Journal of Economic Psychology 48 (2015) 72–88

Contents lists available at ScienceDirect

Journal of Economic Psychology

journal homepage: www.elsevier.com/locate/joep

Peer effects and social preferences in voluntary cooperation: A theoretical and experimental analysis

Christian Thöni [a], Simon Gächter [b,c,*]

[a] University of Lausanne, UNIL-Dorigny, 1015 Lausanne, Switzerland
[b] School of Economics, University of Nottingham, Sir Clive Granger Building, University Park, Nottingham NG7 2RD, United Kingdom
[c] CESifo & IZA, Germany

ARTICLE INFO

Article history:
Received 6 May 2014
Received in revised form 22 January 2015
Accepted 2 March 2015
Available online 18 March 2015

JEL classification:
C92
D03

PsycINFO classification:
2260
2360
3020

Keywords:
Social preferences
Voluntary cooperation
Peer effects
Reflection problem
Gift-exchange
Conformism
Social norms
Social esteem
Experiments

ABSTRACT

Social preferences and social influence effects (“peer effects”) are well documented, but little is known about how peers shape social preferences. Settings where social preferences matter are often situations where peer effects are likely too. In a gift-exchange experiment with independent payoffs between two agents we find causal evidence for peer effects. Efforts are positively correlated but with a kink: agents follow a low-performing but not a high-performing peer. This contradicts major theories of social preferences which predict that efforts are unrelated, or negatively related. Some theories allow for positively-related efforts but cannot explain most observations. Conformism, norm following and social esteem are candidate explanations.

1. Introduction

Is pro-social voluntary cooperation subject to ‘peer effects’, that is, influenced by the behavior of comparison others (‘peers’)? Or is pro-sociality a preference that is largely immune to social influence? These questions are at the heart of this

* Corresponding author at: School of Economics, University of Nottingham, Sir Clive Granger Building, University Park, Nottingham NG7 2RD, United Kingdom.

E-mail addresses: christian.thoeni@unil.ch (C. Thöni), simon.gaechter@nottingham.ac.uk (S. Gächter).
URL: http://sites.google.com/site/christianthoeni (C. Thöni).

http://dx.doi.org/10.1016/j.joep.2015.03.001
0167-4870/

paper. Understanding how peer effects shape social preferences is important because people normally do not act in a social vacuum but are constantly exposed to peers. The purpose of this paper is to understand how social preferences and peer effects are linked both by providing novel experimental evidence and by clarifying what theories of social preferences have to say on peer effects.

We speak of a peer effect if an agent is influenced in his or her actions by what a comparison agent does, *even if there are no material spillovers between agents* and hence no *direct* social preference links exist between peers (indirect links are possible, as our theoretical analysis will show). Our definition is narrower than frequent usage of the term in the literature. Sometimes 'peer effects' is used as an umbrella term to describe behaviors where an agent reacts to other agents' actions; such a behavioral reaction might be motivated by social preferences. As an example, think of voluntary contributions to a public good, where people contribute more the more other group members contribute. Such 'conditional cooperation' can be due to peer effects, but also social preferences.

Understanding the link between social preferences and peer effects is important for two reasons. First, in reality social preferences are often relevant in environments that are potentially rich in possibilities for social influence effects. Think of the workplace as a prime example. In a seminal field experiment, Falk and Ichino (2006) show that work effort is subject to peer effects which were suggested to be important in non-experimental empirical work (Ichino & Maggi, 2000). Yet, their study does not allow linking such peer effects to theories of social preferences, which is our main contribution.

Second, suppose we find evidence for peer effects. What would the implication be for theories of social preferences? For example, in popular theories of inequality aversion people's social preferences are modeled as individually *fixed* distastes for inequitable outcomes. Evidence for peer effects would constitute a *prima facie* challenge to fixed preference assumptions. We deem it important to clarify theoretically whether peer effects indeed provide such a challenge or whether existing theories assuming fixed preferences can explain the data before resorting to changes in social preference parameters as an explanation.

Our tool to measure peer effects is a one-shot three-person gift-exchange game where a principal pays his two agents i and j a wage w (the same for both) and the agents choose efforts e_i and e_j. The material incentive structure gives both agents an incentive to choose minimal effort irrespective of w and irrespective of the other agent's effort. However, from numerous two-person gift-exchange games we expect that many agents choose efforts that increase in wage. In this situation we will speak of a 'peer effect' if $e_i = f(e_j)|_w$, that is, holding the common wage w constant, agent i's effort depends on agent j's effort ($f' \neq 0$), despite the absence of any earnings interdependency between agents.

A major problem of measuring peer effects empirically is the "reflection problem" (e.g., Manski, 2000): $e_i = f(e_j)|_w$ and $e_j = f(e_i)|_w$. If i is influenced by j and j is influenced by i it is impossible to disentangle the causal influences i and j have on each other. Here we propose a novel design that avoids the reflection problem. The main innovation is to make the effort of the other agent exogenous. To achieve this, both agents first choose their efforts simultaneously and then, after having learned the effort decision of their co-agent, are given the opportunity to *revise* their effort, *holding their co-agent's effort constant*. Since the design removes any material and strategic incentives to revise effort, revision decisions (compared to a control condition with no effort information) tell us about the extent to which people change their effort *because* of the effort chosen by the co-agent.

Our results provide unambiguous evidence for peer effects. Effort revisions are more likely and bigger when agents are informed about their co-agent's effort than when they are uninformed. When the co-agent has provided lower effort than them they revise their efforts downwards, but they barely increase their effort when the co-agent provided higher effort.

To see whether these peer effects are a novel phenomenon that is incompatible with existing theories of social preferences, we analyze the theoretical predictions of widely used theories of social preferences that model various distributional and/or intentional concerns under the assumption that preferences are fixed. We focus on the best-reply predictions with regard to effort changes, that is, de_i/de_j. To our knowledge, no such analysis has been done in the context of explaining peer effects in voluntary cooperation.

The most robust predictions of these standard theories of social preferences are that either there are no peer effects (efforts are unrelated in models of reciprocity), or if there are peer effects, efforts are negatively related (in all other models). Our finding of peer effects with positively correlated efforts seems therefore inconsistent with most models. However, this evidence is not fully conclusive because predictions are about the agents' best-reply functions, which our simple revision decisions do not reveal.

To have a conclusive test we ran experiments where we also elicited the agents' *beliefs* about the initial effort choice of their co-agent. Thus, we now observe two points on each agent's best-response which allows us to draw conclusions about the slope of the best replies. The results reject the prediction of most theories that efforts are negatively related. In the peer effect we observe, efforts are strategic complements, not substitutes. Also the theories that predict positively correlated efforts are only consistent with a minority of choices.

While standard theories of social preferences which model equality concerns and/or intentions, typically predict the opposite of what we find, recent theories that incorporate social motives like conformism (Sliwka, 2007), norm-following (López-Pérez, 2008), or social esteem (Ellingsen & Johannesson, 2008) can explain the peer effects we observe.

Our paper is structured as follows. Section 2 places our paper in the related literature. Section 3 introduces our experimental game and procedures. Section 4 presents our results on peer effects. Section 5 discusses the implications of various models of social preferences in our game. In Section 6 we analyze data on agents' beliefs to identify the slopes of the reaction function which allows us to contrast data and theories. Section 7 discusses the significance of the findings and alternative theoretical explanations. Section 8 summarizes the paper.

2. Related literature and our contribution

Hitherto the literature[s] on social influence effects and social preferences are, with a few exceptions, largely unconnected. The literature on social preferences shows that many people are willing to act against their self-interest even in anonymous one-shot situations with incentives to behave selfishly and no possibilities for social influence (e.g., Gintis, Bowles, Boyd, & Fehr, 2005). The literatures on social influence effects show that people's behavior in many economically important domains is often strongly shaped by what comparison others do even in the absence of material payoff spillovers.[1] Similarly, social psychologists have long argued that situational cues (provided by the environment or the behavior of others) are often more important than personality traits (Asch, 1952; Ross & Nisbett, 1991). Both social preferences and social influence effects are firmly established empirically, but little is known about how social preferences are influenced by, and related to, peer effects.

In connecting these literatures, our paper makes two main and intertwined contributions. Our first contribution is to provide novel and causal evidence for peer effects in reciprocity by using the gift-exchange game run in the direct response method and by ruling out confounding factors such as strategic incentives that might be present, for instance, in repeated interactions (as in Falk, Fischbacher, & Gächter, 2013, who used finitely repeated simultaneous public goods games to study peer effects in voluntary cooperation). The use of the direct-response method is one distinguishing feature of this paper relative to Gächter, Nosenzo, and Sefton (2012) and Gächter, Nosenzo, and Sefton (2013) who also use a three-person gift-exchange game but play it in the strategy method. Both studies find evidence for peer effects which take the form of positively correlated efforts, but by their strategy-method designs are unable to observe the kink in efforts we are able to observe in our design. Moreover, both papers also allow for wage inequality, which our design precludes for simplicity.

Our use of the gift-exchange game separates our study from papers that investigate and find peer effects (in the absence of any payoff spillovers like in our case) in sharing in the dictator game (Bicchieri & Xiao, 2009; Cason & Mui, 1998; Krupka & Weber, 2009) and in the ultimatum game (Ho & Su, 2009). Papers more closely related to the experimental part of our paper are Bardsley and Sausgruber (2005) and Mittone and Ploner (2011). Bardsley and Sausgruber provide evidence that contributions to a one-shot public good can be shaped by observing the contributions of an unrelated other group. Mittone and Ploner also study a sequential game of reciprocity, in their case the trust game, and unlike us, they use the strategy method. Their central tool is an observer/observed design with the difference to our study being that the relevant group of recipients is comprised of groups of four, two of whom are observed and two who are the observers. The return rates ('trustworthiness') of the observed and observer recipients are positively correlated, which is evidence for peer effects in reciprocity. In our design, everyone is an observer and is observed, while strategic incentives to influence others are absent.

Our second contribution is to clarify the power of a range of theories of social preferences for explaining peer effects in reciprocity. We provide a formal analysis and then run further experiments to compare the best-reply predictions with the empirically estimated slope. Such an analysis is methodologically important to safeguard against premature declarations of peer effects as a phenomenon that requires separate explanations from those of existing theories.

We deem it important not to pre-select one particular theory (most typically, the Fehr & Schmidt, 1999 model) but to look at a whole range of theories, which capture different psychological motivations. Such an analysis can narrow the range of theories that are candidate explanations for peer effects, even if we acknowledge that one lesson of tests of theories of social preferences is that all fail in some dimensions (e.g., Charness & Rabin, 2002; Cox, Friedman, & Gjerstad, 2007; Daruvala 2010; Engelmann & Strobel, 2004; Falk, Fehr, & Fischbacher, 2008) and maybe especially so in three-person games (e.g., Kagel & Wolfe, 2001). A comprehensive comparative analysis can provide insights about patterns of failures across classes of theories that are informative in their own right.

3. Design and procedures

3.1. A design to avoid the reflection problem: the three-person gift-exchange game with a revision stage

Our three-person gift-exchange game is an extension of the two-player gift-exchange game (Fehr, Kirchsteiger, & Riedl, 1993) – there is one principal and two identical agents. The principal first chooses the same wage $w \in \{50, 100, 200\}$ for both agents. After observing w, the two agents decide simultaneously about their effort, that is, they choose $e_i \in \{1, 2, \ldots, 20\}$.

Our central tool to avoid the reflection problem is the introduction of a "revision stage". Agents learn about the *revision stage* where they are informed about the 'initial effort' decision of their co-agent, e_j.[2] In light of this new information agents

[1] Some examples comprise deviant behavior (Sampson, Morenoff, & Gannon-Rowley, 2002), academic success (e.g., Sacerdote, 2001), savings behavior (e.g., Duflo & Saez, 2002); corruption (e.g., Dong, Dulleck, & Torgler, 2012); tax evasion (e.g., Prinz, Muehlbacher, & Kirchler, 2014); health-related issues like alcohol consumption (Kremer & Levy, 2008) and obesity (Christakis & Fowler, 2007) and behavior in the workplace (Bandiera, Barankay, & Rasul, 2010; Ichino & Maggi, 2000; Mas & Moretti, 2009). A few studies have investigated field settings where observed results are consistent with peer effects, but also social preferences, such as conditional cooperation in naturally occurring public goods (Chen, Harper, Konstan, & Li, 2010; Frey & Meier, 2004); and charitable donations (Croson & Shang, 2008; Shang & Croson, 2009).

[2] When agents decide on their initial effort they do not yet know about the possibility to revise effort and about the fact that their effort choice would be communicated to the other agent. This is necessary to avoid that the initial effort is strategically biased, which would preclude a clean measurement of peer effects. The information about the revision possibility and its description appeared on a separate screen (see Appendix A). The reader may ask why this procedure rather than letting the agents choose their efforts sequentially. We could then test whether the second mover's effort depends on the effort of the first moving agent. However, it is difficult to disentangle peer effects from the second-moving agent's disposition to reciprocate toward the principal. The first mover might have set his or her effort strategically, to influence the second mover's effort. Our design avoids these problems.

are told that they can, but do not have to, revise their effort. Both agents simultaneously choose a revised effort $\hat{e}_i \in \{1, 2, \ldots, 20\}$. To make the revision decision incentive compatible, agents are told that only for *one* randomly selected agent the *revised* effort is relevant for calculating earnings, while for the *other* agent the *initial* effort remains relevant for the payoffs. A random device generates $r \in \{0, 1\}$ with equal probability. In case $r = 1$ agent 1's revised effort and agent 2's initial effort are payoff relevant (that is, agent 2's revised effort has no effect on any of the earnings). In case of $r = 0$, agent 2's revised effort and agent 1's initial effort are payoff relevant (and agent 1's revised effort has no effect on any of the earnings). Thus, our design ensures that (1) only one agent's revised effort matters, and therefore the other agent's initial effort is unaltered, and (2) it allows us collecting data from both agents. Subjects know this procedure (but not yet the outcome of the random draw) when choosing the revised effort. The expected earnings of the principal are

$$x_P = vE_r[r(\hat{e}_1 + e_2) + (1 - r)(e_1 + \hat{e}_2)] - 2w, \tag{1}$$

where $v > 0$ is the constant marginal product of the agents' efforts. The expected earnings of the two agents are calculated as

$$x_1 = w - E_r[rc(\hat{e}_1) - (1 - r)c(e_1)] \quad \text{and} \quad x_2 = w - E_r[(1 - r)c(\hat{e}_2) - rc(e_2)], \tag{2}$$

where the cost of effort is equal to $c(e_i) = 7(e_i - 1)$ for both agents.[3] Note that we impose uniform wages and identical marginal productivity (v), because we want to observe the two agents in an identical situation.

The revised effort is our main instrument to identify social interaction effects. The only change between the initial effort decision and the revision stage is the additional information about the co-agent's effort. We use $\Delta e_i = \hat{e}_i - e_i$ as a measure for the reaction to effort information, that is, as an indication for a peer effect.

It is important to note that we measure peer effects in a situation where the co-agent's effort remains *unchanged*. This design feature avoids the reflection problem. When choosing the revised effort, agent *i* knows that either his decision has no effect ($r = 0$) *or* the effort of the co-agent remains unchanged ($r = 1$). The random selection of either the initial effort or the revised effort ensures that the co-agent's effort is exogenous and has the added advantage that it allows us to collect revision decisions from all agents.

We cannot rule out the possibility, however, that subjects might want to change their effort decision in the *revision stage* for reasons unrelated to peer effects, such as 'experimenter demand effect' (e.g., Zizzo, 2010), 'virtual learning' (Weber, 2003), or change of mind or even non-stable social preferences (Volk, Thöni, & Ruigrok, 2012). Thus, in order to isolate peer effects from other sources of effort revisions we need a control treatment in addition to the 'Effort Information treatment' (EIT). Our control treatment, called the 'No Information treatment' (NIT), is identical to the game explained above except when reaching the revision stage subjects are *not* informed about the effort choice of the co-agent.

3.2. Further design features and procedural details

In some of the sessions we elicit beliefs about the co-agent's initial effort choice. This allows us to observe two points of an agent's reaction function (Section 6). Subjects enter their belief in the same screen as they choose their initial effort. Subjects earn additional 100 ECU for a correct belief, nothing otherwise. Because eliciting beliefs might influence effort (e.g., Gächter & Renner, 2010) we include the belief elicitation only in some of the sessions.

To check for the robustness of our results we change several contextual parameters across sessions. First, we vary the level of the agents' productivity ($v = 18$ or $v = 35$ for both agents) and therefore the gains from cooperation. Second, most of our participants (14 sessions, 357 participants) played a one-shot, three-person gift-exchange game *prior* to the experiment we report in this paper. In this Experiment 1 agents made their effort decision in the strategy method. Subjects did not receive any information about other subjects' decisions prior to the experiment presented in this paper. Another group of subjects (3 sessions, 108 participants) played Experiment 1′, consisting of eight rounds of a three-person gift-exchange game with random matching (in matching groups of 12 subjects). These subjects have more experience with the game prior to the start of the experiment at hand. During the eight rounds agents received information about their principal's wage offerings but agents did not receive any direct information about their co-agent's effort choices. We label these subjects as *Experienced* and use this contextual variation to check whether increased experience with the game influences peer effects. See Gächter and Thöni (2010) for the results of Experiment 1 and 1′. We use the data from Experiment 1 and 1′ to classify our subjects into *Selfish* and *Non-selfish* types. This provides us with a measure for other-regarding preferences that is not derived from the decisions in the experiments reported here.

We conducted the experiment at the Universities of St. Gallen and Zurich in computerized laboratories where subjects were separated by partitions and thus made their decisions in isolation and without communication. All decisions were anonymous. We used the software z-Tree (Fischbacher, 2007) to run our experiments and ORSEE (Greiner, 2004) for recruiting the subjects. We framed the experiment in a 'buyer–seller' terminology because we deem it to be more neutral than a labor relations frame.

Our research question requires a one-shot experiment. We therefore took great care to ensure that subjects understand the rules, as well as the pecuniary payoff consequences of their decisions. Subjects had to answer a set of control questions

[3] To rule out overall losses all players were endowed with 400 ECU (experimental currency unit).

on payoff consequences. To help subjects calculate earnings for all players and all possible combinations of efforts and wages, the software provided a 'What-if calculator' (see Fig. B1 in Appendix B for an illustration of a decision screen).

We have observations from 18 sessions with a total of 489 participants, 326 agents and 163 principals. The majority (330) decided in the EIT. The remaining 159 subjects decided in the NIT. We imposed no time limit for decisions. The experiment lasted about 30 min and the average earnings were CHF 13.8 (€ 8.8). See Table C1 (Appendix C) for further details.

4. Results I: Existence and direction of peer effects

4.1. Initial effort choices

Recall that the EIT and the NIT are identical up to the *revision stage*. For analyzing initial effort choices we therefore pool the data. As expected from numerous gift-exchange experiments (surveyed in, e.g., Charness & Kuhn, 2011), efforts increase in wages.[4] Regression analyses confirm this observation (see Appendix D). These analyses also show that subjects who we classify as *Selfish* (always chose minimal effort; 28.1%) and *Non-selfish* (chose at least one non-minimal effort; 71.9%) based on their behavior in Experiment 1 and 1' make different initial effort choices in this experiment. Later we use this classification to look at the subgroup of subjects who were sufficiently reciprocal toward the principal to choose non-minimal efforts. With regard to initial effort we find that *non-selfish* subjects are significantly more likely to choose a non-minimal initial effort; they also choose higher initial effort levels than subjects classified as *Selfish*.

4.2. Existence of peer effects in voluntary cooperation

The left panel of Fig. 1 shows the frequency of effort revisions in the two treatments. In EIT agents revise their effort in 73 out of 220 of the cases (33.2%). Effort revisions also occur in the NIT: 24 out of 106 agents (22.6%) revise their effort. The differences between EIT than in NIT are significant at $p = .050$.[5]

Peer effects presumably matter most among agents who care about others' well-being at all. Agents with no other-regarding preferences will always choose minimum effort. Among the *Non*-selfish subjects 62 out of 141 (44.0%) revise effort in the EIT while 21 out of 82 (25.6%) do so in the NIT ($p = .012$).[6] Even more frequent are revisions among the subjects who chose a non-minimal initial effort. Sixty-eight percent of agents in EIT revise their effort. In NIT the corresponding number is 45% ($p = .009$).

Because observations within a triad are not independent, we treat a triad as an independent cluster of observation. Table 1 reports the results of Probit estimations (coefficients, standard errors, and marginal effects). 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 EIT increases the probability of an effort revision significantly. The marginal effect is a 14.2 percentage point increase of the probability in the EIT compared to the NIT. We introduce the initial effort by two variables in order to allow for changes in the behavior of agents with minimal and non-minimal effort. *Initial effort* is the effort chosen (e_i) and *Minimal initial effort* is a dummy for $e_i = 1$. Higher initial efforts increase the probability of an effort revision, whereas having chosen a minimal initial effort decreases the likelihood to revise (by 41.8 percentage points). None of the other parameters matters significantly.

In Model 2 we repeat the estimation for the restricted sample of agents classified as non-selfish. The marginal effect of effort information on revision increases to 21.5%. None of the other contextual variables has a significant effect on the probability of effort revisions.

The magnitude of effort revisions ($|\Delta e_i|$, right panel of Fig. 1) is considerably larger in EIT (.97 on average) than in NIT (.37). This effect is not only driven by the fact that agents revise effort more frequently when information about their co-agent's effort is provided. In the subsample of agents who actually do revise effort ($\Delta e_i \neq 0$) the difference between the average absolute effort revisions increases to 1.31 effort units. If we repeat the estimates of Table 1 but apply a Tobit regression with the absolute effort revision as dependent variable we get very similar results. We summarize these findings as follows:

Result 1: *We find evidence for peer effects in voluntary cooperation: Information about the other agent's effort causes significantly more and substantially larger effort revisions compared to the No Information treatment.*

4.3. Direction of peer effects

We first investigate whether effort information has a systematic effect on revised efforts. We apply a Tobit estimate for the revised effort, dependent on the observed other agent's effort and controlling for own effort. Table 2 reports the results of these estimates (EIT only).

[4] The average effort chosen at the lowest wage is 1.53; the intermediate wage triggered an average effort of 2.97 and the highest wage an average effort of 5.53. Minimal efforts occurred in 68.4%, 49.0% and 37.5% of the cases in which principals paid the low, intermediate and high wage, respectively. Among the 163 principals in our sample 46.6% paid the lowest possible wage of 50. Another 31.3% paid the intermediate wage of 100 and the remaining 22.1% offered the highest wage of 200.

[5] *p*-values are from χ^2 tests with correction for dependence within matching group (Rao & Scott, 1984).

[6] Among the subjects classified as *Selfish* 11% of the subjects revise their effort, while 37% of the *Non-selfish* subjects revise their effort ($p < .000$).

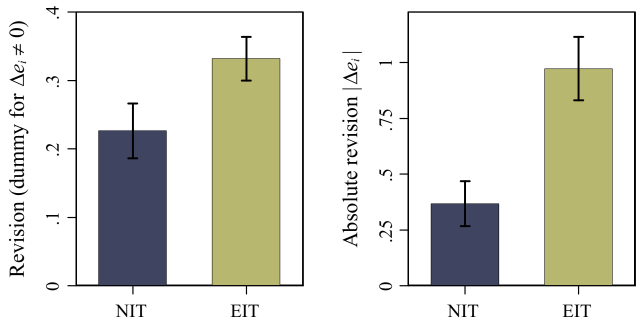

Fig. 1. Frequency (left panel) and magnitude of absolute effort revisions (right panel) in NIT and EIT, with clustered standard errors.

Table 1
Probit estimations for the decision to revise effort.

	Dependent variable: Revise (dummy for $\Delta\hat{e}_i \neq 0$)					
	Model 1 (all agents)			Model 2 (non-selfish agents only)		
	Coef	SE	ME	Coef	SE	ME
EIT (D)	0.522***	0.193	0.142	0.658***	0.224	0.215
Initial effort	0.136***	0.039	0.040	0.133***	0.043	0.046
Minimal initial effort (D)	−1.398***	0.257	−0.418	−1.441***	0.314	−0.446
Experienced (D)	−0.297	0.211	−0.082	−0.194	0.256	−0.065
Belief (D)	0.242	0.216	0.071	0.080	0.257	0.028
High productivity (D)	0.236	0.374	0.064	0.007	0.441	0.003
Zurich (D)	0.048	0.329	0.014	0.227	0.369	0.076
Constant	−1.041***	0.362		−0.957**	0.436	
N	326			223		
Log-likelihood	−124.786			−94.851		
$p > \chi^2$	0.000			0.000		

Notes: The NIT is the omitted benchmark. Apart from *Initial effort* all independent variables are dummies (D). We report coefficients, standard errors (SE), and marginal effects (ME). Model 1 uses all agents, and Model 2 only agents classified as *Non-selfish*. We apply a robust estimation of the SE clustered within a matching group.

* $p < 0.10$.
** $p < 0.05$.
*** $p < 0.01$.

Table 2
Tobit estimations for revised effort.

	Dependent variable: Revised effort ($\hat{e}_i$)			
	Model 1 (all agents)		Model 2 (non-selfish agents only)	
	Coef	SE	Coef	SE
Co-agent's initial effort (e_j)	0.292***	0.081	0.411***	0.115
Initial effort	0.489***	0.092	0.373***	0.109
Minimal initial effort (D)	−4.542***	0.737	−3.634***	0.605
Experienced (D)	0.146	0.882	1.020	0.809
Belief (D)	−0.377	0.922	−0.838	0.847
High productivity (D)	−1.230	1.379	−1.707	1.134
Zurich (D)	1.323	1.149	1.065	0.930
Constant	0.236	0.909	0.923	0.848
Σ	2.910		2.221	
N	220		141	
Log-likelihood	−248.8		−180.3	
$p > F$	0.000		0.000	

Notes: Except for the *Co-agent's initial effort* and the *Initial effort* all independent variables are dummies (D). Data from EIT only. Model 1 uses all agents; Model 2 uses only agents classified as *Non-selfish*. Robust standard errors clustered within a triad.

* $p < .10$.
** $p < .05$.
*** $p < .01$.

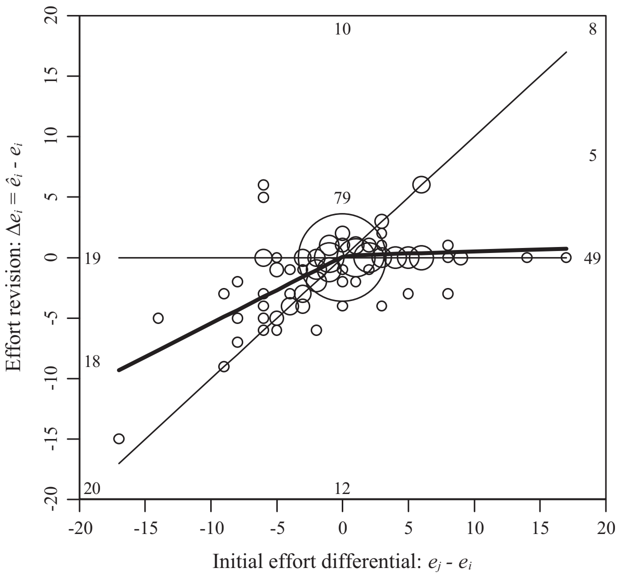

Fig. 2. Scatter plot of effort revisions dependent on the difference between the agents' initial efforts. The thin lines show the limit cases of no effort revisions (horizontal line) and 'perfect' effort revisions (45-degree line). The bold line depicts a trend line. The total number of observations is 220. Numbers next to lines indicate number of cases on the respective line, and numbers between lines indicate number of cases between lines.

Model 1 shows that the co-agent's effort significantly influences the revised effort. The effect is positive, that is, high co-agent's efforts *ceteris paribus* increase the agent's effort and vice versa. In Model 2 we restrict our sample to the *Non-selfish* types. The estimate for the subgroup is qualitatively similar to the estimate with the whole sample.

The strong and positive influence of the observed co-agent's effort on the revised effort suggests that, on average, efforts are complements. In a next step we take a closer look at how the observed *difference* between *j*'s effort and *i*'s effort influences *i*'s revision decision.

Fig. 2 provides a scatter plot of the differences in initial efforts e_i and e_j and the effort revision in EIT. The size of dots is proportional to the number of underlying observations. Observations on the thin horizontal line stem from agents who left their effort unchanged. The second thin line is the 45-degree line. Observations on this line mean that an agent matched the co-agent's effort exactly. The numbers in the scatter plot indicate the number of observations within a region. Numbers at the end of the thin lines count the observations on the line for negative or positive effort differences, respectively. Numbers in areas between lines count observations within the regions between the thin lines.

For *negative initial effort differentials* ($e_j < e_i$), 20 effort revisions are on the diagonal, and 19 are on the zero-revision line. Eighteen observations are between the zero-revision line and the diagonal. These agents revise their effort toward the other agent's effort but do not match it. The number in the middle of the graph (79) indicates the number of observations with *no initial effort difference* and no effort revision. Ninety percent of these observations are from agents choosing minimal initial effort. In case of *positive effort differentials* ($e_j > e_i$) agents either match the other agent's effort (in 8 cases), adjust toward the other agent's effort but not fully (in 5 cases), or, in most cases (49), do not revise their effort.

The observations in Fig. 2 suggest asymmetric reactions to positive and negative effort differentials. When fitting the data with a regression line we therefore allow for different slopes and different intercepts. This trend line (the bold line in Fig. 2) shows a kink at $e_j - e_i = 0$, which suggests that on average people only react to their co-agent's effort if the co-agent chooses a lower effort than them.

Table 3 shows regressions of effort revision on the initial effort differential ($e_j - e_i$), the initial effort e_i, and the contextual parameters. Model 1 disregards any kink in the revision response. The effort differential has a positive and highly significant impact on effort revisions. An increase of the effort differential by one unit induces an agent to increase his effort in the revision stage by .18 units, *ceteris paribus*.

However, as Fig. 2 suggests, there are substantial differences between positive and negative effort differentials. In Model 2 we allow for different slopes by adding two additional variables for the initial effort differential. The variable *Initial effort difference if* > 0 is calculated as max[e_j - e_i, 0].

The results of Model 2 confirm the impression from Fig. 2. The coefficient of *Initial effort difference* is significant and positive. Agents who learn that their co-agent had chosen a lower effort reduce their effort on average by .38 effort units per unit of the differential. The interaction variable *Initial effort difference if* > 0 has a significant negative coefficient, indicating that the reaction to the effort differential is lower in the positive domain. The net effect in the domain of positive effort differentials is the sum of the first and second coefficient. The effect is positive (.04, the sum of the first two coefficients)

Table 3
Effort revisions as a function of effort differences.

	Dependent variable: Δei					
	Model 1		Model 2	Model 3		
	Coef	SE	Coef	SE	Coef	SE
Initial effort difference	0.177***	0.046	0.379***	0.115	0.372***	0.118
Initial effort difference if > 0			−0.337**	0.134	−0.329**	0.138
Initial effort difference > 0 (D)			0.181	0.196	0.212	0.278
Experienced × Initial effort difference					0.184	0.294
Experienced × Initial effort diff. if > 0					−0.340	0.277
Experienced × Initial effort diff. > 0 (D)					0.036	0.342
Initial effort	−0.319***	0.074	−0.222***	0.044	−0.228***	0.045
Minimal initial effort (D)	−0.550	0.349	−0.542*	0.319	−0.584*	0.343
Experienced (D)	0.104	0.327	−0.106	0.302	0.021	0.289
Belief (D)	−0.045	0.407	0.065	0.396	0.068	0.399
High productivity (D)	−0.876	0.661	−0.634	0.548	−0.646	0.556
Zurich (D)	0.951	0.598	0.611	0.461	0.617	0.469
Constant	0.768	0.522	0.852*	0.482	0.879*	0.498
N	220		220		220	
Log-likelihood	−418.1		−411.6		−411.3	
p > F	0.000		0.000		0.000	
r^2	0.419		0.453		0.454	

Notes: OLS regression of the effort revision Δe_i dependent on the difference in the initial efforts (*Effort difference* e_j - e_i). Model 2 allows for different slopes in the positive and negative domain by including the effort difference in the positive range, i.e., $\max[e_j - e_i, 0]$ and a dummy for positive effort differences. In Model 3 we add interaction variables between *Experienced* and the measures for effort differences. Robust standard errors in parentheses, two agents in a group are clustered.

* $p < 0.10$.
** $p < 0.05$.
*** $p < 0.01$.

but not significantly different from zero ($p = .436$, F-test). Thus, the interaction between the two efforts is mainly driven by effort reductions of the high-effort agents.

Among the remaining variables only *Initial effort* has a significant impact on the effort revision; the coefficient is negative. Thus, unsurprisingly, the higher the initial effort the larger is the downward revision.

Model 3 allows for the possibility that agents who are experienced with the gift-exchange game react differently to effort information than inexperienced agents. The interaction variables are insignificant. Thus, the observed peer effects are robust to experience.

We summarize our findings in Result 2.

Result 2: *Overall, effort revisions and differences in initial efforts are positively correlated. Agents who learn that their co-agent has provided less effort than them reduce their effort significantly, whereas we do not find systematic effects for agents who chose a lower initial effort than their co-agent.*

4.4. Peer effects or learning?

The fact that people tend to revise their efforts downwards when facing an agent with a lower effort can be explained by peer effects, but might as well be an indication of the importance of learning. For two reasons we think that our results cannot be explained by learning effects. First, Model 3 of Table 3 shows that experienced subjects do not exhibit weaker reactions to effort information. Second, recall that subjects had access to a 'What-if Calculator'. Our software recorded subjects' calculations. All but 11 of our 489 subjects calculated the payoffs for the Nash equilibrium efforts and therefore should not have been surprised by the fact that a co-agent with a lower effort earns a higher payoff. Thus, Results 1 and 2 are most likely not due to learning about the money maximizing strategy.

5. What standard models of social preferences predict about peer effects in the trilateral gift-exchange game

In this section we explore what existing theories of social preferences have to say on explaining peer effects, given that all theories we look at in this section can explain initial effort choices. We focus our analysis on the subgame starting when the two agents choose their effort. We derive agent i's reaction function to agent j's effort decision, that is, $e_i = R(e_j)$ and focus on the derivative with respect to e_j. A particular model predicts a peer effect if $de_i/de \neq 0$; no peer effect is predicted if $de_i/de = 0$. Because (i) role allocation was random (ii) we explained the game to the subjects as a three-player game and made them aware of the earnings consequences of each other's choices (see instructions and Fig. B1), and (iii) subjects were not informed about any decision other three-player groups took, we assume that a group of three players forms the reference group.

In the following we derive the basic results and briefly discuss the underlying intuitions. For all details see Appendix E. Readers not interested in the details can directly refer to the summary Table 4 at the end of this section.

Table 4
Summary of theoretical predictions with regard to de_i/de_j.

Class		Model	Slope of $R(e_j)$ for interior solutions ($1 < e_i < 20$) Numerical range ($v = 35$)
Money maximizing			0 (no interior solutions)
Distributional preferences	Altruism	Cox et al. (2007)	(−.63, −.98)
		Charness and Rabin (2002)	
		– Maximin	−.83 ∪ −1.2
		– Intermediate	−.83 ∪ 1
		– Utilitarian	0 (no interior solutions)
	Inequity aversion	Bolton and Ockenfels (2000)	(−.57, −.70)
		Fehr and Schmidt (1999)	
		– Aheadness averse (*AA*)	−.83
		– Behindness averse (*BA*)	−.83 ∪ 1
(Type based) reciprocity		Dufwenberg and Kirchsteiger (2004)	0
		Levine (1998)	0
Hybrid models		Cox et al. (2007)	No additional slopes to altruism prediction
		Charness and Rabin (2002)	No additional slopes to altruism prediction
		Falk and Fischbacher (2006)	−.83

Notes: Predictions for the slope of agent *i*'s reaction function to agent *j*'s effort, $e_i = R(e_j)$. For the piecewise linear models (Charness & Rabin; Fehr & Schmidt) we can calculate the slopes directly from the formula for $v = 35$. For models predicting a continuum of reaction functions we derive the range of possible slopes. For Cox et al. (2007) we report the range of slopes of best replies that lead to interior solutions. In case of Bolton and Ockenfels (2000) there is no closed-form solution for the best replies. The numbers reported are from numerical calculations using parameterized utility function $u = x_i - b(\sigma_i - 1/n)2$. Details are in Online Appendix E.

5.1. Distributional preferences

We consider models of altruism and inequity aversion. Players have a utility function $u_i(x_i, x_j, x_P)$ defined on the monetary earnings x_i, x_j and x_P of the two agents *i* and *j* and the principal *P*, respectively (see Eqs. (1) and (2) in Section 3.1). The models differ in the assumptions about the derivatives of u_i with respect to other players' earnings. Models of altruism like Charness and Rabin (2002) or Cox, Friedman, and Sadiraj (2008) assume that these derivatives are positive. Models of inequity aversion (Bolton & Ockenfels, 2000; Fehr & Schmidt, 1999) assume that these derivatives are positive as long as other players are poorer than player *i* but negative otherwise. We now discuss these models in turn.

5.2. Models of altruism

Assume agent *i* maximizes a utility function $u_i(x_i, x_j, x_P)$. The other agent's payoff x_j is independent of *i*'s actions. Agent *i* chooses e_i to set x_i and x_P at the level which maximizes her utility given x_j. The left panel in Fig. 3 illustrates the utility maximization problem of agent *i* in the (x_i, x_P) space. The lower thick line represents the choice set for agent *i* in the example case where agent *j* chooses $e_j = 4$ If agent *i* chooses maximum effort the principal's earnings are highest and *i*'s earnings are lowest. The slope of the graph representing the choice set is $-v/7$, the marginal benefit of effort divided by the marginal cost of effort. The thin lines show two indifference curves of agent *i*. The slope of the indifference curves indicates the marginal rate of substitution between agent *i*'s earnings and the principal's earnings. If agent *i* cares sufficiently for the earnings of the principal (as reflected in agent *i*'s indifference curves) then the optimal effort choice e_i^* is non-minimal, as depicted in the left panel of Fig. 3.

How does an altruistic agent react to changes in e_j? An increase in the other agent's effort, for instance, shifts the choice set in Fig. 3 upwards because it increases the principal's income. The location of the new optimum depends on the 'income elasticity' of the demand for x_i, the own income. If the own income is a 'normal good', the new optimum lies northeast of the old optimum as depicted in the left panel of Fig. 3. In this case agent *i* reduces her effort whenever agent *j* increases his effort. In the following we focus on agent *i*'s reaction function to *j*'s effort. The right panel of Fig. 3 depicts the corresponding reaction function. If x_i is a normal good for agent *i*, then the slope of the reaction function is negative for interior solutions, that is, the two efforts are strategic substitutes ($de_i/de_j < 0$).[7]

For expositional purposes we derive the reaction functions using a parameterized version proposed by Cox et al. (2007).[8] They use a CES utility function

[7] Assuming that x_i is a normal good is reasonable for most people and situations. Otherwise, a subject facing a windfall gain should be ready to transfer more than the gain to another subject. In a trust game, for instance, a subject should return more money than received by the trustor, a behavior which is hardly ever observed.

[8] The model builds on Cox and Sadiraj (2012) who introduced (in the working paper version of 2003) the CES function as shown in Eq. (3) and call their approach a model of egocentric altruism.

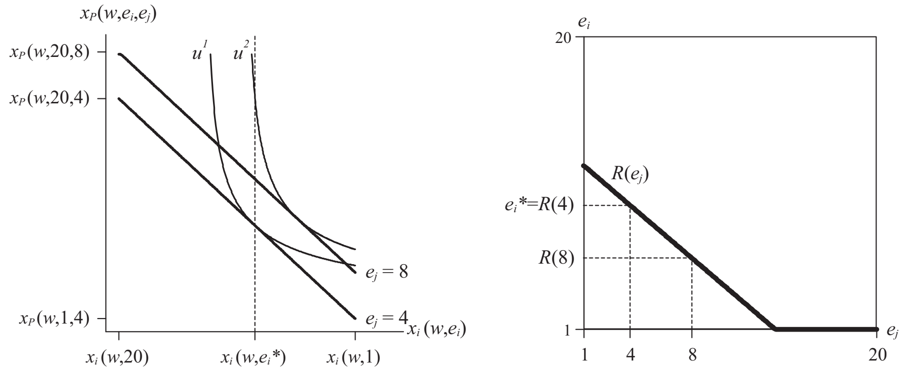

Fig. 3. Peer effects in a model of altruism. Left panel: Utility maximization in the (x_p, x_i) space for an altruistic agent i. Thick lines represent two choice sets for two levels of the other agent's effort; u^1 and u^2 indicate indifference curves of an altruistic agent i; e_i^* denotes the optimal effort of agent i in case agent j chooses an effort of 5 and the lower thick line is i's choice set. Right panel: Corresponding reaction function of agent i to agent j's effort.

$$u(x_i, x_j, x_P) = (x_i^\alpha + \theta_j x_j^\alpha + \theta_P x_P^\alpha)/\alpha, \tag{3}$$

which allows varying the elasticity of substitution between an agent's own payoff and the other players' payoffs by $\alpha \in (-\infty, 0) \cup (0, 1]$; θ_j and θ_P measure the emotional state of player i toward the other two players. Suppose for a moment that θ_j and θ_P are positive. For $\alpha = 1$ the payoffs are perfect substitutes and agent i chooses either maximal effort (for $\theta_P > 7/v$) or minimal effort (otherwise), irrespective of e_j. In this case the slope of the reaction function is zero and there is no interior solution. Panel A of Fig. 4 shows these reaction functions as horizontal lines at the bottom and the top of i's action space. Another extreme case is when the payoffs are perfect complements and weighted equally (Leontief case $\alpha \to -\infty$ and $\theta_P = 1$). In this case agent i chooses e_i to ensure $x_i = x_P$ (if feasible). The reaction function can be derived by solving this equation for e_i which gives us a linear function with a slope of $-v/(v + 7)$.

Between these extremes is a continuum of negatively-sloped reaction functions. The lines in Fig. 4A show some examples. All reaction functions are linear functions that intersect at a point far to the upper left of the admissible effort space. Both the slope and the intercept of the reaction function are jointly determined by α and θ_P. Optimal efforts of agent i might lead to situations where the principal earns more than agent i, which is the case above the thick line in Fig. 4A. In all cases the slope lies in $(-1, 0)$ for interior solutions.

Another model of altruism is Charness and Rabin (2002) who – in the basic version – propose a utility function that captures preferences for efficiency (utilitarian) and/or care for the least fortunate (maximin). Utility is case-wise linear in all arguments:

$$u(x_i, x_j, x_P) = (1 - \lambda)x_i + \lambda[\delta \min\{x_i, x_j, x_P\} + (1 - \delta)(x_i + x_j + x_P)], \tag{4}$$

with λ weighing the importance of distributional preferences ($\lambda \in [0, 1]$) and δ measuring the type of distributional preferences, ranging from $\delta = 0$ for pure efficiency concerns to $\delta = 1$ for pure maximin concerns. Unlike the Cox et al. (2007) model, Charness and Rabin do not predict a continuum but only four distinct kinds of reaction functions: For a low enough λ an agent always chooses minimal effort. For high λ and low δ an agent seeks to maximize joint income and chooses maximum effort (thereby minimizing her own income). This reaction function is labeled as 'Utilitarian' in Fig. 4B. The most interesting cases are the intermediate ones in between the utilitarian and maximin cases. If the maximin motive dominates (high δ), agent i increases her effort if and only if she can increase the minimal earnings in the group.

In Fig. 4A we already introduced the locus where agent i earns the same as the principal ($x_i = x_P$) as a downward-sloping linear function. In panel B we add two loci: a steeper downward-sloping function indicating where agent j earns the same as the principal ($x_j = x_P$) and the 45-degree line indicating where the two agents earn the same. The intersection of the three lines is the situation in which all three earnings are equal, which is the case when both agents choose $\bar{e} = (3w + 7)/(2v + 7)$.

In case of $e_j < \bar{e}$ a maximin agent chooses her effort along $x_i = x_p$, to prevent the principal from being the uniquely poorest. In case of $e_j > \bar{e}$ agent i is always richer than agent j. Agent i cannot influence x_j; however, i's choice determines whether agent j or the principal is poorest. Agent i then chooses her effort such that the principal does not earn less than agent j, that is, the reaction function follows $x_j = x_p$. Finally, for intermediate values of δ there is a type with a v-shaped best reply. She promotes efficiency under the restriction that her own earnings do not become the unique minimum. For $e_j < \bar{e}$ this agent acts as a maximin type; for $e_j > \bar{e}$ she matches agent j's effort and, hence, the reaction function follows the 45-degree line.

5.3. Models of inequity aversion

Theories of inequity aversion assume that agents dislike unequal payoff distributions, *ceteris paribus*. Consider the model of inequity aversion by Bolton and Ockenfels (2000). Utility is $u(x_i, \sigma_i)$ where σ_i is i's *share of total earnings*. For a given x_i

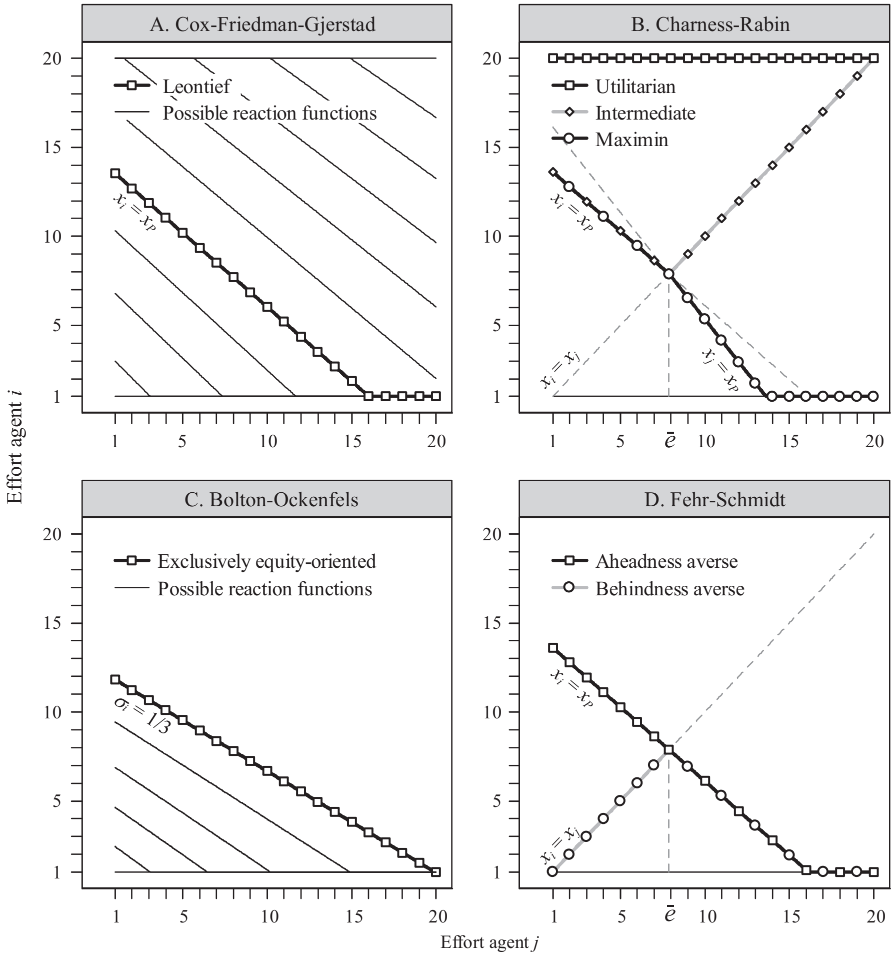

Fig. 4. Illustrations of peer effects (Reaction functions $ei = R(ej)$) predicted by theories of distributional preferences. *Note:* The reaction functions are drawn for $w = 200$ and $v = 35$.

players are assumed to prefer their share to be equal to one third. Deviations from equality reduce utility.[9] To get an intuition consider first the case of a very strongly inequity averse player, who only cares about her payoff share. In the role of agent *i*, this player chooses her effort such that her share of total earnings equals one third:

$$\frac{x_i(w, e_i)}{x_i(w, e_i) + x_j(w, e_j) + x_P(w, e_i, e_j)} = \frac{1}{3}. \quad (5)$$

For such a player the two efforts are strategic substitutes. To see this, consider an *increase* of player *j*'s effort. This decreases x_j and increases x_P. Since providing more effort is efficient the sum of x_j and x_P increases by $v - 7 > 0$ and the left-hand expression in (5) drops below 1/3. To re-establish equality agent *i* must *decrease* her effort in order to increase x_i. The reaction function of such an agent is depicted in Fig. 4C, labeled as 'Exclusively equity-oriented'. Using the payoff functions (1) and (2) and solving (5) for e_i one can show that the slope of the reaction function is $(7 - v)/(14 + v) < 0$ (note that this is not the same reaction function as the limit case in panel A where $x_i = x_p$). Players with weaker inequity aversion face a tradeoff 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.*

The thin lines in Fig. 4C show five examples of reaction functions. There is a lower limit of inequity aversion under which behavior is identical to money-maximization. However, it generally holds that, for interior solutions ($1 < e_i < 20$), the slope of the reaction function is always in $(-1, 0)$, that is, the two efforts are always strategic substitutes.

The intuition is that an inequity-averse agent providing low effort suffers from earning more than the equal share. To relieve this adverse feeling there are two possibilities: (i) she increases her own effort and thereby lowers her earnings,

[9] An early strand of literature incorporated envy into the utility function (Bolton, 1991; Kirchsteiger, 1994). These models cannot explain non-minimal efforts in our game, because players seek to maximize their absolute and relative monetary payoff.

or (ii) the co-agent increases his effort and thereby increases the total payoff, which in turn brings the (unchanged) income of the agent 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) players make bilateral comparisons with all group members. Players get utility from their own monetary payoff and disutility from any payoff difference with comparison partners (see also Loewenstein, Thompson, and Bazerman (1989)). The utility function in the Fehr–Schmidt model is

$$u(x) = x_i - \frac{\alpha}{2}\left([x_j - x_i]^+ + [x_P - x_i]^+\right) - \frac{\beta}{2}\left([x_i - x_j]^+ + [x_i - x_P]^+\right), \quad (6)$$

where $[a]^+ \equiv \max(a,0)$. The disutility of earning less than another group member is linear and equal to α times the payoff difference. Earning more than another group member also leads to a disutility, weighted by β (but $\beta < \alpha$). We illustrate the reaction functions of Fehr–Schmidt agents in Fig. 4D. Two loci are important: the negatively-sloped line where agent i earns the same as the principal and the 45-degree line where the two agents earn the same. In the intersection of the two lines all three players earn the same, which is the case at $e = \bar{e}$.

The Fehr–Schmidt model predicts three types: A player with low concern for advantageous inequality ($\beta < \beta' = 14/(v + 14) \approx 0.29$ for $v = 35$) always choose minimal effort. For higher β there are two possibilities depending on the relative importance of α and β: If a player is relatively intolerant toward disadvantageous inequality compared to his intolerance of advantageous inequality, we call her 'Behindness averse' (*BA*; $\beta' < \beta < (14 + 7\alpha)/(v + 7)$). Such an agent chooses non-minimal efforts under the condition that she does not fall behind another player. For low co-agent's efforts ($e_j < \bar{e}$) the best reply is $e_i = e_j$ up to $\bar{e}$. For high co-agent's efforts ($e_j \geqslant \bar{e}$) she chooses the effort that equalizes his earnings to the principal's earnings. A third type called 'Aheadness averse' (*AA*; $\beta' < \beta > (14 + 7\alpha)/(v + 7)$) is an agent i who (i) suffers heavily from the fact that the principal earns less than her (high β), and (ii) is relatively tolerant to the fact that she earns less than agent j (low α). Such an agent always seeks to match her payoff with the principal's payoff. The resulting reaction function is identical to the Leontief case shown in panel A of Fig. 4.

5.4. Models of reciprocity

Theories of reciprocity model the idea that people reward kind acts with kindness and mean acts with unkindness (Rabin, 1993). A theory of reciprocity that is adequate for our sequential gift-exchange game is the sequential reciprocity model by Dufwenberg and Kirchsteiger (2004). This theory does not predict peer effects because agent j's effort has no influence on agent i's earnings. Thus agent j is neither kind nor unkind to agent i. Hence, $de_i/de_j = 0$. The only reason for choosing a non-minimal effort is to reward the principal for a high wage, irrespective of the other agent's actions.

In case of *type-based reciprocity* (Levine, 1998) the results are similar. In this model players gain (dis)utility from other agents' income if they are altruistic (spiteful) types. However, since the agents cannot influence their co-agent's income they cannot act altruistically (or spitefully) toward them and thus, do not take their actions into account. Hence, no peer effects are predicted: $de_i/de_j = 0$.

5.5. Hybrid models

Cox et al. (2007) and Charness and Rabin (2002) enrich their models of altruism with reciprocity. In both cases reciprocity does, however, not change the qualitative predictions about the shape of the reaction functions discussed so far. Reciprocity means that if the agent is treated unkindly she weighs the earnings of the unkind player less or even negatively in her utility function. In both models intentions play a role only with respect to the wage offer. Low wage offers are perceived as unkind, high wages as kind. In case of Cox et al. (2007) a low wage leads to a negative θ_P. A player with $\theta_P < 0$ chooses minimal effort irrespective of e_j, thus acting like a money-maximizing agent. Also in Charness and Rabin (2002) there is a reciprocity part by which payoff-based concerns are reduced when a player is treated unkindly by another player. Negative emotions toward the principal shift the reaction functions downwards but do not qualitatively change the characteristics derived above.

Finally, the model of Falk and Fischbacher (2006) combines interpersonal payoff comparisons with intentionality. Like in Dufwenberg and Kirchsteiger (2004), reciprocity does not predict a direct link between the two efforts. However, agent i wants to reciprocate to the principal *and* cares about earnings differences. The predictions of the Falk-Fischbacher model are very similar to the predictions of the *AA*-type in the Fehr–Schmidt model. For very strong reciprocal preferences the reaction function is again identical to the *AA*-type, weaker reciprocal preferences result in a parallel downward shift.

5.6. Summary

Table 4 summarizes the models and their predicted peer effects (the predicted slope of the reaction function(s)). The rightmost column of Table 4 provides numerical boundaries for the slope of the reaction function(s) of the respective model.

With two exceptions all models of social preferences predict either that efforts are unrelated (no peer effects) or that peer effects take the form of efforts being strategic substitutes. The intuition for the latter is simple: with distributional preferences agents will choose non-minimal efforts either because they (i) enjoy the principal's earnings (altruism), or (ii) they seek equitable outcomes (inequity aversion). A co-agent who puts in high effort reduces i's need to put in high effort. In none of the models agent i cares about whether the increase in the principal's earnings is caused by her own or some other player's actions. There are two notable exceptions that allow for strategic complementarity between the two efforts, the Fehr–Schmidt *BA*-type, and the Charness–Rabin intermediate type. In both cases efforts have to be one-to-one complements, that is, the two agents choose identical efforts.

Our analysis shows that the most robust prediction about peer effects is that the agents' efforts are strategic substitutes. By contrast, Fig. 2 *suggests* that efforts are strategic complements. This observation of positively correlated efforts is not yet conclusive, however, because the predictions concern the slope of the reaction function. In the following we report experiments that provide qualitative conclusions about the sign of peer effects.

6. Results II: Can standard models of social preferences explain peer effects?

In order to measure the slope of the reaction function we make use of agent i's *belief* about agent j's initial effort. In a subset of our data ($n = 110$) we elicited agents' beliefs about their co-agent's initial effort decision. Given the belief we can observe two points on an agent's reaction function in case the belief was wrong. This provides a direct measure of the sign of the slope of a monotonic reaction function by estimating the function $\Delta e_i = f(e_j - e_j')$, where e_j' denotes agent i's belief about j's initial effort. We call the difference between the true co-agent's effort and the belief 'surprise'.

We use OLS to estimate the average $\Delta e_i = \alpha + \beta(e_j - e_j') + \varepsilon$. According to the theories discussed above, the difference between the belief and the actual effort of the co-agent is the only reason to revise 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 $v = 35$. As said, a robust prediction is that efforts are strategic substitutes and the slope of the reaction function is between $-.98$ and $-.57$ (see Table 4).

In contrast, the estimation results show that the slope coefficient is positive and significant ($\beta = .261$, $p = .003$). The constant is insignificant ($\alpha = -.230$, $p = .216$). Fig. 5 shows a scatter plot of the relevant data and the OLS regression line. We allow for different slopes in the negative and positive domain. Again we observe a kink when we go from a negative to a positive surprise. The shaded area and the diagonal in Fig. 5 show the range of slopes predicted by the various models of social preferences. Irrespective of whether we estimate the reaction as a whole or piecewise we can rule out a negative slope for the reaction function.

Apart from the large number of observations in the origin (which are compatible with any slope of the reaction function) there are very few observations that are compatible with a negatively-sloped reaction function. The Fehr–Schmidt model can account for a positively-sloped reaction function (see Table 4 and Fig. 4D). A *BA*-type Fehr–Schmidt agent chooses to match the other agent's effort up to a certain threshold. According to the parameter calibration suggested by Fehr and Schmidt (1999) we should observe only 10% *BA*-type agents in the experiments with the high productivity parameter; the estimates provided by Blanco, Engelmann, and Normann (2011) suggest 21% *BA*-type agents.

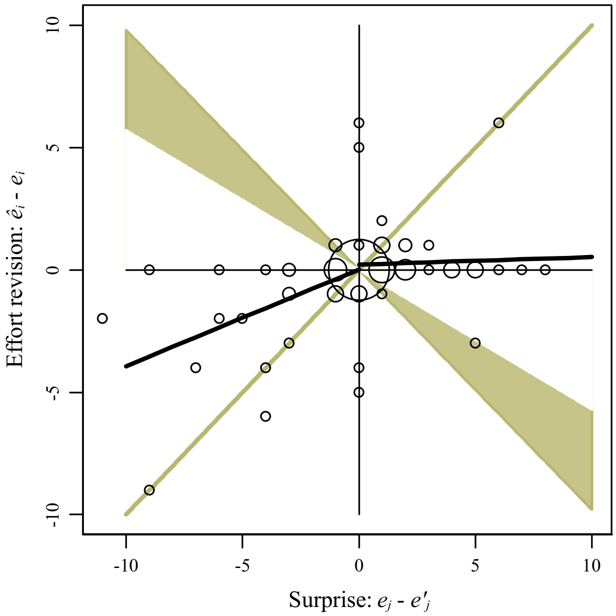

Fig. 5. Effort revisions dependent on the difference between the actual e_j and agent i's belief e_j'. The shaded areas and the diagonal are predictions consistent with theories of social preferences.

Could it be that the *BA*-type agent is much more frequent among our subjects? An observation is called *BA-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. Of the 220 observations in the EIT, 95 (43%) are compatible with the *BA*-prediction. However, a lot of these observations are minimal efforts and therefore also compatible with the standard prediction. If we restrict our sample to agents with non-minimal initial efforts then only 16 out of 96 (17%) choose their efforts in accordance with the *BA* type. Another possibility is to look at the fraction of effort revisions ($\Delta e_i \neq 0$) that are explained by the *BA*-type behavior. Among the 73 agents who do revise their effort only 14 agents (19%) do so according to the prediction.

The second theory that predicted positively-sloped reaction functions is the intermediate type in Charness and Rabin (2002). Thirteen (6%) out of the 220 observations in the EIT follow this pattern. Among the 73 agents who do revise their effort six (8%) do so as predicted by the intermediate type. Furthermore, although the slope estimates are positive, they are nowhere near unity, as the two theoretical cases would predict. An *F*-test rejects the hypothesis $\beta = 1$ for both the linear and the piecewise linear estimate ($p < .01$).

We summarize our findings in Result 3:

Result 3: *Peer effects in voluntary cooperation predominately take the form of efforts being strategic complements rather than substitutes as predicted by most theories of social preferences.*

The explanations offered by standard theories of social preferences do not capture the peer effects we observe. One apparent possibility to account for this is to alter the definition of the reference group. Suppose that for some reason, agents only compare among themselves, e.g., because they feel more attached to the co-agent than to the principal. Yet, redefining the reference group to comprise only the agents in not convincing. To see why, assume an agent with distributional preferences $u(x_i, x_j)$. By design, agent *i* cannot influence x_j, the profit of player *j*. Therefore, even with altruistic preferences agent *i* chooses minimal effort irrespective of e_j. A Fehr–Schmidt agent would as well always choose minimal effort, because adjusting to a higher co-agent's effort would require $\beta > 1$, which is ruled out by the Fehr and Schmidt (1999) model. The only model that predicts strategic complementarity in this case is Bolton and Ockenfels (2000). In the extreme case of an exclusively equity-oriented agent the reaction function matches the other agent's effort. But even in this situation it is unclear why two such agents should coordinate on non-minimal efforts, because they could increase utility by choosing minimal effort.

7. Discussion: Economic significance and explanations for peer effects in voluntary cooperation

Before we investigate potential explanations of peer effects we discuss the economic significance of efforts as strategic complements. Positively correlated efforts due to peer effects allow for the possibility of multiple equilibria, i.e., high and low effort cultures, depending on the team composition. The 'kinked effort function' in Fig. 2 suggests that a low effort situation is more likely to emerge because people are not willing to increase effort if others have exerted higher efforts but are ready to reduce effort if others provided less effort. More research is needed, however, to understand the nature of positively correlated efforts.

Our findings contrast with the field (experimental) results reported by Falk and Ichino (2006) and Mas and Moretti (2009) who find the opposite, namely that mutual observability increases the productivity of the low effort agents. We can only speculate about the reasons for this result. One explanation is that our experiment is conducted in full anonymity whereas in these studies agents could observe and talk to each other, which might have exerted some social pressure on low-performing agents. An additional possibility is that our effort choice is an abstract decision, whereas in these studies effort is linked to a real task. 'Social facilitation' (Zajonc, 1965) suggests that the mere presence of another person can improve performance (see also Falk & Ichino, 2006, p. 48, for a related discussion).

We now turn to the theoretical discussion of our empirical results which are opposite to the predictions of a host of standard theories of social preferences. If standard models of social preferences cannot explain the peer effects we see, the question arises whether motivations that are not captured by the standard models can explain our results. One possibility is that people are conformists - a tendency long established by social psychologists (Asch, 1952). Conformism "refers to the act of changing one's behavior to match the responses of others" (Cialdini & Goldstein, 2004, p. 606). Conformity is a potential channel because it is a common and deeply rooted human predisposition (Henrich & Boyd, 1998) that can also explain important economic phenomena (see, e.g., Bernheim, 1994; Clark & Oswald, 1998) including ones related to our research question (Sliwka, 2007).

Another possibility is that people follow a norm of reciprocity but take the behavior of others as a cue about what is an appropriate reciprocal response. Such norm-following is empirically plausible (e.g., Krupka & Weber, 2013; López-Pérez, 2008).

Positively correlated efforts can also result if people are motivated by social esteem (Ellingsen & Johannesson, 2008) whereby effort choices are made to create favorable impressions in the other players. In the remainder we sketch three recent new generation models of social preferences that, respectively, incorporate conformity, norm-following, and social esteem as relevant social motivations into their frameworks. All three models predict that peer effects take the form of positively correlated efforts (see online Appendix E).

Sliwka (2007) presents a model that allows for *conformity*. *Selfish* agents have a utility function $u_S(x_i)$, while *fair* agents have distributional social preferences: $u_F(x_i, x_j, x_P)$. Conformism is introduced by assuming a third type, a *conformist* agent, whose utility is either u_S or u_F, depending on which type she thinks is more frequent in the population. All conformist agents have a prior about the distribution of types in the population. The revision stage in our experiment provides the agents with additional information about the distribution of types. An agent with non-minimal effort might thus conform to money-maximizing behavior if she is paired with an agent with minimal effort and vice versa.

Agents might also derive utility for *norm-following* (or disutility from breaking them). López-Pérez (2008) provides a formalization of this possibility. He starts with the simple idea that players share a common norm which guides behavior. The norm demands from an agent to choose an effort $\tilde{e}$. Utility is given as

$$u_i = \begin{cases} x_i & \text{if } e_i = \tilde{e} \\ x_i - \gamma r & \text{else} \end{cases} \tag{7}$$

where x_i denotes the earnings, $\gamma > 0$ is a preference parameter and r denotes the number of players who did not (yet) break the norm. If the agent sticks to the norm then her utility is equal to her earnings. If the agent deviates and chooses $e_i < \tilde{e}$ then she enjoys higher earnings but suffers a psychological cost of γr, which can be interpreted as a feeling of guilt or shame. These costs do not depend on the size of the deviation, which implies that whenever an agent deviates she chooses $e_i = 1$. Furthermore, breaking a norm is assumed to be less costly to the agent if others do so as well, that is, if r is low.

López-Pérez posits an efficiency and equity norm where a social welfare function is maximized which contains (i) the sum of all earnings and (ii) the difference between the best-off and worst-off player. For the three-person gift-exchange game the norm demands the principal to pay the highest wage and the two agents to choose either maximum effort ($\tilde{e} = 20$) or the effort which equalizes all earnings at the highest wage ($\tilde{e} = \bar{e}|_{w=200} = 607/(2\upsilon + 7)$), depending on the relative weight of argument (i) and (ii).

An agent's effort depends on the strength of her preference parameter (γ) and on whether she observes the co-agent violating the norm. If γ is large (small), the agent always (never) follows the norm. There is an interesting intermediate range of γ where an agent starts by obeying the norm and thus chooses non-minimal effort, but revises to minimal effort if she learns that the co-agent did not follow the norm.

Considerations of *social esteem* can also explain positively correlated efforts. Ellingsen and Johannesson (2008) model players who care about what others think of them.[10] Their model is only defined for two players. We assume a utility function adapted for our three-player gift-exchange game:

$$u_i = x_i + \theta_i(x_j + x_P) + \theta_{ij}\sigma_j + \theta_{iP}\sigma_P. \tag{8}$$

The first two terms concern the material outcomes of the game for which the model assumes altruistic preferences ($\theta_i \geqslant 0$, similar to the model of Cox et al. (2007) for $\alpha = 1$) and θ_{ij} measures agent j's estimation about θ_i, interpreted as j's esteem for i. Finally σ_j measures how important j's opinion is for i, and it is assumed that σ_j is increasing in θ_j, i.e., the higher the altruism of the other player the more his opinion matters to agent i. Taken together, the third term in (8) represents agent i's pride from the interaction with the other agent and the fourth term is i's pride from the interaction with the principal.

The model turns the gift-exchange game into a signaling game. Ellingsen and Johannesson assume that there are two types of agents, altruists with θ_H and selfish players with θ_L ($\theta_H > \theta_L$). They show that in a separating equilibrium selfish agents choose minimal effort while altruistic agents signal their type to the principal by choosing a non-minimal effort $\tilde{e}_i$. In the three-player gift-exchange game agents not only signal their type to the principal but also to the other agent. When choosing the initial effort agents do not know the type of the other agent but have a prior probability p of expecting an altruistic type. In online Appendix E9 we show that the effort necessary to signal altruistic preferences is increasing in p, that is, there is a function $\tilde{e}_i(p)$ with $\tilde{e}_i' > 0$. Intuitively, the more likely it is that i's co-agent is altruistic the more pride agent i can gain by being regarded as an altruist, irrespective of whether she actually is an altruist or not. Thus, to credibly demonstrate her altruism, i must become more generous to the principal. In our three-player gift-exchange game selfish players always choose minimal effort. Altruistic agents initially choose $e_i = \tilde{e}_i(p) > 1$. In the revision stage agents learn the type of their co-agent and update their prior probability to either 0 or 1. When paired with a selfish player they lower their effort to $\hat{e}_i = \tilde{e}_i(0)$, else they increase their effort to $\hat{e}_i = \tilde{e}_i(1)$. Thus, concerns for social esteem can explain positively correlated efforts, but cannot explain the kink in effort revisions observed in the experiment.

In summary, newer theories of social preferences that expand social motives beyond distributional concerns and intentions by incorporating desires for conformity, norm-following, or social esteem, can rationalize the empirically observed positively correlated efforts. The kink in the reaction to observed effort differences shown in Fig. 2 is best captured by the model of López-Pérez (2008). Of the three models considered here this is the only model that predicts a distinct asymmetric effect of the effort information in the three-person gift exchange game: some agents initially choose their effort according to

[10] Closely related is the model by Bénabou and Tirole (2006), which assumes that players differ in two dimensions, their preference for (i) the social good and (ii) money. Players choose their actions to signal high interest in (i) and low interest in (ii). Andreoni and Bernheim (2009) present a model which formulates the players' desire to be perceived as fair and apply it to dictator games. Guilt aversion (Battigalli & Dufwenberg, 2007) can also explain positively correlated efforts. A guilt averse agent may form a belief about what the principal expects and choose effort to conform to this expectation. Observing the co-agent's effort allows updating the belief about what the principal might expect and this may induce the agent to change effort.

a norm and turn to a selfish strategy once they observe others breaking the norm (this is reminiscent of 'moral wiggle room' (Dana, Weber, & Kuang, 2007). If they find it optimal to break the norm in the first place then observing high co-agent's efforts does not turn them into norm-abiding players. This suggests that peer effects as observed in our experiment are better explained by models of conformity, norm-guided behavior or considerations of social esteem than more direct motives such as altruism or inequity aversion.

It is also possible to enrich standard theories by conformity or norm abiding motives. In models of reciprocity the necessary steps seem obvious: the parameter which governs positive or negative reciprocity toward the principal needs to depend somehow on the information about the co-agent's action. For example, in Cox et al. (2007) we could reformulate the emotional state θ_P as being positively related to agent *i*'s belief about agent *j*'s emotional state toward the principal, thus introducing a desire to conform into the model. In the revision stage the agent would then update his information about the emotional state of the co-agent and we could generate positively sloped reaction functions.

One may object that these newer theories also allow for more social motives than the standard theories of social preferences, and one standard theory, the Fehr and Schmidt (1999) model (the *BA*-type, see Table 4), actually can, at least qualitatively, explain positively correlated efforts without resorting to additional motives. Put differently, Fehr and Schmidt (1999) might provide a parsimonious explanation of peer effects in voluntary cooperation, if we are prepared to relax the prediction that agents choose the same efforts to positively correlated efforts. Whether this is a valid argument is a task for future research and Gächter, Nosenzo, and Sefton (2013) provide a first step in this direction. They build on the theoretical results of this paper (that Fehr–Schmidt preferences can, at least qualitatively, explain peer effects) and also provide evidence that the measured descriptive norms are consistent with the peer effects we see as well. Interestingly, a quantitative comparison between norms and Fehr–Schmidt preferences for peer effects suggests a surprisingly strong explanatory power of Fehr–Schmidt preferences, despite unambiguous evidence that descriptive norms exhibit peer effects.

8. Summary

We developed a novel experimental gift-exchange game with effort revision to study the role of peer effects in social preferences. Our design avoids the reflection problem and allows studying the reaction of an agent to the information about a co-agent's effort in isolation. We find that efforts are strategic complements, in particular when the co-agent's effort is lower than the agent's effort. This poses a challenge to the standard theories of social preferences which predict either unrelated efforts or strategic substitutes. Newer theories of conformity, social norms, and esteem, do a better job in explaining the peer effects we see.

Acknowledgments

We are grateful for helpful comments by Benjamin Beranek, Friedrich Breyer, Stefan Bühler, Rachel Croson, Tore Ellingsen, Urs Fischbacher, Yukihiko Funaki, Martin Kolmar, Tatiana Kornienko, Michael Kosfeld, Bentley MacLeod, Daniele Nosenzo, Rupert Sausgruber, Martin Sefton, Karl Schlag, Rudi Stracke, Jean-Robert Tyran, and participants at various seminars, conferences and workshops. We acknowledge financial support from the Grundlagenforschungsfonds of the University of St.Gallen. Simon Gächter also acknowledges support under the European Research Council Advanced Investigator Grant ERC-AdG 295707 COOPERATION, the ESRC-funded research Network of Integrated Behavioural Science (NIBS, ES/K002201/1), and the hospitality of the Institute for Advanced Studies at Hebrew University in Jerusalem while working on this paper.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.joep.2015.03.001.

References

Andreoni, J., & Bernheim, D. B. (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica, 77*, 1607–1636.
Asch, S. E. (1952). *Social psychology*. Englewood Cliffs: Prentice Hall.
Bandiera, O., Barankay, I., & Rasul, I. (2010). Social incentives in the workplace. *Review of Economic Studies, 77*, 417–458.
Bardsley, N., & Sausgruber, R. (2005). Conformity and reciprocity in public good provision. *Journal of Economic Psychology, 26*, 664–681.
Battigalli, P., & Dufwenberg, M. (2007). Guilt in games. *The American Economic Review, 97*, 170–176.
Bénabou, R., & Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review, 96*, 1652–1678.
Bernheim, B. D. (1994). A theory of conformity. *Journal of Political Economy, 102*, 841–877.
Bicchieri, C., & Xiao, E. (2009). Do the right thing: But only if others do so. *Journal of Behavioral Decision Making, 22*, 191–208.
Blanco, M., Engelmann, D., & Normann, H. T. (2011). A within-subject analysis of other-regarding preferences. *Games and Economic Behavior, 72*, 321–338.
Bolton, G. E. (1991). A comparative model of bargaining - theory and evidence. *American Economic Review, 81*, 1096–1136.
Bolton, G. E., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review, 90*, 166–193.
Cason, T., & Mui, V.-L. (1998). Social influence in the sequential dictator game. *Journal of Mathematical Psychology, 42*, 248–265.
Charness, G., & Kuhn, P. (2011). Lab labor: What can labor economists learn from the lab? In O. Ashenfelter & D. Card (Eds.), *Handbook of labor economics* (pp. 229–330). Amsterdam: Elsevier.
Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics, 117*, 817–869.

Chen, Y., Harper, F. M., Konstan, J., & Li, S. X. (2010). Social comparisons and contributions to online communities: A field experiment on movielens. *American Economic Review, 100*, 1358–1398.
Christakis, N. A., & Fowler, J. H. (2007). The spread of obesity in a large social network over 32 years. *New England Journal of Medicine, 357*, 370–379.
Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology, 55*, 591–621.
Clark, A. E., & Oswald, A. J. (1998). Comparison-concave utility and following behaviour in social and economic settings. *Journal of Public Economics, 70*, 133–155.
Cox, J. C., Friedman, D., & Gjerstad, S. (2007). A tractable model of reciprocity and fairness. *Games and Economic Behavior, 59*, 17–45.
Cox, J. C., Friedman, D., & Sadiraj, V. (2008). Revealed altruism. *Econometrica, 76*, 31–69.
Cox, J. C., & Sadiraj, V. (2012). Direct tests of individual preferences for equity and efficiency. *Economic Inquiry, 50*, 920–931.
Croson, R., & Shang, J. (2008). The impact of downward social information on contribution decisions. *Experimental Economics, 11*, 221–233.
Dana, J., Weber, R., & Kuang, J. (2007). Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness. *Economic Theory, 33*, 67–80.
Daruvala, D. (2010). Would the right social preference model please stand up! *Journal of Economic Behavior & Organization, 73*, 199–208.
Dong, B., Dulleck, U., & Torgler, B. (2012). Conditional corruption. *Journal of Economic Psychology, 33*, 609–627.
Duflo, E., & Saez, E. (2002). Participation and investment decisions in a retirement plan: The influence of colleagues' choices. *Journal of Public Economics, 85*, 121–148.
Dufwenberg, M., & Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and Economic Behavior, 47*, 268–298.
Ellingsen, T., & Johannesson, M. (2008). Pride and prejudice: The human side of incentive theory. *American Economic Review, 98*, 990–1008.
Engelmann, D., & Strobel, M. (2004). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *American Economic Review, 94*, 857–869.
Falk, A., Fehr, E., & Fischbacher, U. (2008). Testing theories of fairness–intentions matter. *Games and Economic Behavior, 62*, 287–303.
Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior, 54*, 293–315.
Falk, A., Fischbacher, U., & Gächter, S. (2013). Living in two neighborhoods - social interaction effects in the laboratory. *Economic Inquiry, 51*, 563–578.
Falk, A., & Ichino, A. (2006). Clean evidence on peer effects. *Journal of Labor Economics, 24*, 38–57.
Fehr, E., Kirchsteiger, G., & Riedl, A. (1993). Does fairness prevent market clearing? An experimental investigation. *Quarterly Journal of Economics, 108*, 437–459.
Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics, 114*, 817–868.
Fischbacher, U. (2007). Z-tree: Zurich toolbox for readymade economic experiments. *Experimental Economics, 10*, 171–178.
Frey, B. S., & Meier, S. (2004). Social comparisons and pro-social behavior. Testing 'conditional cooperation' in a field experiment. *American Economic Review, 94*, 1717–1722.
Gächter, S., Nosenzo, D., & Sefton, M. (2012). The impact of social comparisons on reciprocity. *The Scandinavian Journal of Economics, 114*, 1346–1367.
Gächter, S., Nosenzo, D., & Sefton, M. (2013). Peer effects in pro-social behavior: Social norms or social preferences? *Journal of the European Economic Association, 11*, 548–573.
Gächter, S., & Renner, E. (2010). The effects of (incentivized) belief elicitation in public goods experiments. *Experimental Economics, 13*, 364–377.
Gächter, S., & Thöni, C. (2010). Social comparison and performance: Experimental evidence on the fair wage-effort hypothesis. *Journal of Economic Behavior & Organization, 76*, 531–543.
Gintis, H., Bowles, S., Boyd, R., & Fehr, E. (Eds.). (2005). *Moral sentiments and material interests: The foundations of cooperation in economic life*. Cambridge: MIT Press.
Greiner, B. (2004). An online recruitment system for economic experiments. In K. Kremer & V. Macho (Eds.). *Forschung und wissenschaftliches rechnen gwdg bericht* (Vol. 63, pp. 79–93). Göttingen: Gesellschaft für Wissenschaftliche Datenverarbeitung.
Henrich, J., & Boyd, R. (1998). The evolution of conformist transmission and the emergence of between-group differences. *Evolution and Human Behavior, 19*, 215–241.
Ho, T. H., & Su, X. M. (2009). Peer-induced fairness in games. *American Economic Review, 99*, 2022–2049.
Ichino, A., & Maggi, G. (2000). Work environment and individual background: Explaining regional shirking differentials in a large italian firm. *Quarterly Journal of Economics, 115*, 1057–1090.
Kagel, J. H., & Wolfe, K. W. (2001). Tests of fairness models based on equity considerations in a three-person ultimatum game. *Experimental Economics, 4*, 203–219.
Kirchsteiger, G. (1994). The role of envy in ultimatum games. *Journal of Economic Behavior and Organization, 25*, 373–389.
Kremer, M., & Levy, D. (2008). Peer effects and alcohol use among college students. *Journal of Economic Perspectives, 22*, 189–206.
Krupka, E., & Weber, R. A. (2009). The focusing and informational effects of norms on pro-social behavior. *Journal of Economic Psychology, 30*, 307–320.
Krupka, E. L., & Weber, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association, 11*, 495–524.
Levine, D. K. (1998). Modeling altruism and spitefulness in experiments. *Review of Economic Dynamics, 1*, 593–622.
Loewenstein, G., Thompson, L., & Bazerman, M. (1989). Social utility and decision making in interpersonal contexts. *Journal of Personality and Social Psychology, 57*, 426–441.
López-Pérez, R. (2008). Aversion to norm-breaking: A model. *Games and Economic Behavior, 64*, 237–267.
Manski, C. F. (2000). Economic analysis of social interactions. *Journal of Economic Perspectives, 13*, 115–136.
Mas, A., & Moretti, E. (2009). Peers at work. *American Economic Review, 99*, 112–145.
Mittone, L., & Ploner, M. (2011). Peer pressure, social spillovers, and reciprocity: An experimental analysis. *Experimental Economics, 14*, 203–222.
Prinz, A., Muehlbacher, S., & Kirchler, E. (2014). The slippery slope framework on tax compliance: An attempt to formalization. *Journal of Economic Psychology, 40*, 20–34.
Rabin, M. (1993). Incorporating fairness into game-theory and economics. *American Economic Review, 83*, 1281–1302.
Rao, J. N. K., & Scott, A. J. (1984). On chi-squared tests for multiway contingency tables with cell proportions estimated from survey data. *Annals of Statistics, 12*, 46–60.
Ross, L., & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology*. New York: McGraw Hill Inc.
Sacerdote, B. (2001). Peer effects with random assignment: Results for dartmouth roommates. *Quarterly Journal of Economics, 116*, 681–704.
Sampson, R. J., Morenoff, J. D., & Gannon-Rowley, T. (2002). Assessing "neighborhood effects": Social processes and new directions in research. *Annual Review of Sociology, 28*, 443–478.
Shang, J., & Croson, R. (2009). A field experiment in charitable contribution: The impact of social information on the voluntary provision of public goods. *Economic Journal, 119*, 1422–1439.
Sliwka, D. (2007). Trust as a signal of a social norm and the hidden costs of incentive schemes. *The American Economic Review, 97*, 999–1012.
Volk, S., Thöni, C., & Ruigrok, W. (2012). Temporal stability and psychological foundations of cooperation preferences. *Journal of Economic Behavior & Organization, 81*, 664–676.
Weber, R. A. (2003). 'Learning' with no feedback in a competitive guessing game. *Games and Economic Behavior, 44*, 134–144.
Zajonc, R. B. (1965). Social facilitation. *Science, 149*, 269–274.
Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics, 13*, 75–98.