PEER EFFECTS IN PRO-SOCIAL BEHAVIOR: SOCIAL NORMS OR SOCIAL PREFERENCES?

Simon Gächter
University of Nottingham

Daniele Nosenzo
University of Nottingham

Martin Sefton
University of Nottingham

Abstract
We compare social preference and social norm–based explanations for peer effects in a three-person gift exchange experiment. In the experiment a principal pays a wage to each of two agents, who then make effort choices sequentially. In our baseline treatment we observe that the second agent's effort is influenced by the effort choice of the first agent, even though there are no material spillovers between agents. This peer effect is predicted by the Fehr–Schmidt (1999) model of social preferences. As we show from a norms elicitation experiment, it is also consistent with social norms compliance. A conditional logit investigation of the explanatory power of payoff inequality and elicited norms finds that the second agent's effort is best explained by the social preferences model. In further experiments we find that the peer effects change as predicted by the social preferences model. Again, a conditional logit analysis favors an explanation based on social preferences, rather than social norms. Our results suggest that, in our context, the social preferences model provides a parsimonious explanation for the observed peer effect. (JEL: A13, C92, D03)

1. Introduction

A large body of experimental evidence challenges the assumption that individuals act exclusively to maximize their narrowly defined self-interest. In simple games such as

The editor in charge of this paper was Stefano DellaVigna.

Acknowledgments: We thank Yan Chen, Stefano DellaVigna, Erin Krupka, Stephen Leider, Ilana Ritov, Christian Thöni, Roberto Weber, Luca Zarri, three anonymous referees, participants at the 2010 ESA Conference in Tucson, the 2010 ESA Conference in Copenhagen, the Thurgau Experimental Economics Meeting, the University of Southampton workshop on "The Determinants and Implications of Pro-social Behavior", the University of Forlí workshop on "Social Economics", seminar audiences in Jerusalem, Haifa, Bar-Ilan University, Michigan, Pittsburgh and Tokyo for helpful comments. Lu Dong, Lingbo Huang, and Josh Welsh provided excellent research assistance. Erin Krupka kindly provided software for the conditional logit analysis. Simon Gächter is grateful for the hospitality of the Institute for Advanced Studies at Hebrew University in Jerusalem while working on this paper. Daniele Nosenzo acknowledges support from the Leverhulme Trust (ECF/2010/0636) and Simon Gächter from the European Research Council Grant ERC-AdG 295707 COOPERATION. Gächter is a Research Fellow at IZA and a Research Network Fellow at CESifo.

E-mail: simon.gaechter@nottingham.ac.uk (Gächter); daniele.nosenzo@nottingham.ac.uk (Nosenzo); martin.sefton@nottingham.ac.uk (Sefton)

Journal of the European Economic Association June 2013 11(3):548–573

DOI: 10.1111/jeea.12015

the dictator game, the investment game and the gift exchange game, individuals often choose to reduce their own earnings in order to increase the earnings of their opponent, even in nonrepeated and anonymous interactions where there are no positive future consequences associated with behaving pro-socially (see, e.g., Camerer 2003). More recently, several studies of peer effects show that the extent to which people behave pro-socially is sensitive to whether they observe others doing so.[1]

Previous studies have typically interpreted these peer effects by referring to different behavioral mechanisms. One view is that pro-social behavior originates from pressures to comply with norms of "socially appropriate" behavior in a given context (e.g., Bicchieri 2005; Keizer, Lindenberg, and Steg 2008). Here observing what others do can affect individuals' actions by influencing what they perceive to constitute appropriate behavior in a given situation. An alternative view emphasizes the role of stable distributional social preferences, which can also predict nonselfish behavior and, as we show in what follows, peer effects. Our paper reports experiments designed to separate these leading explanations of pro-social behavior and peer effects in one particular context, involving how employee effort is sensitive to the efforts of other employees.

Our starting point is a three-person gift exchange game described in Section 2. In the experiment subjects are grouped in experimental "firms" composed of three members: "Employer", "Employee 1" and "Employee 2". Employers choose wages for their employees, who choose costly effort in return. A key feature of the game is that, before making an effort choice, Employee 2 is informed of the co-worker's effort choice. Previous studies with this experimental setting have found that the availability of information about others' efforts has a systematic influence on Employee 2's decisions: in particular, when the Employer pays high wages to both employees, Employee 2 expends more effort when she observes her co-worker supplying more effort, despite the fact that there are no earnings interdependencies between the employees. In this paper we replicate these results and show that a model of social preferences where individuals dislike inequitable distributions of earnings (Fehr and Schmidt 1999) can explain this effect. However, social norm compliance may also explain the positive relation between employees' efforts. To verify this, in Section 3 we identify social norms in the three-person gift exchange game by applying the elicitation procedure introduced by Krupka and Weber (2013) using incentivized coordination games this procedure allows to determine which actions are viewed as most socially appropriate in a given situation. We find that the co-worker's effort decisions systematically affect the perception of what constitutes an appropriate effort response by Employee 2 to a high wage offer. Thus, the behavioral mechanism underlying this peer effect is unclear.

1. See, for example, Cason and Mui (1998); Bicchieri and Xiao (2009), and Krupka and Weber (2009) in the dictator game; Gächter, Nosenzo, and Sefton (2012) and Thöni and Gächter (2011) in the gift exchange game; Mittone and Ploner (2011) in the investment game. Related, in public goods game experiments many people are "conditional cooperators" who are willing to contribute in proportion to what others contribute (see, e.g., Falk, Fischbacher, and Gächter 2013; Fischbacher and Gächter 2010). The importance of conditional cooperation has also been documented in a number of recent field experiments (Chen et al. 2010; Shang and Croson 2009; Alpizar, Carlsson, and Johansson-Stenman 2008; Croson and Shang 2008; Martin and Randal 2008; Frey and Meier 2004; Rustagi, Engel, and Kosfeld 2010).

In Section 4 we compare the explanatory power of social norms versus distributional social preferences using a conditional logit framework. We consider models in which Employee 2's effort choice depends on their own payoff associated with each possible effort level, and then augment this model by including additional explanatory variables measuring the payoff inequality and/or social appropriateness of each possible effort level. We find that a social preferences model in which players trade off own payoff and payoff inequality offers a parsimonious account of choice behavior. In our most general specification we find that payoff inequalities have a significant effect on choice behavior, whereas social appropriateness does not.

These findings stand in contrast with those of two other recent studies on the relative importance of social norms and preferences in other contexts. Krupka and Weber (2013) find that social norm compliance can explain variations in Dictator game behavior that cannot be accounted for by social preference models, and Krupka, Leider, and Jiang (2012) find substantial explanatory power from elicited social norms in Dictator and Bertrand games.

We examine the robustness of our finding in Section 5 by conducting two new treatments with modified games. In the ASYMMETRIC treatment we introduce small wage asymmetries in the high-wage combination, weakening the peer effect predicted by the social preferences model. These asymmetries may also affect norm compliance by creating an opportunity for "moral wiggle room" exploitation (Dana, Weber, and Kuang 2007). Commensurate with this, we also find a weaker peer effect in the behavioral data. We also compare the explanatory power of payoff inequalities and elicited norms for this treatment and find that the social preferences model performs better in explaining choices. In our RANDOM treatment, only the wage and effort of one randomly selected employee are used for computing earnings, eliminating entirely the peer effect predicted by the social preference model. In this treatment, again consistent with the social preferences model, we do not observe a significant peer effect in the behavioral data. Again, an econometric comparison of the explanatory power of payoff inequalities and elicited norms favors the social preference model.

Overall, the distributional social preferences model organizes somewhat well peer effects in pro-social behavior in our context. However, we also observe differences across treatments in the proportion of individuals who are willing to engage in pro-social behavior, and this is more difficult to reconcile with the social preference model. In Section 1 we discuss our results and conclude.

2. Peer Effects in Trilateral Gift Exchange Games

We examine peer effects in pro-social behavior using a three-person version of the standard bilateral gift exchange game (Fehr, Kirchsteiger, and Riedl 1993). In the trilateral gift exchange game "firms" are composed of three members: one "Employer" and two "Employees". The game begins with the Employer choosing wages for the two employees. Wages are publicly observed before employees choose costly effort which is beneficial to the Employer. Previous experiments with this type of game show the existence of a positive "own-wage effect": employees often reciprocate higher wage offers with higher effort, as in standard two-person gift exchange game experiments. In addition, two previous studies (Gächter, Nosenzo, and Sefton 2012, henceforth GNS; Thöni and Gächter 2011) show that the presence of information about the co-worker's effort can systematically affect the magnitude of this own-wage effect even when there are no earnings interdependencies between the employees. GNS, and Thöni and Gächter (2011) find that when the effort of the co-worker is publicly observable employees' willingness to reciprocate high wages depends on whether the co-worker is also willing to do so. Thus, a positive and systematic correlation between employees' efforts has been documented in some trilateral gift-exchange game experiments.[2]

In this paper we begin by reporting data from new sessions using 81 subjects playing the same three-person gift exchange game as in GNS. This game, which we refer to as the BASELINE game, has three players labeled "Employer", "Employee 1" and "Employee 2". The structure of the game, which is common information to all players, is as follows. All players move sequentially: the Employer moves first and chooses a wage $w_i \in \{16, 32\}$ for each Employee $i \in \{1, 2\}$. Employee 1 observes both wages and then chooses an effort level $e_1 \in \{1, 2, 3, 4\}$. Finally, Employee 2 observes both wages *and* the effort chosen by Employee 1, and then chooses an effort level $e_2 \in \{1, 2, 3, 4\}$. After Employee 2's choice, the game ends and the Employer's earnings are computed as

$$\pi_{ER} = 10\,(e_1 + e_2) - w_1 - w_2 \tag{1}$$

and employee i's earnings are computed as

$$\pi_i = w_i - 5\,(e_1 - 1). \tag{2}$$

As in GNS, the new sessions implemented the game as a one-shot game using the strategy method (Selten 1967)—that is, subjects had to make contingent decisions for all nodes at which they may have to play.[3] Thus, for each Employee 2 we observe sixteen effort choices, one for each possible combination of wages and effort chosen by the Employer and Employee 1. Figure 1 shows, for each possible combination,

2. A number of studies also examine the effect of information about a co-worker's wage. In some cases wage comparisons systematically affects effort decisions (e.g., Abeler et al. 2010; Gächter and Thöni 2010; Nosenzo 2011), while in other cases such effects are weak or absent (e.g., GNS; Charness and Kuhn 2007; Güth et al. 2001).

3. Details on experimental procedures and instructions are provided in an Online Appendix. The GNS sessions were conducted in 2007 and the new sessions in 2009 using different subjects, but using the same University of Nottingham subject pool, recruitment procedures (using ORSEE, Greiner 2004), and for the most part, the same experimental procedures and software (using *z*-Tree, Fischbacher 2007). The only substantive difference in procedures was that the new sessions were implemented in two parts, where in the second part subjects played the game in direct-response mode. This was done to address concerns that decisions elicited by the strategy method may differ systematically from those elicited by the direct-response method (see Brandts and Charness 2011 for a discussion of this issue). Across all treatments reported in this paper 80% of the subjects made the same direct-response choice as was implied by their submitted strategies, and we could not detect any systematic pattern in deviations from submitted strategies. The empirical analysis presented in this paper is based on the strategy method choices.

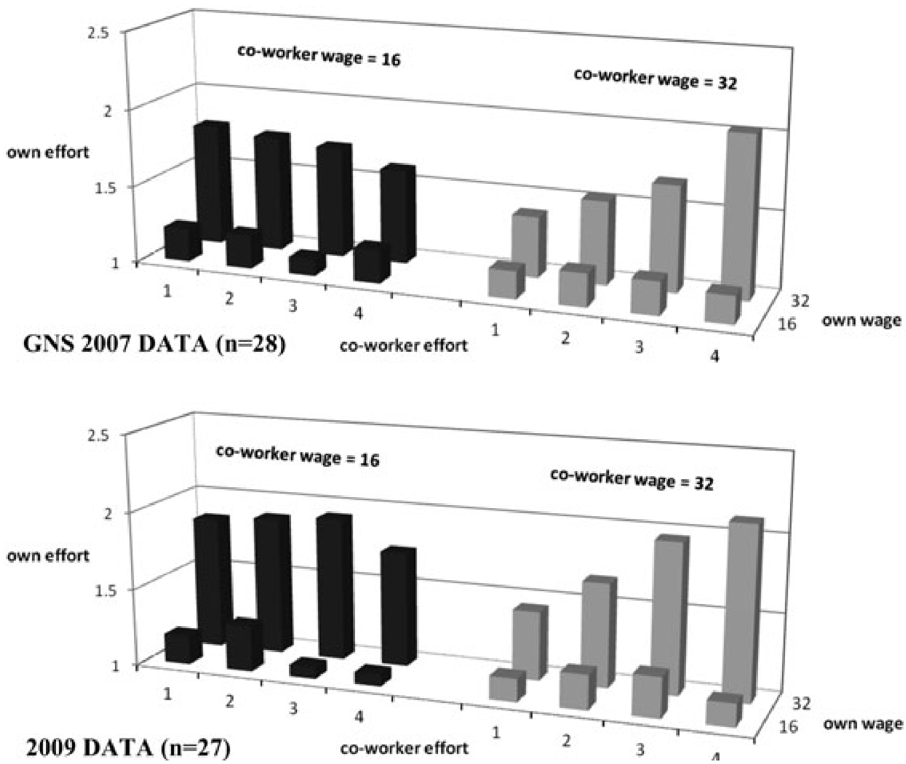

FIGURE 1. Employee 2 average effort in BASELINE.

the average effort expended by the 28 Employees 2 who participated in the GNS experiment and the 27 Employees 2 from the new sessions.

A striking feature of Figure 1 is the similarity of the data in the two panels. In fact for none of the 16 contingencies shown in the figure is effort significantly different across panels ($p > 0.316$ across all 16 cases according to two-sided Wilcoxon rank sum tests). Thus, in the remainder of the paper we pool the data from GNS and the new sessions and refer to this as our BASELINE treatment.

Two other features are striking. First, as in many other gift exchange game experiments, Figure 1 shows a positive own-wage effect: Employees 2 reward the Employer with higher effort when they are paid a higher wage. Employee 2's effort when the own wage is 32 is significantly higher than when the own wage is 16 ($p \leq 0.004$ across all eight cases according to two-sided Wilcoxon signed-rank tests). Second, Figure 1 shows a peer effect when both employees are paid a high wage: focusing on the extreme cases where the co-worker increases her effort from 1 to 4 units, Employee 2's average effort increases by 0.65 units (from 1.42 to 2.07). This peer effect is highly significant ($p < 0.001$ according to a one-tailed Page test for ordered alternatives).

What underlies this peer effect? In the next sections we discuss two alternative explanations: a model of social preferences and social norm compliance.

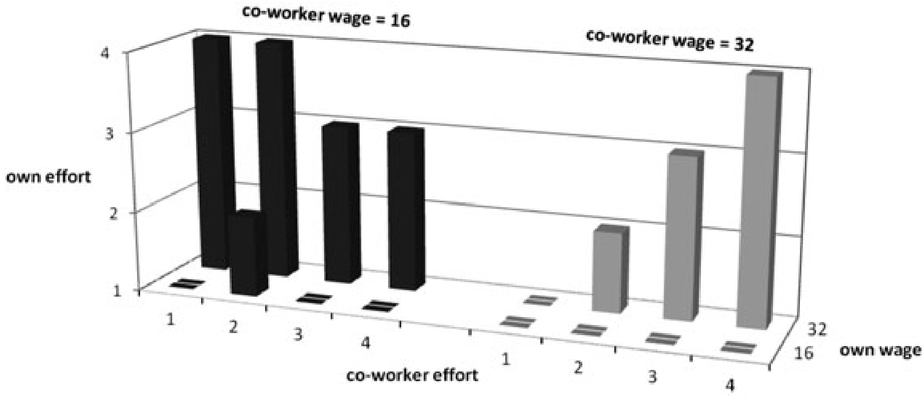

FIGURE 2. FS prediction of Employee 2 effort for BASELINE ($\beta_2 \geq 1/2$).

2.1. Explaining Peer Effects: the Social Preferences Approach

A leading model of social preferences that can explain the peer effect observed in the BASELINE game is the Fehr and Schmidt (1999) (FS) model of inequity aversion.[4] According to the FS model Employee 2's utility function is

$$U_2(\pi_2, \pi_{ER}, \pi_1) = \pi_2 - \frac{\alpha_2}{2}(\max\{\pi_{ER} - \pi_2, 0\} + \{\pi_1 - \pi_2, 0\})$$
$$- \frac{\beta_2}{2}(\max\{\pi_2 - \pi_{ER}, 0\} + \max\{\pi_2 - \pi_1, 0\}),$$

where π_{ER} and $\pi_{i \in \{1,2\}}$ are respectively the Employer's and employee i's material payoff from the game. The parameter α_2 measures the strength of the employee's aversion to disadvantageous payoff inequality, and β_2 measures the strength of the employee's aversion to advantageous inequality. As in FS, we assume $\alpha_2 \geq \beta_2$ and $0 \leq \beta_2 < 1$.

The model prediction for the BASELINE parameters depends only on the degree of "superiority aversion"β_2. If Employee 2 is not sufficiently superiority averse (i.e., $\beta_2 < 1/2$) she will choose minimal effort irrespective of her wage and the wage and effort of the co-worker, but a sufficiently superiority averse Employee 2 ($\beta_2 \geq 1/2$) may choose non-minimal effort as shown in Figure 2.[5]

Note how these predictions bear out the two most distinctive features of the experiments reproduced in Figure 1: firstly, the positive own-wage effect whereby higher effort is chosen in response to a higher wage, and secondly, the strong positive

4. Many other social preference models do not predict this positive peer effect. See Thöni and Gächter (2011) for a detailed analysis of the extent to which social preference models can accommodate peer effects in a closely related game. We return to this point in Section 6.

5. In a recent paper Blanco, Engelmann, and Normann (2011) observe behavior in a variety of experimental games and estimate that 56% of their subjects have a superiority aversion parameter larger than 0.5.

relation between employees' efforts in the wage combination where both wages are high. In the model, the own-wage effect results from Employee 2's superiority aversion: when she is paid a high wage, Employee 2 is generally better off than the other firm members. By choosing nonminimal effort levels she can reduce the advantageous payoff differences relative to the Employer and Employee 1 and attain more equitable earnings distributions. The one-to-one complementary relation between employees' efforts when both employees are paid a high wage results from the fact that Employee 2 is generally unwilling to choose nonminimal effort if by doing so her payoff falls (too much) short of the payoff of another firm member. Thus, when matched with a highly paid co-worker, Employee 2 is willing to expend effort and reduce advantageous payoff inequalities only to the extent that the co-worker is also willing to do so.[6]

2.2. *Explaining Peer Effects: the Social Norms Approach*

Peer effects could also be the result of a desire to comply with social norms—that is, collectively recognized rules of conduct that prescribe socially acceptable behaviors in a given situation (e.g., López-Pérez 2008). In the context of the trilateral gift exchange game experiment, a relevant norm is the "norm of reciprocity" whereby employees are expected to reciprocate a kind action by the employer (i.e., a high wage) with kindness (high efforts). However, observing how the co-worker responds to a high wage offer may systematically influence Employees 2's perception of what constitutes an appropriate reciprocal response to a high wage offer. Thus, if a descriptive social norm exists in this game and behavior is guided by such norms, Employees 2's effort choice may be systematically influenced by the co-worker's effort choice.

Explaining phenomena by appealing to the influence of social norms can be problematic because of the difficulties of precisely identifying and measuring norms. In particular, often what may or may not constitute a norm is based on intuition or casual empiricism. For a more objective approach toward identifying whether social norms are relevant to the experimental environment in BASELINE, and whether they can explain the patterns of behavior discussed previously, we adapted an experimental norms-elicitation procedure recently introduced by Krupka and Weber (2013) for dictator games. This procedure uses incentivized coordination games to identify which actions are viewed as most socially appropriate in a given situation. We describe this procedure for our trilateral gift exchange game in the next section.

3. Identifying Social Norms in the Trilateral Gift-Exchange Game

The norms elicitation experiment consisted of one session with 51 students recruited at the University of Nottingham and paid £5 for their participation. Upon arrival subjects learnt that their task was to read a description of four different "situations" in which a

6. Note that the model generates a negative relation between employees' efforts when Employee 2 is paid a high wage and the co-worker a low wage. This effect is not clearly observed in the BASELINE data.

"decision-maker" had to choose among a number of possible alternative actions.[7] The situations reproduced the following four possible contingencies faced by Employees 2 in our trilateral gift exchange game experiment:[8]

(1) Employer pays a wage of 32 to Employee 1 and a wage of 16 for Employee 2. Employee 1 chooses 1 unit of effort;
(2) Employer pays a wage of 32 to Employee 1 and a wage of 16 for Employee 2. Employee 1 chooses 4 units of effort;
(3) Employer pays a wage of 32 to both employees. Employee 1 chooses 1 unit of effort;
(4) Employer pays a wage of 32 to both employees. Employee 1 chooses 4 units of effort.

These situations were chosen to identify any norm of reciprocity (Situation 1 versus Situation 3, and Situation 2 versus Situation 4), and to examine how normative judgments are affected by observation of co-workers' actions (Situation 1 versus Situation 2, and Situation 3 versus Situation 4).

For each of these situations and for each effort choice available to Employee 2 subjects had to rate whether the effort choice was "very socially appropriate", "somewhat socially appropriate", "somewhat socially inappropriate", or "very socially inappropriate". Subjects were told that at the end of the session the experimenter would randomly select one of the four situations, and one possible effort choice in that situation. Subjects received an additional £3 if their appropriateness rating for the selected situation and choice matched the modal rating of other subjects in the session. Thus, subjects were given incentives to reveal their perception of the *most prevalent* appropriateness judgment in the session, and not their personal judgment.[9] The results of our norms elicitation experiment are shown in Table 1, which reports, for each situation, the average social appropriateness rating of the effort choices available to Employee 2.

The patterns of the social appropriateness ratings are consistent with the existence of a norm of reciprocity in BASELINE: when Employee 2 receives a high rather than low wage (Situations 3 and 4 versus Situations 1 and 2), the appropriateness of choosing low effort (i.e. 1 or 2 units of effort) decreases whereas the appropriateness of choosing high effort (i.e. 3 or 4 units of effort) increases.[10] Moreover, the perception

7. Instructions for the norms elicitation experiment are available in an Online Appendix.

8. We only elicited judgments about four of the 16 possible situations faced by Employees 2 in order to keep the size of the task and the duration of the experiment reasonable for subjects.

9. Thus, the structure of material incentives generates a coordination game with multiple equilibria. See Krupka and Weber (forthcoming) for a discussion of how the use of such "pure matching" coordination games constitutes an incentive-compatible tool to elicit social norms, and see Burks and Krupka (2012) and Krupka, Leider, and Jiang (2012) for other recent applications of this procedure. The use of coordination games to provide incentives for classifying natural language messages are used and discussed in Xiao and Houser (2005) and Houser and Xiao (2011).

10. For Situation 1 versus Situation 3 we find statistically significant differences in the ratings for low effort (two-sided Wilcoxon signed-rank tests $p = 0.004$ for effort $= 1$ and $p = 0.074$ for effort $= 2$). For

TABLE 1. Elicited norms—BASELINE.

Employee 2 effort	Situation 1 ($w_1 = 32$; $w_2 = 16$; $e_1 = 1$)	Situation 2 ($w_1 = 32$; $w_2 = 16$; $e_1 = 4$)	Situation 3 ($w_1 = 32$; $w_2 = 32$; $e_1 = 1$)	Situation 4 ($w_1 = 32$; $w_2 = 32$; $e_1 = 4$)
1	1.10	1.02	0.65	0.14
	(1.06)	*(0.84)*	*(0.98)*	*(0.40)*
2	1.57	2.31	1.33	0.92
	(0.83)	*(0.71)*	*(0.74)*	*(0.48)*
3	1.74	1.88	1.80	1.96
	(0.87)	*(0.68)*	*(0.72)*	*(0.44)*
4	1.82	1.45	2.04	2.96
	(1.24)	*(1.27)*	*(1.13)*	*(0.20)*

Notes: We transformed subjects' appropriateness ratings into numerical scores using the following scale: very socially inappropriate = 0; somewhat socially inappropriate = 1; somewhat socially appropriate = 2; very socially appropriate = 3. Standard deviations in parentheses.

of what constitutes an appropriate reciprocal response to a high wage is systematically influenced by how the co-worker behaves in similar circumstances. The last two columns of Table 1 show the appropriateness ratings for Situations 3 and 4, where both employees receive the same high wage and the co-worker's effort is either low (Situation 3) or high (Situation 4). The table shows that choosing low effort is more appropriate in Situation 3, when the co-worker also does so, than in Situation 4. Conversely, the appropriateness of choosing high effort is higher in Situation 4 than 3.[11]

The implications of our elicited norms can be gauged by the following exercise. Assume that, if asked to make an effort choice in the role of Employee 2 in each of the four situations they evaluated, the 51 raters in our experiment would select the level of effort that they evaluated as the "most" socially appropriate for that situation.[12] Figure 3 shows the average effort that our 51 raters would then expend in each of the four situations that they were asked to evaluate.

Note how the levels of effort reported in Figure 3 are higher than in the corresponding contingencies of the BASELINE experiment reported in Figure 1. This suggests that in the experiment individuals do not base their choices exclusively on social appropriateness, but take other motives (e.g. own payoff considerations) into account as well. Nevertheless, Figure 3 reproduces two distinctive features of Figure 1: firstly, for a given effort choice of the co-worker, the average effort increases with own wage (two-sided Wilcoxon signed-rank tests: $p = 0.003$ when the co-worker chooses one unit of effort; $p < 0.001$ when the co-worker chooses four units of effort).

Situation 2 versus Situation 4 we detect highly significant differences for effort choices of 1, 2, and 4 ($p < 0.001$) but not for effort = 3 ($p = 0.558$).

11. The differences in ratings between Situation 3 and Situation 4 are highly significant according to two-sided Wilcoxon signed-rank test for effort choices of 1, 2, and 4 ($p < 0.001$). The difference is not statistically significant for effort = 3 ($p = 0.215$).

12. If a subject assigns the highest appropriateness rating to more than one effort level, we select the effort level which gives him/her the highest material payoff.

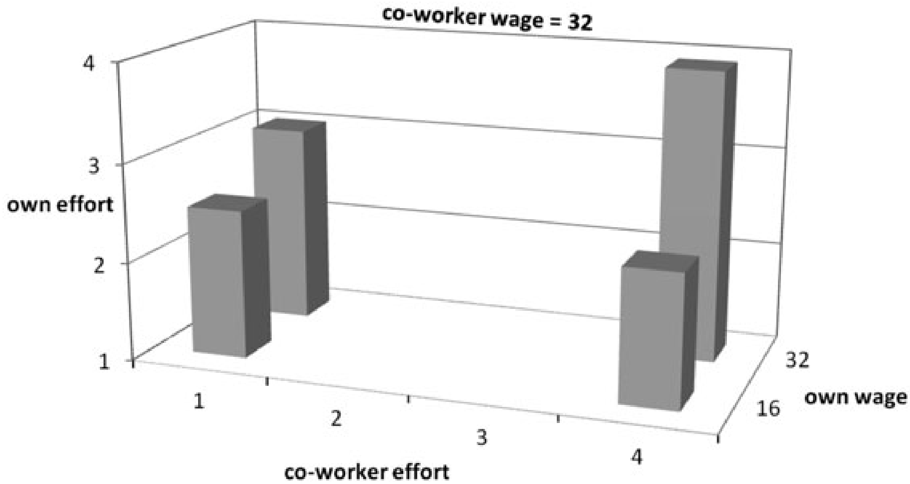

FIGURE 3. Socially most appropriate effort choices in BASELINE. Effort choices that the raters would make on average if they chose the most socially appropriate effort level in the four situations they evaluated.

This reproduces the positive own-wage effect shown in Figure 1. Secondly, when the Employer pays a high wage to both employees, average effort depends positively on the co-worker's effort: when the co-worker's effort increases from 1 to 4 the average effort increases by 0.90 units of effort (from 3.02 to 3.92, two-sided Wilcoxon signed-rank test: $p < 0.001$). This reproduces the peer effect shown in Figure 1.

Overall, our norms elicitation experiment reveals that if subjects comply with social norms this can also explain the behavioral regularities observed in the BASELINE game.

4. Disentangling Social Preferences and Social Norms in the Trilateral Gift Exchange Game

To investigate whether subjects are actually guided by distributional concerns and/or by a desire to comply with social norms we follow Krupka and Weber (2013) and Krupka, Leider, and Jiang (2012) and use a conditional logit model to examine the explanatory power of these motives in the BASELINE game.

In the model the probability that a subject in the role of Employee 2 chooses effort level k depends on the utility associated with that choice, $U(k)$, relative to the utility associated with the other alternatives:

$$\Pr\{e_2 = k\} = \frac{exp\{U(k)\}}{\sum_{i=1,\ldots,4} exp\{U(i)\}}, \quad k = 1, \ldots 4,$$

For each combination of wages and co-worker's effort, the four possible effort levels available to Employee 2 offer different outcomes in terms of own payoff, distribution of payoffs, and social appropriateness. Our first specification assumes that utility depends only on own payoff:

$$U(k) = \theta_1 \pi_2 = \theta_1(w_2 - 5(k-1)), \quad \text{[SELFISH]}$$

The selfish model predicts a distribution of effort choices for Employee 2 that is independent of own wage or co-worker's effort. The shape of the distribution depends on the θ_1 parameter. When $\theta_1 = 0$ each effort choice is equally likely and as θ_1 increases lower efforts (that give a higher own payoff) are more likely.

To investigate whether distributional concerns guide behavior we consider a second specification that augments SELFISH with a payoff inequality term as an additional explanatory variable:[13]

$$U(k) = \theta_1 \pi_2 + \theta_2(|\pi_2 - \pi_1| + |\pi_2 - \pi_{ER}|). \quad \text{[FS]}$$

To investigate whether social norms guide behavior we consider an alternative specification that augments SELFISH with the average norm rating associated with that action as an additional explanatory variable:

$$U(k) = \theta_1 \pi_2 + \theta_3 N(k). \quad \text{[NORMS]}$$

Finally, in our most general specification we assume that utility depends on own payoff, payoff inequality, and social appropriateness:

$$U(k) = \theta_1 \pi_2 + \theta_2(|\pi_2 - \pi_1| + |\pi_2 - \pi_{ER}|) + \theta_3 N(k). \quad \text{[FS + NORMS]}$$

Since, in order to keep the norm elicitation task manageable for our raters, we only elicited norms for four contingencies (see Section 3), the regressor $N(k)$ is only available for these four contingencies. Thus, we estimate the models using choice data from the BASELINE treatment for these four contingencies. As in Krupka, Leider, and Jiang (2012) we use bootstrapped standard errors for the models containing the average norm ratings.

Table 2 presents the results. In all models the coefficient on own payoff is positive and highly significant, showing that subjects are more likely to choose actions that give higher own payoffs. Augmenting the SELFISH model with either norm ratings or payoff inequality substantially improves explanatory power. The FS specification shows that effort levels that imply larger payoff inequality are chosen significantly less

13. The general FS model specification allows for different weights to be placed on disadvantageous and advantageous inequality. Estimating a model with separate regressors for disadvantageous and advantageous inequality is problematic due to collinearity of regressors. Thus, in the analysis we use a utility specification with just one term for the payoff inequality associated with a choice, and this requires specifying weights for disadvantageous and advantageous inequality. We report the simplest specification with equal weights on disadvantageous and advantageous inequality. We also estimated the models by forming the inequality term with higher weight placed on disadvantageous inequality relative to advantageous inequality. These estimations give similar results; details available from the authors on request.

TABLE 2. Conditional logit regressions of Employee 2 effort in BASELINE.

	SELFISH	FS	NORMS	FS + NORMS
Own payoff (θ_1)	0.218***	0.481***	0.350***	0.479***
	(0.018)	(0.058)	(0.047)	(0.065)
			[0.035]	[0.016]
Payoff inequality (θ_2)		–0.101***		–0.101***
		(0.018)		(0.021)
				[0.006]
Norm rating (θ_3)			1.008***	–0.023
			(0.294)	(0.407)
			[0.245]	[0.208]
Observations	880	880	880	880
Log likelihood	–196.326	–175.281	–189.892	–175.279

Notes: dependent variable is the chosen action; standard errors in parentheses, with bootstrapped standard errors in brackets for the specifications with norm ratings; ***$p < 0.01$.

TABLE 3. Average and predicted mean effort for Employee 2 in BASELINE.

	Own wage = 16 Co-worker effort = 1	Own wage = 16 Co-worker effort = 4	Own wage = 32 Co-worker effort = 1	Own wage = 32 Co-worker effort = 4
Average effort	**1.163**	**1.163**	**1.418**	**2.073**
SELFISH	1.454	1.454	1.454	1.454
	[1.356, 1.553]	[1.356, 1.553]	[1.356, 1.553]	[1.356, 1.553]
FS	1.307	1.172	1.312	2.027
	[1.210, 1.404]	[1.122, 1.221]	[1.213, 1.412]	[1.753, 2.302]
NORMS	1.317	1.469	1.412	1.620
	[1.218, 1.415]	[1.395, 1.544]	[1.316, 1.507]	[1.450, 1.790]
FS + NORMS	1.309	1.169	1.313	2.027
	[1.183, 1.435]	[1.076, 1.263]	[1.212, 1.413]	[1.752, 2.302]

Notes: Employee 1 wage is 32 in all cases. 95% prediction interval in brackets.

often, whereas the NORMS specification shows that effort levels that are deemed more socially appropriate are chosen more often. When both norm ratings and payoff inequalities are incorporated (FS + NORMS model) the coefficient on payoff inequality is still negative and highly significant, but the coefficient on norm ratings is no longer significant.

Table 3 presents predicted mean effort of Employee 2 based on the conditional logit models.[14] As we discussed in previous sections the two distinctive features of the BASELINE results are the existence of own-wage effects and a peer effect when both employees are paid a high wage. Although own payoff is a highly significant explanatory variable in all four specifications reported in Table 2, the model where utility depends only on own payoff (SELFISH) cannot explain either of these features

14. That is, we calculate mean effort for each contingency using the estimates in Table 2 to compute predicted relative frequencies of each effort level.

of the data. The augmented models do explain some of these features, although to different extents. First, the models exhibit a predicted own-wage effect when the co-worker's effort is high. However, the magnitude of the effect is much milder in NORMS than in either the actual data or the specifications including payoff inequality, and in fact for this model the relevant prediction intervals overlap.[15] Second, the peer effect when both employees receive a high wage is exhibited in the predicted efforts of all augmented models. However, again, the effect is milder in NORMS, and again, the relevant prediction intervals in NORMS overlap. Overall, the conditional logit results suggest that the social preferences model offers a parsimonious explanation of the behavioral regularities observed in the BASELINE setting.

5. Disentangling Social Preferences and Social Norms: Two Further Treatments

The results from the BASELINE game suggest that the peer effects and pro-social behavior observed in our trilateral gift-exchange game can be explained by the social preferences approach. In this section we put this conclusion to a stress test by introducing two further treatments, ASYMMETRIC and RANDOM, where we modify the BASELINE game so as to manipulate the social preferences predictions of peer effects. Since these manipulations may also affect norms we conduct new norms elicitation experiments for these treatments.[16]

In the ASYMMETRIC treatment we recruited 153 new subjects to play the same game used in BASELINE except that, while the Employer can still choose a wage of 16 or 32 for Employee 1, the wage the Employer pays to Employee 2 can be either 16 or 30. Thus, in the wage combination where both wages are high Employee 1 is paid a wage of 32 and Employee 2 a slightly lower wage of 30.

In the RANDOM treatment we also recruited 153 new subjects to play the game used in BASELINE except that, after all members of the firm have made their decisions, one of the two employees is randomly selected with equal probability. Only the wage and effort decisions regarding the selected employee are implemented. In RANDOM the payoff functions (1) and (2) above are then modified such that the Employer's earnings are computed as[17]

$$\pi_{ER} = \begin{cases} 20 + 10 \cdot (e_1) - w_1 & \text{if Employee 1 is selected} \\ 20 + 10 \cdot (e_2) - w_2 & \text{if Employee 2 is selected} \end{cases} \tag{1'}$$

15. Note, however, that the augmented models predict only a mild own-wage effect when the co-worker chooses minimal effort, and in fact the associated prediction intervals overlap. This appears to be because all models overpredict effort for the contingency where own wage is low and co-worker's effort is low.

16. These additional treatments use the same subject pool, recruitment, and experimental procedures as for the 2009 BASELINE sessions and the BASELINE norms elicitation session. Details can be found in the Online Appendix.

17. Note that a constant is added to the Employer's payoff in RANDOM. As explained in detail in the next section, this is done to refine our theoretical predictions for the RANDOM treatment.

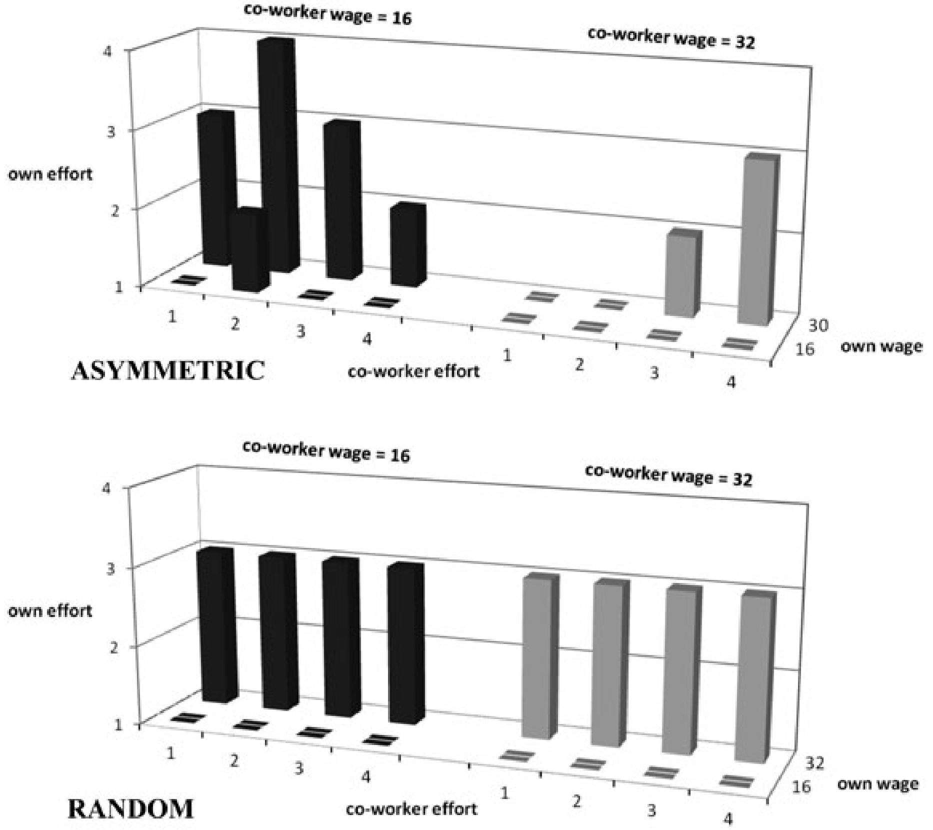

FIGURE 4. FS prediction of Employee 2 effort for ASYMMETRIC & RANDOM Predicted Employee 2 effort choices for $\beta_2 \geq 1/2$. For ASYMMETRIC the exact prediction depends also onα_2. The figure displays predicted effort for $1/2 \leq \beta_2 \leq 10/19 + 1/19\,\alpha_2$. The predictions for other cases are similar and are reproduced in the Online Appendix. In particular, for all cases there is a positive correlation between employees' effort in the high wages combination.

and employee i's earnings are computed as

$$\pi_i = \begin{cases} w_i - 5 \cdot (e_i - 1) & \text{if Employee } i \text{ is selected} \\ 0 & \text{if Employee } i \text{ is not selected.} \end{cases} \tag{2$'$}$$

5.1. Explaining Peer Effects in ASYMMETRIC and RANDOM

What are the implications of these manipulations for the social preferences predictions of peer effects? As in BASELINE, the FS model predicts minimal effort if $\beta_2 < 1/2$. Otherwise, if $\beta_2 \geq 1/2$, Employee 2 may choose non-minimal effort; see Figure 4.

Relative to the BASELINE game, the small perturbation in Employee 2's high wage in ASYMMETRIC, weakens, but does not eliminate, the peer effect predicted by the FS model in the wage combination where both employees receive a high wage. On the other hand, in the RANDOM treatment the FS model yields dramatically different predictions relative to BASELINE: while the FS model still predicts a positive own-wage effect, it predicts *no* relation between employees' efforts. Intuitively, in the RANDOM treatment the decision of Employee 2 can only affect earnings and the distribution of earnings in the state of the world where she is selected. In this state of the world the co-worker's effort and wage have no impact on earnings or the distribution of earnings, and hence cannot affect the choices of an individual with FS preferences.[18]

These modifications to the BASELINE setting may also affect social norms. For example, the small wage asymmetries introduced in ASYMMETRIC may affect norm compliance due to the existence of "moral wiggle" opportunities (Dana, Weber, and Kuang 2007) that may weaken or eliminate the descriptive norm regulating the relationship between employees' efforts. The reason is that employees are now in different situations and what one employee does may no longer provide a relevant benchmark for what is appropriate behavior by the other.[19] To elicit social norms in the two new treatments, we conducted two further norms elicitation experiments using the same procedures used for the norms elicitation experiment described in Section 3. We recruited 93 new subjects who were given a description of four situations reproducing four of the 16 possible contingencies faced by Employees 2 either in the ASYMMETRIC (46 subjects) or in the RANDOM (47 subjects) treatment. The situations corresponded to those used in the norms-elicitation experiment for the BASELINE game.

Table 4 shows the results of the norms elicitation experiments. For each situation, the Table shows the average social appropriateness rating of the effort choices available to Employee 2. As in BASELINE, low effort receives a higher appropriateness rating when Employee 2 receives a low wage (i.e. in Situations 1 and 2) than when Employee 2 receives a high wage (i.e. in Situations 3 and 4). In contrast high effort receives a higher appropriateness rating when the own wage is high rather than low.[20] Also as in BASELINE, the co-worker's effort systematically affects appropriateness ratings when both employees receive a high wage: low effort is generally deemed more appropriate when the co-worker chooses low effort (Situation 3) than when the co-worker chooses

18. The prediction of no correlation between employees' efforts in RANDOM is robust across models of social preferences where utilities are defined over outcomes. However, if utilities are defined over expected outcomes, as in Trautmann (2009), employees' efforts may be correlated. In order to exclude this possibility we added 20 points to the Employer's payoff such that Employee 2 is worse off than the Employer in expectation and thus is generally unwilling to supply effort above the minimal level.

19. Dana, Weber, and Kuang (2007) discuss the existence of moral wriggle room across variants of the dictator game. For other studies providing evidence of exploitation of moral wriggle room possibilities see Haisley and Weber (2010), Xiao and Bicchieri (2010), and Thöni and Gächter (2011). Also related are the studies on exit options in dictator games by Dana, Cain, and Dawes (2006); Broberg, Ellingsen, and Johannesson (2007); and Lazear, Malmendier, and Weber (2012). Other related studies also point to the existence of a *self-serving* or *egocentric bias* when one's self-interest is at stake: see Konow (2005) for an overview and related literature. In the context of the gift exchange game, see Charness and Haruvy (2000).

20. According to two-sided Wilcoxon signed-rank tests the differences in ratings between Situation 1 and 3 are only significant in RANDOM ($p \leq 0.001$ for each possible effort choice). The differences between Situation 2 and 4 are highly significant in both treatments for effort equals 1, 2, or 4 ($p < 0.001$), but not for effort equals 3: $p = 0.180$ in ASYMMETRIC; $p = 0.105$ in RANDOM).

TABLE 4. Elicited norms—ASYMMETRIC & RANDOM.

ASYMMETRIC				
Employee 2 effort	Situation 1 ($w_1 = 32$; $w_2 = 16$; $e_1 = 1$)	Situation 2 ($w_1 = 32$; $w_2 = 16$; $e_1 = 4$)	Situation 3 ($w_1 = 32$; $w_2 = 30$; $e_1 = 1$)	Situation 4 ($w_1 = 32$; $w_2 = 30$; $e_1 = 4$)
1	0.98	1.15	0.98	0.26
	(1.18)	*(1.03)*	*(1.26)*	*(0.57)*
2	1.43	2.28	1.39	1.00
	(0.69)	*(0.65)*	*(0.74)*	*(0.52)*
3	1.67	1.78	1.63	1.98
	(0.67)	*(0.76)*	*(0.71)*	*(0.54)*
4	1.63	0.93	1.65	2.74
	(1.32)	*(1.04)*	*(1.21)*	*(0.68)*
RANDOM				
Employee 2 effort	Situation 1 ($w_1 = 32$; $w_2 = 16$; $e_1 = 1$)	Situation 2 ($w_1 = 32$; $w_2 = 16$; $e_1 = 4$)	Situation 3 ($w_1 = 32$; $w_2 = 32$; $e_1 = 1$)	Situation 4 ($w_1 = 32$; $w_2 = 32$; $e_1 = 4$)
1	1.96	1.38	0.77	0.25
	(1.08)	*(1.01)*	*(1.07)*	*(0.71)*
2	2.02	2.19	1.53	1.02
	(0.74)	*(0.88)*	*(0.69)*	*(0.57)*
3	1.42	1.62	2.11	1.87
	(0.77)	*(0.78)*	*(0.73)*	*(0.57)*
4	0.87	1.06	2.00	2.70
	(1.26)	*(1.31)*	*(1.14)*	*(0.69)*

Notes: Subjects' appropriateness ratings were converted into numerical scores using the following scale: very socially inappropriate = 0; somewhat socially inappropriate = 1; somewhat socially appropriate = 2; very socially appropriate = 3. Standard deviations in parentheses.

high effort (Situation 4), while high effort is deemed more appropriate when the co-worker chooses high effort than when she chooses low effort.[21]

Figure 5 shows the implications of our elicited norms for the ASYMMETRIC and RANDOM treatments, using the same approach as that for the BASELINE game described in Section 3. That is, we examine the efforts that our raters would expend if, in each situation, they chose the effort they deemed to be most socially appropriate.

As in the BASELINE game, our norm-elicitation experiments reveal a reciprocity norm: the average effort increases with own wage.[22] Also as in the BASELINE game, our norm-elicitation experiments reveal a peer effect in that effort depends positively on the co-worker's effort when the own wage is high. In ASYMMETRIC, when

21. According to two-sided Wilcoxon signed-rank tests, the differences in ratings between Situation 3 and 4 are significant for each possible effort choice, both in ASYMMETRIC ($p \leq 0.005$) and RANDOM ($p \leq 0.031$).

22. The exception to this is when the co-worker's effort is low in the ASYMMETRIC treatment. In this case the difference in average efforts is not significant (two-sided Wilcoxon signed-rank test: $p = 0.382$). The own-wage effect is highly significant in RANDOM when the co-worker's effort is low, and in either treatment when the co-worker's effort is high ($p < 0.001$).

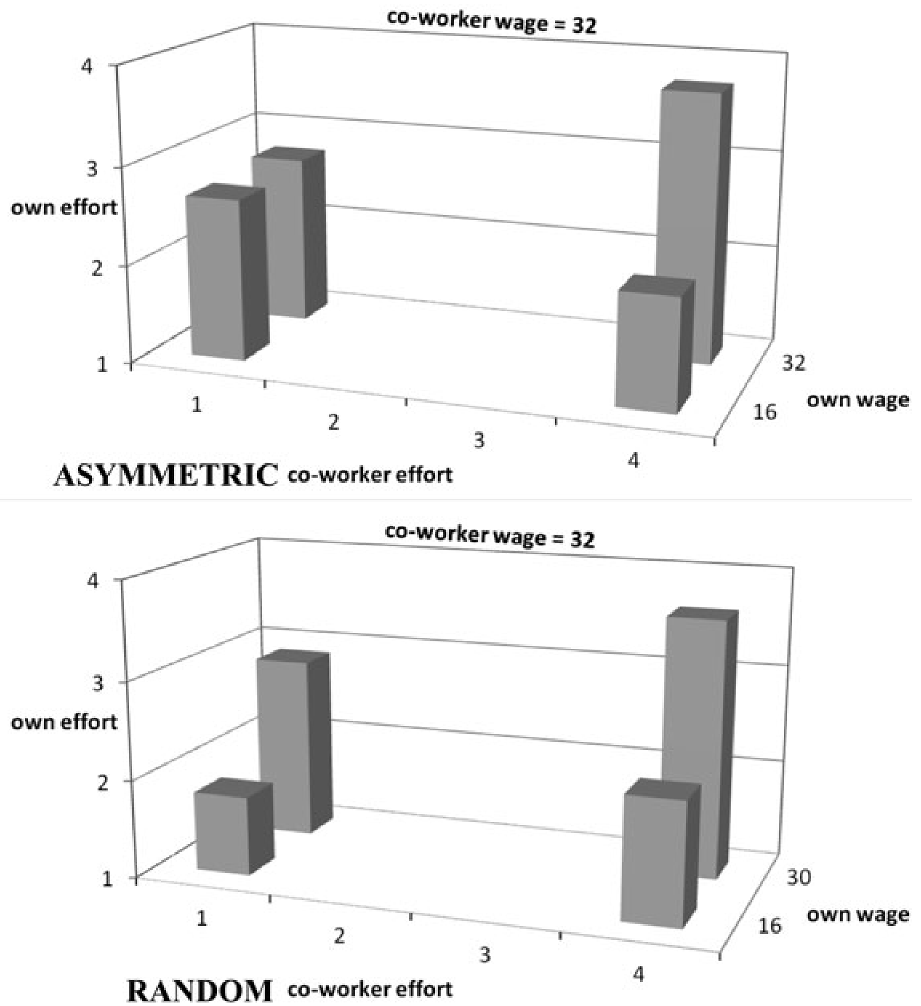

FIGURE 5. Socially most appropriate effort choices in ASYMMETRIC & RANDOM. Effort choices that the raters would make on average if they chose the most socially appropriate effort level in the four situations they evaluated.

the co-worker's effort increases from 1 to 4 units, average effort increases by 1 unit (from 2.74 to 3.74, two-sided Wilcoxon signed-rank test: $p < 0.001$). In RANDOM, the corresponding increase in average effort is 0.75 units (from 2.87 to 3.62, two-sided Wilcoxon signed-rank test: $p < 0.001$). The strength of these peer effects is not different from that inferred from most socially appropriate behavior in BASELINE (where average effort increases by 0.90 units) in either ASYMMETRIC ($p = 0.724$) or RANDOM ($p = 0.511$) according to two-sided Wilcoxon rank-sum tests. The comparison between BASELINE and ASYMMETRIC is particularly interesting since it shows that the elicited norms do not capture a "moral wiggle-room" effect.

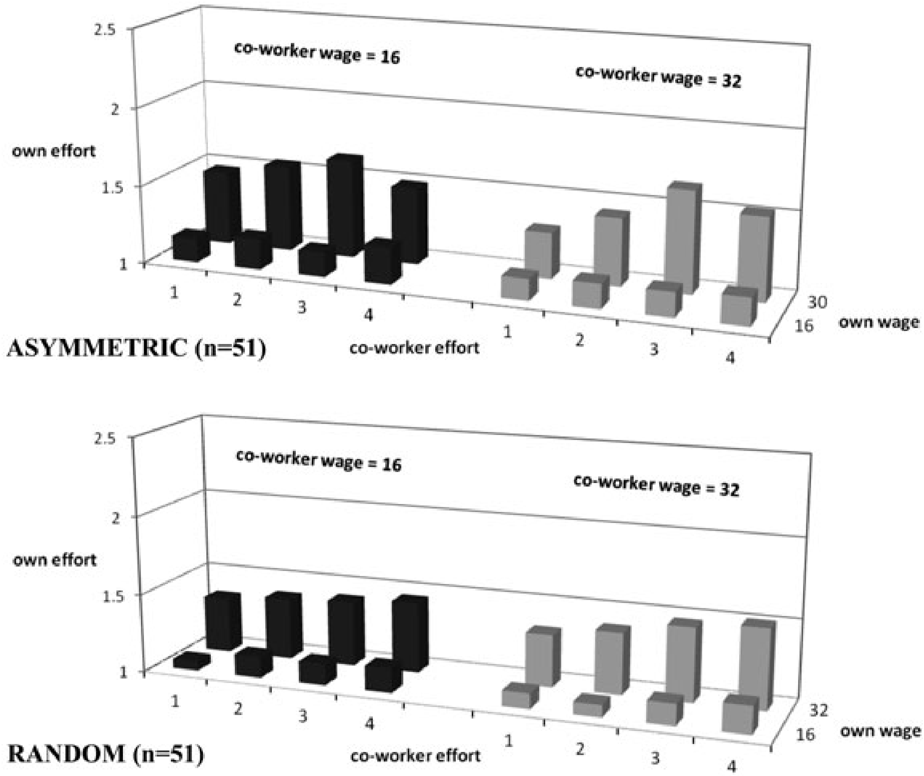

FIGURE 6. Employee 2 average effort—ASYMMETRIC & RANDOM.

5.2. Behavioral Data

We turn now to the actual effort choices of Employees 2 in the ASYMMETRIC and RANDOM treatments. Figure 6 shows the average effort choices for each contingency.

As in BASELINE, there are clear own-wage effects in the ASYMMETRIC and RANDOM treatments. In each of these treatments, Employee 2's effort when the own wage is high is significantly higher than when the own wage is low, across all eight cases shown in both panels of Figure 6 ($p \leq 0.038$ according to two-sided Wilcoxon signed-rank tests). Note, however, that in general efforts tend to be lower than in BASELINE. In fact, the fraction of Employees 2 who behave "nonselfishly" (i.e. who make at least one nonminimal effort choice) is significantly lower in RANDOM than in BASELINE (29% versus 60%, $\chi^2(1) = 9.99$, $p = 0.002$).[23] The higher proportion of Employees 2 choosing minimal effort in RANDOM relative to BASELINE is difficult to reconcile with the FS model, as the threshold value of the superiority aversion parameter β_2 by which Employees 2 are predicted to expend non-minimal effort is the same in both

23. The fraction of "nonselfish" Employees 2 in ASYMMETRIC is instead not different than in BASELINE (49%, $\chi^2(1) = 1.288$, $p = 0.256$).

treatments. A possible explanation is that individuals may dislike inequality in expected outcomes, as in Trautmann (2009). In RANDOM these individuals are predicted to choose minimal effort most of the time, and this may explain a higher fraction of minimal effort choices in RANDOM.

Turning to peer effects, Figure 6 shows that in ASYMMETRIC Employee 2's effort tends to increase, albeit nonmonotonically, with the co-worker's effort. A one-tailed Page test for ordered alternatives shows that this positive correlation is statistically significant ($p = 0.033$). Nevertheless, the observed peer effect appears weaker than in BASELINE: focusing on the extreme cases where both wages are high and the co-worker's effort increases from 1 to 4 units, Employee 2's effort increases by 0.24 units in ASYMMETRIC (from 1.29 to 1.53). The corresponding average effort increase in BASELINE is 0.65. The apparently weaker peer effect in ASYMMETRIC compared to BASELINE is consistent with the predictions of the social preferences model or a moral wiggle-room effect (although recall that, somewhat surprisingly, the elicited norms did not capture any moral wiggle-room effect). However, the difference between the strength of peer effects in ASYMMETRIC and BASELINE is just insignificant according to a two-sided Wilcoxon rank-sum test: $p = 0.108$.

In the RANDOM treatment Employee 2's effort slightly increases as the co-worker expends higher effort: again focusing on the cases where the co-worker's effort increases from 1 to 4 units, Employee 2's effort increases by 0.18 units (from 1.33 to 1.51). This peer effect is *not* significant according to a one-tailed Page test for ordered alternatives ($p = 0.198$).[24] The absence of a peer effect in RANDOM is consistent with the social preferences model.

To examine formally the explanatory power of the social preferences model vis-à-vis the social norms approach, we apply the conditional logit models introduced in Section 4 to the choice data from ASYMMETRIC and RANDOM.[25] Table 5 presents the results.

As in BASELINE, the own payoff variable is positive and highly significant in both treatments, confirming the importance of own payoff considerations for Employee 2 effort decisions. Also as in BASELINE, adding a payoff inequality regressor improves explanatory power in both treatments. Turning to the NORMS specification, the coefficient on norms rating is positive and highly significant in RANDOM, but is not significant in ASYMMETRIC. In the most general FS + NORMS specification the coefficient on payoff inequality is negative and highly significant in both treatments

24. The strength of the peer effect in RANDOM is significantly different from BASELINE according to a two-sided Wilcoxon rank sum test: $p = 0.011$. In fact, in RANDOM only four out of 51 Employees 2 monotonically increase effort when the co-worker chooses higher effort in the high-wages combination. By comparison, in BASELINE 19 out of 55 Employees 2 monotonically increase their effort with the co-worker's effort in the high wages combination. This difference in proportions is highly significant ($\chi^2(1) = 11.10, p = 0.001$).

25. As in BASELINE, we have only elicited norms ratings in ASYMMETRIC and RANDOM for four of the possible contingencies available to Employee 2, and we therefore estimate the conditional logit models only using data from these contingencies. For RANDOM we use expected own payoffs and expected inequality in payoffs as regressors.

TABLE 5. Conditional logit regressions of Employee 2 effort in ASYMMETRIC & RANDOM.

ASYMMETRIC	SELFISH	FS	NORMS	FS + NORMS
Own payoff (θ_1)	0.296***	0.408***	0.327***	0.389***
	(0.024)	(0.061)	(0.048)	(0.066)
			[0.019]	[0.025]
Payoff inequality (θ_2)		−0.049**		−0.058***
		(0.022)		(0.026)
				[0.006]
Norm rating (θ_3)			0.297	−0.366
			(0.372)	(0.495)
			[0.193]	[0.284]
Observations	816	816	816	816
Log likelihood	−137.985	−134.677	−137.660	−134.398
RANDOM	SELFISH	FS	NORMS	FS + NORMS
Own payoff (θ_1)	0.597***	0.686***	0.711***	0.669***
	(0.049)	(0.067)	(0.091)	(0.092)
			[0.041]	[0.044]
Payoff inequality (θ_2)		−0.054***		−0.060***
		(0.021)		(0.031)
				[0.016]
Norm rating (θ_3)			0.587***	−0.130
			(0.343)	(0.515)
			[0.200]	[0.342]
Observations	816	816	816	816
Log likelihood	−136.499	−132.732	−134.773	−132.700

Notes: Dependent variable is the chosen action; standard errors in parentheses, with bootstrapped standard errors in brackets for the specifications with norm ratings; $^{***}p < 0.01$, $^{**}p < 0.05$.

while the norms rating coefficient has the "wrong" sign and is statistically insignificant in both treatments.

These results mirror our findings from BASELINE, and show that, for each of our settings choice behavior can be explained parsimoniously by the social preferences model. Note, however, that the magnitude of coefficient estimates changes across treatments in all specifications. Compare for example the BASELINE and RANDOM results for the FS specification. The coefficient on payoff inequality in RANDOM is roughly half that of BASELINE, while the coefficient on own payoff is about one and a half times greater in RANDOM than BASELINE. This reflects the greater proportion of selfish choices in the RANDOM treatment noted earlier. For a formal test of parameter stability across treatments we estimated conditional logit models using data from all treatments, comparing models that restrict parameters to be the same across treatments with unrestricted models allowing parameters to vary.[26] We reject easily the hypothesis that model parameters are invariant to treatment ($p < 0.001$ for all specifications, based on likelihood-ratio tests).

Table 6 presents predicted mean effort of Employee 2 from the conditional logit models for the ASYMMETRIC and RANDOM treatments. The predicted efforts

26. See the Appendix for details on the regressions output.

TABLE 6. Average and predicted mean effort for Employee 2 in ASYMMETRIC & RANDOM.

ASYMMETRIC	Own wage = 16 Co-worker effort = 1	Own wage = 16 Co-worker effort = 4	Own wage = 30 Co-worker effort = 1	Own wage = 30 Co-worker effort = 4
Average effort	**1.137**	**1.176**	**1.294**	**1.529**
SELFISH	1.284	1.284	1.284	1.284
	[1.204, 1.364]	[1.204, 1.364]	[1.204, 1.364]	[1.204, 1.364]
FS	1.258	1.176	1.260	1.442
	[1.175, 1.342]	[1.110, 1.243]	[1.176, 1.344]	[1.261, 1.623]
NORMS	1.268	1.290	1.266	1.314
	[1.182, 1.354]	[1.214, 1.366]	[1.177, 1.354]	[1.198, 1.429]
FS + NORMS	1.277	1.145	1.283	1.432
	[1.174, 1.380]	[1.047, 1.244]	[1.173, 1.393]	[1.253, 1.611]
RANDOM	own wage = 16 co-worker effort = 1	own wage = 16 co-worker effort = 4	own wage = 32 co-worker effort = 1	own wage = 32 co-worker effort = 4
Average effort	**1.098**	**1.176**	**1.333**	**1.510**
SELFISH	1.279	1.279	1.279	1.279
	[1.200, 1.358]	[1.200, 1.358]	[1.200, 1.358]	[1.200, 1.358]
FS	1.173	1.173	1.386	1.386
	[1.089, 1.257]	[1.089, 1.257]	[1.258, 1.513]	[1.258, 1.513]
NORMS	1.188	1.267	1.314	1.349
	[1.089, 1.286]	[1.201, 1.333]	[1.223, 1.405]	[1.223, 1.474]
FS + NORMS	1.180	1.166	1.390	1.382
	[1.077, 1.283]	[1.065, 1.266]	[1.256, 1.525]	[1.252, 1.511]

Notes: Employee 1 wage is 32 in all cases. 95% prediction interval in brackets.

bear out the results of nonparametric statistical tests for some specifications but not others. First, nonparametric tests detect clear own-wage effects in both treatments, but models that do not include a term for payoff inequality (i.e. SELFISH and NORMS) either do not predict own-wage effects or the relevant prediction intervals overlap. The FS + NORMS specification does show a clear own-wage effect in one case (the ASYMMETRIC treatment when the co-worker's effort is 4), while the FS specification shows clear own-wage effects in both treatments when the co-worker chooses high effort and in the RANDOM treatment when the co-worker chooses low effort. Second, nonparametric tests detect a peer effect in the high wage combination of the ASYMMETRIC treatment. In all specifications except SELFISH, where the predicted efforts are necessarily constant across situations, predicted effort in the high wage combination of ASYMMETRIC is higher when the co-worker chooses high effort than when the co-worker chooses low effort. However, for all three specifications, the associated prediction intervals overlap. This is perhaps consistent with our earlier observation that the peer effect in ASYMMETRIC is weaker than in BASELINE.[27]

27. This might also reflect the fact that the conditional logit regressions use less data than the nonparametric tests. The peer effect in ASYMMETRIC, significant at the 5% level according to a one-sided Page test, compares effort choices across contingencies where the co-worker chooses 1, 2, 3, or 4 units

6. Discussion and Conclusions

Our study examines whether peer effects in reciprocity are best organized by social norms compliance or by a social preferences model (the Fehr and Schmidt 1999 inequity aversion model, FS). Our data favor the social preferences approach. Although we do find evidence for the empirical relevance of social norms in our experiments, an econometric horse-race reveals that norm compliance does not explain behavior once we control for payoff inequality.

In our analysis we concentrated on inequity aversion as formulated by FS because a theoretical analysis of the three-player gift exchange game by Thöni and Gächter (2011) revealed that most other standard theories of social preferences (theories that model various distributional preferences and/or intentions) predict either no peer effect (Dufwenberg and Kirchsteiger 2004; Levine 1998) or unambiguously negatively-related efforts (Cox, Friedman, and Gjerstad 2007; Cox, Friedman, and Sadiraj 2008; Bolton and Ockenfels 2000; Falk and Fischbacher 2006). Two theories are generally consistent with positively as well as negatively related efforts: Charness and Rabin (2002) and FS. However, for our parameters, only the latter predicts a peer effect.[28]

While the FS model gives a reasonable account of the observed patterns of pro-social behavior within each treatment of the experiment, we also observe differences across treatments in the proportion of individuals who engage in pro-social behavior. These differences occur despite the fact that the threshold value of the advantageous inequity parameter by which an employee is predicted to expend non-minimal effort is the same in all treatments, and that subjects are randomly assigned to treatment conditions. Moreover, our conditional logit analysis shows that a stable distribution of preferences parameters is inconsistent with the pooled choice data from our three treatments. These results are more difficult to reconcile with the FS model predictions.[29] As suggested by Thöni and Gächter (2011), newer models that take into account other motives than distributional concerns and/or intentions may also be successful in explaining peer effects in reciprocity: for example, models of conformity (Sliwka 2007) and social esteem (Bénabou and Tirole 2006; Ellingsen and Johannesson 2008). While our experiments are not designed to assess the predictive power of models of conformity or social esteem, this seems a promising task for future research.

of effort, while the conditional logit regressions only use the data from the contingencies where the co-worker's effort is 1 or 4. It is possible to estimate the FS conditional logit model using all the choice data, in which case the estimated peer effect is weaker but the standard errors are smaller implying that the prediction intervals do not overlap (the prediction interval when the co-worker chooses effort 1 is 1.313 ± 0.046, while the prediction interval when co-worker's chooses effort 4 is 1.461 ± 0.074).

28. The Charness and Rabin (2002) model assumes that Employee 2 directly compares her payoff with Employee 1 payoff when Employee 1 is the least well-off firm member. In our setting this does not occur in the wage combination where both wages are high, and so the model does not predict a peer effect for this wage combination. Of course, suitably modified versions of the model may be able to generate predicted peer effects for our game.

29. Other studies have noticed the limited success of social preference models in predicting when and whether an individual will deviate from self-interest (e.g., Blanco, Engelmann, and Normann 2011; Herrmann and Orzen 2008).

Appendix: Conditional Logit Models Using Pooled Data from All Treatments

Table A.1 presents the results of conditional logit models that use pooled data from the three treatments of the experiment. The restricted models hold the coefficients constant across treatments. The unrestricted models allow different coefficients in the different treatments.

TABLE A.1. Conditional logit regressions of Employee 2 effort, pooled data.

Restricted models	**SELFISH**	**FS**	**NORMS**	**FS + NORMS**
Own payoff	0.301***	0.502***	0.346***	0.460***
	(0.015)	(0.034)	(0.024)	(0.034)
			[0.015]	[0.016]
Payoff inequality		−0.086***		−0.113***
		(0.011)		(0.014)
				[0.006]
Norm rating			0.359***	−0.779***
			(0.143)	(0.230)
			[0.124]	[0.164]
Observations	2512	2512	2512	2512
Log likelihood	−505.757	−467.523	−502.490	−461.379
Unrestricted models	SELFISH	FS	NORMS	FS + NORMS
Own payoff	0.218***	0.481***	0.350***	0.479***
	(0.018)	(0.058)	(0.047)	(0.065)
			[0.034]	[0.016]
Own payoff * ASYMMETRIC	0.078***	−0.073	−0.023	−0.091***
	(0.030)	(0.084)	(0.067)	(0.092)
			[0.040]	[0.028]
Own payoff * RANDOM	0.379***	0.205**	0.361***	0.190***
	(0.052)	(0.088)	(0.103)	(0.113)
			[0.052]	[0.047]
Payoff inequality		−0.101***		−0.101***
		(0.018)		(0.021)
				[0.006]
Payoff inequality * ASYMMETRIC		0.052*		0.043***
		(0.028)		(0.033)
				[0.008]
Payoff inequality * RANDOM		0.047*		0.042**
		(0.027)		(0.037)
				[0.017]
Norm rating			1.008***	−0.023
			(0.294)	(0.407)
			[0.241]	[0.204]
Norm rating * ASYMMETRIC			−0.711**	−0.343
			(0.475)	(0.640)
			[0.317]	[0.348]
Norm rating * RANDOM			−0.421	−0.107
			(0.452)	(0.657)
			[0.311]	[0.401]
Observations	2512	2512	2512	2512
Log likelihood	−470.810	−442.690	−462.325	−442.377

Notes: Dependent variable is the chosen action; standard errors in parentheses, with bootstrapped standard errors in brackets for the specifications with norm ratings; $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

References

Abeler, Johannes, Steffen Altmann, Sebastian Kube, and Matthias Wibral (2010). "Gift Exchange and Workers' Fairness Concerns: When Equality Is Unfair." *Journal of the European Economic Association*, 8, 1299–1324.

Alpizar, F., F. Carlsson, and O. Johansson-Stenman (2008). "Anonymity, Reciprocity, and Conformity: Evidence from Voluntary Contributions to a National Park in Costa Rica." *Journal of Public Economics*, 92, 1047–1060.

Bénabou, Roland, and Jean Tirole (2006). "Incentives and Prosocial Behavior." *American Economic Review*, 96, 1652–1678.

Bicchieri, Cristina (2005). *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge University Press.

Bicchieri, Cristina, and Erte Xiao (2009). "Do the Right Thing: But Only If Others Do So." *Journal of Behavioral Decision Making*, 22, 191–208.

Blanco, Mariana, Dirk Engelmann, and Hans Theo Normann (2011). "A within-Subject Analysis of Other-Regarding Preferences." *Games and Economic Behavior*, 72, 321–338.

Bolton, Gary E., and Axel Ockenfels (2000). "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review*, 90, 166–193.

Brandts, Jordi, and Gary Charness (2011). "The Strategy Versus the Direct-Response Method: A First Survey of Experimental Comparisons." *Experimental Economics*, 14, 375–398.

Broberg, Tomas, Tore Ellingsen, and Magnus Johannesson (2007). "Is Generosity Involuntary?" *Economics Letters*, 94, 32–37.

Burks, Stephen V., and Erin L. Krupka (2012). "A Multimethod Approach to Identifying Norms and Normative Expectations within a Corporate Hierarchy: Evidence from the Financial Services Industry." *Management Science*, 58, 203–217.

Camerer, Colin F. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press.

Cason, Tim, and Vai- Lam Mui (1998). "Social Influence in the Sequential Dictator Game." *Journal of Mathematical Psychology*, 42, 248–265.

Charness, Gary, and Ernan Haruvy (2000). "Self-Serving Biases: Evidence from a Simulated Labour Relationship." *Journal of Managerial Psychology*, 15, 655–667.

Charness, Gary, and Peter Kuhn (2007). "Does Pay Inequality Affect Worker Effort? Experimental Evidence." *Journal of Labor Economics*, 25, 693–723.

Charness, Gary, and Matthew Rabin (2002). "Understanding Social Preferences with Simple Tests." *Quarterly Journal of Economics*, 117, 817–869.

Chen, Yan, F. Maxwell Harper, Joseph Konstan, and Sherry Xin Li (2010). "Social Comparisons and Contributions to Online Communities: A Field Experiment on Movielens." *American Economic Review*, 100, 1358–1398.

Cox, James C., Daniel Friedman, and Steven Gjerstad (2007). "A Tractable Model of Reciprocity and Fairness." *Games and Economic Behavior*, 59, 17–45.

Cox, James C., Daniel Friedman, and Vjollca Sadiraj (2008). "Revealed Altruism." *Econometrica*, 76, 31–69.

Croson, Rachel, and Jen Shang (2008). "The Impact of Downward Social Information on Contribution Decisions." *Experimental Economics*, 11, 221–233.

Dana, Jason, Daylian M. Cain, and Robyn M. Dawes (2006). "What You Don't Know Won't Hurt Me: Costly (but Quiet) Exit in Dictator Games." *Organizational Behavior and Human Decision Processes*, 100, 193–201.

Dana, Jason, Roberto Weber, and Jason Kuang (2007). "Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness." *Economic Theory*, 33, 67–80.

Dufwenberg, Martin, and Georg Kirchsteiger (2004). "A Theory of Sequential Reciprocity." *Games and Economic Behavior*, 47, 268–298.

Ellingsen, Tore, and Magnus Johannesson (2008). "Pride and Prejudice: The Human Side of Incentive Theory." *American Economic Review*, 98, 990–1008.

Falk, Armin, and Urs Fischbacher (2006). “A Theory of Reciprocity.” *Games and Economic Behavior*, 54, 293–315.
Falk, Armin, Urs Fischbacher, and Simon Gächter (2013). “Living in Two Neighborhoods—Social Interaction Effects in the Laboratory.” *Economic Inquiry*, 51, 563–578
Fehr, Ernst, Georg Kirchsteiger, and Arno Riedl (1993). “Does Fairness Prevent Market Clearing? An Experimental Investigation.” *Quarterly Journal of Economics*, 108, 437–459.
Fehr, Ernst, and Klaus M. Schmidt (1999). “A Theory of Fairness, Competition, and Cooperation.” *Quarterly Journal of Economics*, 114, 817–868.
Fischbacher, Urs (2007). “Z-Tree: Zurich Toolbox for Readymade Economic Experiments.” *Experimental Economics*, 10, 171–178.
Fischbacher, Urs, and Simon Gächter (2010). “Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Good Experiments.” *American Economic Review*, 100, 541–556.
Frey, Bruno S., and Stephan Meier (2004). “Social Comparisons and Pro-Social Behavior. Testing ‘Conditional Cooperation’ in a Field Experiment.” *American Economic Review*, 94, 1717–1722.
Gächter, Simon, Daniele Nosenzo, and Martin Sefton (2012). “The Impact of Social Comparisons on Reciprocity.” *The Scandinavian Journal of Economics*, 114, 1346–1367.
Gächter, Simon, and Christian Thöni (2010). “Social Comparison and Performance: Experimental Evidence on the Fair Wage-Effort Hypothesis.” *Journal of Economic Behavior and Organization*, 76, 531–543.
Greiner, Ben (2004). “An Online Recruitment System for Economic Experiments.” In *Forschung und wissenschaftliches Rechnen GWDG Bericht 63*, edited by K. Kremer, and V. Macho, (Eds.). Gesellschaft für wissenschaftliche Datenverarbeitung.
Güth, Werner, Manfred Königstein, Judit Kovács, and Eniko Zala-Mezõ (2001). “Fairness within Firms: The Case of One Principal and Many Agents.” *Schmalenbach Business Review*, 53, 82–101.
Haisley, Emily C., and Roberto A. Weber (2010). “Self-Serving Interpretations of Ambiguity in Other-Regarding Behavior.” *Games and Economic Behavior*, 68, 614–625.
Herrmann, Benedikt, and Henrik Orzen (2008). “The Appearance of Homo Rivalis: Social Preferences and the Nature of Rent-Seeking.” CeDEx Discussion Paper 2008–10, University of Nottingham.
Houser, Daniel, and Erte Xiao (2011). “Classification of Natural Language Messages Using A coordination Game.” *Experimental Economics*, 14, 1–14.
Keizer, Kees, Siegwart Lindenberg, and Linda Steg (2008). “The Spreading of Disorder.” *Science*, 322, 1681–1685.
Konow, James (2005). “Blind Spots: The Effects of Information and Stakes on Fairness Bias and Dispersion.” *Social Justice Research*, 18, 349–390.
Krupka, Erin, Stephen Leider, and Ming Jiang (2012). “A Meeting of the Minds: Contracts and Social Norms.” Working paper, University of Michigan.
Krupka, Erin, and Roberto A. Weber (2009). “The Focusing and Informational Effects of Norms on Pro-Social Behavior.” *Journal of Economic Psychology*, 30, 307–320.
Krupka, Erin, and Roberto A. Weber (2013). “Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary?” *Journal of the European Economic Association*, 11, 495–524.
Lazear, Edward P., Ulrike Malmendier, and Roberto A. Weber (2012). “Sorting in Experiments with Application to Social Preferences.” *American Economic Journal: Applied Economics*, 4, 136–163.
Levine, David K. (1998). “Modeling Altruism and Spitefulness in Experiments.” *Review of Economic Dynamics*, 1, 593–622.
López-Pérez, Raúl (2008). “Aversion to Norm-Breaking: A Model.” *Games and Economic Behavior*, 64, 237–267.
Martin, Richard, and John Randal (2008). “How Is Donation Behaviour Affected by the Donations of Others?” *Journal of Economic Behavior and Organization*, 67, 228–238.

Mittone, Luigi, and Matteo Ploner (2011). "Peer Pressure, Social Spillovers, and Reciprocity: An Experimental Analysis." *Experimental Economics*, 14, 203–222.

Nosenzo, Daniele (2011). "Pay Secrecy and Effort Provision." *Economic Inquiry*, in press.

Rustagi, Devesh, Stefanie, Engel, and Michael Kosfeld (2010). "Conditional Cooperation and Costly Monitoring Explain Success in Forest Commons Management." *Science*, 330, 961–965.

Selten, Reinhard (1967). "Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperimentes." In *Beiträge zur experimentellen Wirtschaftsforschung*, edited by H. Sauermann, (Ed.). J. C. B. Mohr (Paul Siebeck).

Shang, Jen, and Rachel Croson (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, Dirk (2007). "Trust as a Signal of a Social Norm and the Hidden Costs of Incentive Schemes." *The American Economic Review*, 97, 999–1012.

Thöni, Christian, and Simon Gächter (2011). "Peer Effects and Social Preferences in Voluntary Cooperation." CeDEx Discussion Paper 2011–09, University of Nottingham.

Trautmann, Stefan T. (2009). "A Tractable Model of Process Fairness under Risk." *Journal of Economic Psychology*, 30, 803–813.

Xiao, E., and D. Houser (2005). "Emotion Expression in Human Punishment Behavior." *Proceedings of the National Academy of Sciences of the United States of America*, 102, 7398–7401.

Xiao, Erte, and Cristina Bicchieri (2010). "When Equality Trumps Reciprocity." *Journal of Economic Psychology*, 31, 456–470.