets shows the worker's belief about the principal's kindness. This expression depends on the worker's beliefs about the principal's beliefs about the two efforts. These second order beliefs are denoted as e_i'' and e_j''. If this term is positive, worker *i* perceives the principal's actions as kind. The second expression in squared brackets shows the reciprocal reaction, i.e., the influence of worker *i*'s effort on the principal's income (given worker *i*'s belief about the other worker's effort e_j'). The third addend shows the reciprocity towards the other worker. Since the workers cannot influence each others' incomes the third addend is independent of e_i and therefore irrelevant for worker *i*'s optimization. If we differentiate the utility function (15) with respect to e_i we can write the first order condition as:

$$\frac{\partial u_i}{\partial e_i} = -7 + \rho_i \vartheta_P \left[w - 7(e_i'' - 1) - \left\{ v(e_i'' + e_j'') - 2w \right\} \right] v = 0. \qquad (16)$$

In equilibrium, beliefs must be consistent, i.e., $e_i = e_i' = e_i''$ and $e_j = e_j' = e_j''$. From the first order condition we can therefore derive worker *i*'s reaction function as

$$e_i = R^{FF}(e_j) = \left[\frac{1}{v+7} \left(3w + 7 - \frac{7}{v \rho_i \vartheta_P} - v e_j \right) \right]_{[1}^{20]}. \qquad (17)$$

Thus, the model of reciprocity by Falk and Fischbacher predicts a reaction function that is linear in the other worker's effort with a slope of $-1 < \partial e_i / \partial e_j = -v/(v+7) < 0$. For interior solutions the slope is independent of the preference parameter ρ_i and the intention factor ϑ_F. The intention factor is equal to one if the principal pays a wage of 100 or 200. This is an intentionally kind act, since it could have made the worker worse off by paying a wage of 50. Being paid a wage of 50, on the other hand, is perceived as non-intentional and therefore we set $\vartheta_P = \varepsilon_i$. The preference parameter ρ_i has a very straightforward influence on the reaction function: no concern for reciprocity $(\rho_i \to 0)$ leads to minimal effort, a higher concern for reciprocity shifts the reaction function upwards. The limit case $(\rho_i \to \infty)$ is identical to the reaction function of a *SE* worker in the Fehr-Schmidt model.

E. Charness and Rabin (*2002*)

Charness and Rabin (2002) start by proposing a reciprocity-free utility function:

$$u_i(\pi) = (1-\lambda)\pi_i + \lambda \left[\delta \max\left\{ \pi_i, \pi_j, \pi_P \right\} + (1-\delta)\left(\pi_i + \pi_j + \pi_P \right) \right]. \qquad (18)$$

Thereby λ and δ are preference parameters; $\lambda \in [0,1]$ weighs a subjects social preferences in general and $\delta \in [0,1]$ determines whether the social preferences are rather maximin preferences $(\delta = 1)$ or utilitarian preferences $(\delta = 0)$. The derivative with respect to e_i is:

$$\frac{\partial u_i}{\partial e_i} = (1-\lambda)(-7) + \lambda \delta x + \lambda(1-\delta)(v-7) \quad \text{with } x = \begin{cases} v & \text{if } \pi_F = \min\left\{ \pi_i, \pi_j, \pi_P \right\} \quad \text{(i)} \\ 0 & \text{if } \pi_j = \min\left\{ \pi_i, \pi_j, \pi_P \right\} \quad \text{(ii)} \\ -7 & \text{if } \pi_i = \min\left\{ \pi_i, \pi_j, \pi_P \right\} \quad \text{(iii)} \end{cases} \qquad (19)$$

We have to consider three cases depending on whose income is minimal. Irrespective of the case we can calculate thresholds for the preference parameters resulting in a positive derivative:

$$\lambda > \frac{7}{\delta\left(7 - v + x\right) + v} > 0 \tag{20}$$

Note that the expression is monotonically decreasing in x. There are four different reaction functions depending on the number of inequalities that are satisfied given the three different values of x.[24] If the inequality is never satisfied, then the worker will choose minimal effort. If only the 'weakest' inequality with $x = v$ is satisfied then worker i is ready to choose a non-minimal effort whenever the principal's income is minimal. We call this type a "Maximin" worker. The reaction function for such a type is the upper boundary of the area A and F in Figure A1.

If the inequality is also satisfied for $x = 0$ then the worker wants to increase his effort whenever one of the other player's income is minimal. This is the case in the areas A, B, C, and F in Figure A1. Like before, the reaction function is the upper boundary of this area. We call this type a "Utilitarian" worker.[25] If the preference parameters are such that all three inequalities are satisfied then the worker chooses the maximum effort irrespective of what the other players do. To conclude the reaction functions predicted by Charness and Rabin are either minimum or maximum effort or one of the follwing two:

$$\begin{aligned} e_i = R^{CR,\text{maximin}}(e_j) &= \begin{cases} R^{FS,SE}(e_j) & \text{if } e_j \le \left(3w+7\right)/\left(2v+7\right) \\ \left(3w+7-(v+7)e_j\right)/v & \text{else} \end{cases} \\ e_i = R^{CR,\text{utilitarian}}(e_j) &= \begin{cases} R^{FS,SE}(e_j) & \text{if } e_j \le \left(3w+7\right)/\left(2v+7\right) \\ e_j & \text{else} \end{cases} \end{aligned} \tag{21}$$

In a second step Charness and Rabin enrich the model with reciprocity. They introduce demerit parameters, in our case $d_j, d_F \in [0,1]$, indicating (inversely) how much the other player deserves to be treated nicely. The utility function is

$$\begin{aligned} u_i(\pi) = \left(1-\lambda\right)\pi_i + \lambda\Big[&\delta \max\left\{\pi_i, \pi_j + bd_j, \pi_P + bd_P\right\} + \\ &\left(1-\delta\right)\left(\pi_i + \pi_j[1-kd_j]_{[0} + \pi_P[1-kd_P]_{[0}\right) - f\left(d_j\pi_j + d_P\pi_P\right)\Big], \end{aligned} \tag{22}$$

with the nonnegative parameters b and k for the weight of the demerit parameter in the rawlsian and the utilitarian part of the utility function. The paramter $f > 0$ allows to account for destructive behavior. Note that in case there are no hard feelings with respect to the other players ($d_j, d_P = 0$) the utility function is identical to (18). The derivative with respect to e_i is

[24] Like in the case of Fehr and Schmidt we ignore the cases where conditions on preference parameters are met with equality.

[25] We call the intermediate types "Maximin" and "Utilitarian" because, at any given $\lambda > 7/v$, the range of δ giving rise to the former type is higher than for the latter type. However, perfectly utilitarian behavior (δ=0) will in general result in a reaction function at maximum effort. In this sense it would be more accurate to speak of "moderate utilitarian" for the third type of worker behavior.

$$\frac{\partial u_i}{\partial e_i} = (1-\lambda)(-7) + \lambda\delta x' + \lambda(1-\delta)\left(-7 + [1-kd_P]_{[0} v\right) - fd_P v$$

$$\text{with } x' = \begin{cases} v & \text{if } \pi_P + bd_P = \min\{\pi_i, \pi_j + bd_j, \pi_P + bd_P\} \\ 0 & \text{if } \pi_j + bd_j = \min\{\pi_i, \pi_j + bd_j, \pi_P + bd_P\} \\ -7 & \text{if } \pi_i = \min\{\pi_i, \pi_j + bd_j, \pi_P + bd_P\} \end{cases} \tag{23}$$

The implications of the additional parameters for the reaction function are quite straightforward: For positive demerits the paramters k and f determine (in combination with λ and δ) which one of the four reaction functions is chosen. The parameter b influences the position of the reaction function. In general, a higher b shifts the reaction functions $R^{CR,\ maximin}$ and $R^{CR,\ utilitarian}$ downwards. If we assume $d_j = 0$ (which is plausible since no information about j's actions are available) then a higher b shifts the two reaction functions downwards leaving the kink on the 45 degree line. For very high b the reaction function $R^{CR,\ utilitarian}$ is identical to the 45 degree line. However, none of the parameters has a direct influence on the slope of the reaction function.

F. Cox, Friedman and Gjerstad (2007)

The model of Cox, Friedman and Gjerstad (2007) posits a CES utility function:

$$u_i = \alpha^{-1}\left(\pi_i^\alpha + \theta_j \pi_j^\alpha + \theta_P \pi_P^\alpha\right) \quad \text{for } -\infty < \alpha \le 1 \text{ and } \alpha \ne 0. \tag{24}$$

Thereby, α is a preference parameter allowing for difference marginal rates of substitution (*MRS*) between worker i's income and the other players' incomes.[26] The parameters θ_j and θ_P measure the emotional state of player i towards the other two players, where positive values indicate gratitude and negative values resentment. The derivative with respect to e_i is

$$\frac{\partial u_i}{\partial e_i} = \left[400 + w - 7e_i + 7\right]^{\alpha-1}(-7) + \theta_P\left[400 + v\left(e_i + e_j\right) - 2w\right]^{\alpha-1} v. \tag{25}$$

Because the payoff of worker j does not depend on i's choices, θ_j drops out. The expressions in the squared brackets are positive. Thus, in case of neutral or negative emotions towards the principal $(\theta_P \le 0)$ the derivative is negative and worker i will always choose minimal effort. In case of positive emotions the reaction function results as:

$$e_i = R^{CFG}(e_j) = \left[\frac{400 + w + 7 + a\left(2w - ve_j - 400\right)}{av + 7}\right]_{[1}^{20]} \qquad \text{with } a = \left(\frac{\theta_P v}{7}\right)^{\frac{1}{\alpha-1}} \tag{26}$$

Since positive emotions ($\theta_P > 0$) imply $a > 0$, the slope of the reaction function lies between –1 and zero:

[26] α = 1 corresponds to a constant MRS and $\alpha \to \infty$ corresponds to a Leontief function. For α = 0 the utility function converges to a Cobb-Douglas case. We will not consider this case here.

$$-1 < \frac{\partial e_i}{\partial e_j} = -\frac{av}{av+7} < 0 \quad (27)$$

Note that for either $\theta_P \to 7/v$ or $\alpha \to -\infty$ the expression for a goes to unity and the reaction function is identical to the case of Fehr-Schmid *SE* workers. For other values of α the reaction function is either above $R^{FS,SE}$ and flatter (if $\theta_P > 7/v$) or below $R^{FS,SE}$ and steeper (otherwise).

8. References

BANDIERA, O., I. BARANKAY, and I. RASUL (2007): "Social Incentives in the Workplace," Mimeo, London School of Economics.

BARDSLEY, N., and R. SAUSGRUBER (2005): "Conformity and Reciprocity in Public Good Provision," *Journal of Economic Psychology*, 26, 664-681.

BELLEMARE, C., and B. SHEARER (2007): "Gift Exchange within a Firm: Evidence from a Field Experiment," IZA Discussion Paper No. 2696.

BEWLEY, T. (1999): *Why Wages Don't Fall in a Recession*. Cambridge: Harvard University Press.

BOLTON, G. E., and A. OCKENFELS (2000): "Erc: A Theory of Equity, Reciprocity, and Competition," *American Economic Review*, 90 1, 166-93.

CHARNESS, G. (2004): "Attribution and Reciprocity in an Experimental Labor Market," *Journal of Labor Economics*, 22, 665-688.

CHARNESS, G., and M. RABIN (2002): "Understanding Social Preferences with Simple Tests," *Quarterly Journal of Economics*, 117 3, 817-69.

COX, J. C., D. FRIEDMAN, and S. GJERSTAD (2007): "A Tractable Model of Reciprocity and Fairness," *Games and Economic Behavior*, 59, 17-45.

DUFWENBERG, M., and G. KIRCHSTEIGER (2004): "A Theory of Sequential Reciprocity," *Games and Economic Behavior*, 47, 268-298.

FALK, A., and U. FISCHBACHER (2006): "A Theory of Reciprocity," *Games and Economic Behavior*, 54, 293-315.

FALK, A., U. FISCHBACHER, and S. GÄCHTER (2004): "Living in Two Neighborhoods: Social Interactions in the Lab," IZA Discussion Paper No. 1381.

FALK, A., and A. ICHINO (2006): "Clean Evidence on Peer Effects," *Journal of Labor Economics*, 24, 38-57.

FEHR, E., and S. GÄCHTER (2000): "Fairness and Retaliation: The Economics of Reciprocity," *Journal of Economic Perspectives*, 14, 159-181.

FEHR, E., S. GÄCHTER, and G. KIRCHSTEIGER (1997): "Reciprocity as a Contract Enforcement Device: Experimental Evidence," *Econometrica*, 65, 833-860.

FEHR, E., E. KIRCHLER, A. WEICHBOLD, and S. GÄCHTER (1998): "When Social Norms Overpower Competition: Gift Exchange in Experimental Labor Markets," *Journal of Labor Economics*, 16, 324-351.

FEHR, E., G. KIRCHSTEIGER, and A. RIEDL (1993): "Does Fairness Prevent Market Clearing - an Experimental Investigation," *Quarterly Journal of Economics*, 108, 437-459.

FEHR, E., and K. M. SCHMIDT (1999): "A Theory of Fairness, Competition, and Cooperation," *Quarterly Journal of Economics*, 114 3, 817-68.

FISCHBACHER, U. (2007): "Z-Tree: Zurich Toolbox for Readymade Economic Experiments," *Experimental Economics*, 10, 171-178.

GÄCHTER, S., and C. THÖNI (2008): "Wage Comparison and Performance,"

ICHINO, A., and G. MAGGI (2000): "Work Environment and Individual Background: Explaining Regional Shirking Differentials in a Large Italian Firm," *Quarterly Journal of Economics*, 115, 1057-1090.
LEVINE, D. K. (1998): "Modeling Altruism and Spitefulness in Experiments," *Review of Economic Dynamics*, 1 3, 593-622.
MANSKI, C. F. (1993): "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies*, 116, 607-654.
— (2000): "Economic Analysis of Social Interactions," *Journal of Economic Perspectives*, 13, 115-136.
MAS, A., and E. MORETTI (2007): "Peers at Work," Mimeo, University of Berkeley.
ORNE, M. T. (1962): "On the Social Psychology of the Psychological Experiment: With Particular Reference to Demand Characteristics and Their Implications," *American Psychologist*, 17, 776-783.
SCHLAG, K. (1998): "Why Imitate, and If So, How? A Boundedly Rational Approach to Multi-Armed Bandits," *Journal of Economic Theory*, 78, 130-156.
WEBER, R. A. (2003): "'Learning' with No Feedback in a Competitive Guessing Game," *Games and Economic Behavior*, 44, 134-144.